Building Categories in Interaction

Linguistic resources at work

Edited by

Caterina Mauri

Ilaria Fiorentini

Eugenio Goria

JOHN BENJAMINS PUBLISHING COMPANY

Building Categories in Interaction

Studies in Language Companion Series (SLCS)

This series has been established as a companion series to the periodical *Studies in Language*.

For an overview of all books published in this series, please see
benjamins.com/catalog/slcs

Volume 220

Building Categories in Interaction. Linguistic resources at work
Edited by Caterina Mauri, Ilaria Fiorentini and Eugenio Goria

Building Categories in Interaction

Linguistic resources at work

Edited by

Caterina Mauri
University of Bologna

Ilaria Fiorentini
University of Pavia

Eugenio Goria
University of Turin

John Benjamins Publishing Company
Amsterdam / Philadelphia

Table of contents

CHAPTER 1

Building categories in interaction
Theoretical and empirical perspectives

Caterina Mauri, Ilaria Fiorentini and Eugenio Goria

University of Bologna / University of Pavia / University of Turin

1. On categories and categorization within linguistics

In the second half of the 20th century, path-breaking studies in the field of cognitive sciences radically changed our view of categories and categorization. Eleanor Rosch's seminal works on cognitive psychology (1973, 1975) provided a crucial contribution to a theory of categories with the introduction of key notions such as *prototype* and *basic level* which are now widely used in several subdisciplines of linguistics. This revolution also provided the main tenets for a cognitive-functional approach in linguistics, based on the notion that language is embodied and integrated within other human cognitive abilities (Langacker 1987; Lakoff 1987; Croft & Cruse 2004). Several studies in this field have shown how language reflects the way speakers of different languages categorize reality in a culture-specific way; some well-known examples are the studies on color typology (Berlin & Kay 1969) or spatial frames of reference (Levinson 2003). At the same time, advances in cognitive psychology, especially through the works of Lawrence Barsalou (1983, 1991, 2003, 2010, this volume) have introduced an important distinction between natural (or common) categories on the one hand and ad hoc categories on the other. The first roughly correspond to traditional categories, i.e. context independent intuitions, while the latter respond to the need to categorize reality under particular contextual circumstances and for a specific purpose.

Ad hoc categories, which are the main topic of this volume, although the authors do not always use this label, are involved in the creation of reference to situation-specific objects such as "things I need for a one-month vacation to Alaska" or "magazines you can find in a men's barber shop". Therefore, ad hoc categorization is the process through which concepts, regardless of their entrenchment in the speaker's conceptual system, may be constructed in an ad hoc fashion in order to be adapted to local contexts (Mauri, this volume). According to the general theory, such categories differ from "common" or "established" categories in two main

aspects. First of all, ad hoc categories are ephemeral: being strictly dependent on specific situations, they are not permanently stored in human cognition and are often dismissed immediately after their use. This appears to be the clearest reason why languages have "special" strategies to refer to this type of entities (see Mauri 2017 and this volume; Mauri & Sansò 2018, 2020), such as marked prosodic patterns (Matalon this volume), reduplication (Montaut 2009; Alexandrova & Benigni this volume; Kikvidze this volume), associative and similative plurals (Daniel & Moravcsik 2013; Mauri & Sansò in press), list constructions (Selting 2007; Goria & Masini this volume), general extenders (Overstreet 1999, 2005, 2014; Fiorentini & Magni this volume), exemplification (Barotto & Lo Baido this volume). Secondly, ad hoc categories are strongly dependent on human action, or goal derived. In this view, which is extensively discussed in Lawrence Barsalou's chapter, categorization, on the cognitive side, may be thought of as an emergent phenomenon that arises concomitantly with situated action. Therefore, if we think of categorization in terms of process rather than a pre-existing state, we also expect that the linguistic structures that come into play when speakers are in need to verbalize about categorization will somehow reflect this procedural nature. This may be seen in at least two fundamental approaches, that recur across the chapters of this volume: on the one hand, ad hoc categorization is involved in the contextual adaptation of referential expressions, as argued by most of the current pragmatic theories (see Wilson & Carston 2007; Ariel this volume); on the other hand, it may also be thought of as a conversational practice in its own right, whereby speakers cooperate in order to construct referential meaning (see the contributions in Part 2 of this book).

2.   Aims and scope of the volume

Verbalization plays a central role in the process of ad hoc categories construction. Nonetheless, little attention has been given to the linguistic constructions systematically employed to convey this process. (cf. Mauri 2017: 301). The aim of this volume is to provide a relevant contribution to the current debate on categorization and language, employing converging evidence from different languages, mainly based on spoken and experimental data, to account for how categorization is performed and communicated between speakers in interaction. To this end, the linguistic strategies employed to convey the construction of categories will be discussed in detail, providing evidence for an analysis of categorization as the product of linguistic interaction.

More specifically, the book focuses on the attested linguistic strategies and their actual use, rather than on the distinction between category types (e.g. common

categories vs. ad hoc categories, Barsalou 1983), in the belief that the observation of how language is employed in linguistic interaction may reveal the way in which categorization is achieved and communicated, rather than the exact category type being construed. The contributions to the volume address in particular theoretical perspectives on the linguistic construction of categories, interactional approaches to categorization, as well as categorization strategies across languages. Overall, they provide a comprehensive picture of how categorization is achieved through linguistic interaction.

3.   The contributions in this volume

The idea of this volume was first conceived during the workshop *Building Categories in Interaction: Multidisciplinary Approaches to Categorization*, held in Bologna on 19–20 October 2017, in the framework of a broader project on the synchronic and diachronic aspects of categorization in languages. Most of the topics and research questions presented in this book have first emerged during the workshop and have been widely discussed among the participants.

The first part of the volume aims to summarize and expand the current state of the art in research on categorization and is intended, in a way, to be a preliminary reading for any study on this topic. The chapters by Mauri, Barsalou, Goria and Masini and Zaefferer address a systematic discussion of the theoretical and methodological challenges that appear to be recurrent in most of the recent scholarship. In fact, many of the theoretical issues raised here will return at several points throughout the book, in the discussion of language-specific case studies. Caterina Mauri discusses the role that linguistic interaction plays in ad hoc category construction and communication, by looking at naturally occurring data of spoken language. Ad hoc categorization is analyzed taking the perspective of so-called *languaging*, showing that categorization is frequently instrumental to intersubjective aims, such as mutual agreement, negotiation, and the general management of the speakers' positioning. The paper describes the incrementality of ad hoc categorization in interaction along two dimensions, namely, the identification of the category borders and the progressive anchoring of the category to the interlocutors' experience. The chapter by Lawrence Barsalou, couched within the framework of cognitive psycholinguistics, focuses on the close relationship between categorization and action, showing how categories are inherently bound to human activities and may have various degrees of conventionalization. As a consequence, the linguistic resources that in single languages are recruited for the task of performing categorization will reflect in their structure the procedural nature of categories.

This aspect is investigated in detail by Goria and Masini: the authors show that categorizing constructions typically present form-function asymmetries that are easily handled from a Construction Grammatical perspective. Furthermore, an important aspect dealt with in the same chapter concerns the modeling of similarities across different constructions related to the activity of categorization. Finally, Dietmar Zaefferer investigates the diffusion and conventionalization of category labels within a community: by discussing experimental data, Zaefferer offers a new insight on the dynamics of survival vs abandonment of the terms used for newly constructed categories.

The second part of the volume is dedicated to contributions that share an interest in categorization occurring in spoken interaction. While the authors focus on different languages and adopt different methodologies, all the five chapters by Matalon, Sammarco, Paul, Barotto and Lo Baido and Arcodia are unanimous in considering categorization as an outcome of spoken interaction, that is achieved by recruiting specific linguistic resources that may be either language-specific or even shared cross-linguistically. Thus, Nadav Matalon analyzes in detail one particular prosodic pattern, referred to as "the camel humps", which occurs in spoken Hebrew conversation as a contextualization cue for open lists, one of the main strategies through which categories are created in interaction; similarly, Carmela Sammarco takes into account the use of verbless constructions in Italian and French as an interactional resource to provide a label for contextually relevant categories, often with a discourse-organization function. In the chapter by Christine Paul, the theoretical tools of Conversation Analysis are employed to analyze the phenomenon of semantic undeterminacy: underspecified utterances of various formats may be interpreted by the participants of a conversation as locally meaningful and be exploited as resources for the negotiation of contextually bound categories. Barotto and Lo Baido concentrate on how the construction of categories is cooperatively carried out by the participants in dialogue: they focus on the use of exemplification in the construction of categories, which gives rise to emerging regular structures in the form of adjacency triplets. Finally, Giorgio Arcodia describes the various ways in which relative clauses are used in spoken Mandarin Chinese to construct ad hoc categories, also giving rise to idiosyncratic utterance formats that are not observed in written Chinese.

The third part is dedicated to case-studies focusing on language-specific constructions that have the function of expanding the denotation of particular referential expression, allowing for the creation of a category. Ilaria Fiorentini and Elisabetta Magni provide new insights on how categorizing constructions come into being, by investigating the diachronic development of the Italian general extender *eccetera*, from the Latin referential expression *et cetera*. Anna Alexandrova and

Valentina Benigni describe the behavior of one particular type of Russian compounds: rhyming and imitative co-compounds. These typically occur in spoken discourse and express a number of pragmatic functions that are relevant in an account of categorization. A similar type of non-prototypical reduplication is also found in Georgian, as described in Zaal Kikvidze's chapter: the author provides a three-type account of what he refers to as *echo-word formation* reflecting on their role in the construction of ad hoc categories. Wiltrud Mihatsch investigates the functional development of taxonomic nouns meaning *type* or *kind*, which she considers a case of pragmaticalization, as strategies to convey reference to an ad hoc category.

The two final chapters, by Corona and Pietrandrea and Ariel, overturn the perspective held up to this point, convincingly showing that it is not only the case that strictly locally relevant categories of meaning are constructed ad hoc. Categorization goes far beyond this purpose, to the point that even the so-called "taxonomic" or "established" categories may be treated and constructed in an ad-hoc way when they undergo pragmatic adaptation. Luisa Corona and Paola Pietrandrea address the ad hoc construction and the co-construction of the category of manner in discourse. Manner is indeed traditionally approached as a basic cognitive prime expressing the way in which an action is performed by means of a limited set of linguistic strategies. However, the authors argue that manner can be regarded, when analysed through a corpus-driven methodology, as an interactional, gradable category co-constructed by participants in discourse, used by speakers not only to describe reality but also to reach a shared perception of reality.

Finally, Mira Ariel takes into account number categories, which are taken to be an example of discrete entities. Her chapter must be thought of as an afterword to the whole volume: as we move away from the topic of ad hoc categorization, we close our investigation by asking how much common categories resemble ad hoc ones. Ariel's study reminds us that categorization is not all about linguistic resources being put to work in order to express highly context-specific meanings: all types of meaning are in a way constructed ad hoc, every time they need to be pragmatically adjusted to the context.

Funding

References

Barsalou, Lawrence W. 1983. Ad hoc categories. *Memory and Cognition* 11(3): 211–227.
 https://doi.org/10.3758/BF03196968

Barsalou, Lawrence W. 1991. Deriving categories to achieve goals. In *The Psychology of Learning and Motivation: Advances in Research and Theory*, Gordon H. Bower (ed.), 1–64. San Diego CA: Academic Press. (Reprinted in *Goal-driven Learning*, Ashwin Ram & David Leake (eds), 121–176. 1995. Cambridge MA: The MIT Press/Bradford Books).

Barsalou, Lawrence W. 2003. Situated simulation in the human conceptual system. In *Language and Cognitive Processes* 18: 513–562. (Reprinted in *Conceptual Representation*, Helen Moss & James Hampton (eds), 513–566. Hove: Psychology Press).

Barsalou, Lawrence W. 2010. Ad hoc categories. In *The Cambridge Encyclopedia of the Language Sciences*, Patrick C. Hogan (ed.), 87–88. Cambridge: CUP.

Berlin, Brent & Kay, Paul. 1969. *Basic Color Terms: Their Universality and Evolution* [The David Human Series Philosophy and Cognitive Science Reissues 19]. Berkeley CA: University of California Press.

Croft, William & Cruse, Alan D. 2004. *Cognitive Linguistics*. Cambridge: CUP.
 https://doi.org/10.1017/CBO9780511803864

Daniel, Michael & Moravcsik, Edith. 2013. The associative plural. In *The World Atlas of Language Structures*, Martin Haspelmath, Matthew Dryer, David Gil, Bernard Comrie (eds), Chapter 36. Munich: Max Planck Digital Library. <https://wals.info/chapter/36> (22 March 2021).

Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago IL: The University of Chicago Press.
 https://doi.org/10.7208/chicago/9780226471013.001.0001

Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar, Vol. 1: Theoretical Prerequisites*. Stanford CA: Stanford University Press.

Levinson, Stephen C. 2003. *Space in Language and Cognition: Explorations in Cognitive Diversity*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511613609

Mauri, Caterina. 2017. Building and interpreting ad hoc categories: A linguistic analysis. In *Formal Models in the Study of Language*, Joanna Blochowiak, Cristina Grisot, Stephanie Durrleman, & Christopher Laenzlinger (eds), 297–326. Berlin: Springer.
 https://doi.org/10.1007/978-3-319-48832-5_16

Mauri, Caterina & Sansò, Andrea. 2018. Linguistic strategies for ad hoc categorization: Theoretical assessment and cross-linguistic variation. *Folia Linguistica* 52(s39–1): 1–35.
 https://doi.org/10.1515/flih-2018-0001

Mauri, Caterina & Sansò Andrea (eds). 2020. Ad Hoc Categorization and Language: The Construction of Categories in Discourse. Special Issue of *Languages Sciences* 81.

Mauri, Caterina & Sansò, Andrea. In press. Heterogeneous plurals: A diachronic typology of associative and similative plurals. To appear in *Linguistic Typology*.
 https://doi.org/10.1515/lingty-2021-2072

Montaut, Annie. 2009. Reduplication and 'echo words' in Hindi/Urdu. In *Annual Review of South Asian Languages and Linguistics*, Rajendra Singh (ed.), 21–91. Berlin: De Gruyter.

Overstreet, Maryann. 1999. *Whales, Candlelight and Stuff Like That: General Extenders in English Discourse*. Oxford: OUP.

Overstreet, Maryann. 2005. And stuff und so: Investigating pragmatic expressions in English and German. *Journal of Pragmatics* 37: 1845–1864. https://doi.org/10.1016/j.pragma.2005.02.015

Overstreet, Maryann. 2014. The role of pragmatic function in the grammaticalization of English general extenders. *Pragmatics* 24: 105–129. https://doi.org/10.1075/prag.24.1.05ove

Rosch, Eleanor H. 1973. Natural categories. *Cognitive Psychology* 4(3): 328–350.
https://doi.org/10.1016/0010-0285(73)90017-0

Rosch, Eleanor H. 1975. The nature of mental codes for color categories. *Journal of Experimental Psychology: Human Perception and Performance* 1(4): 303–322.

Selting, Margret. 2007. Lists as embedded structures and the prosody of list construction as an interactional resource. *Journal of Pragmatics* 39(3): 483–526.
https://doi.org/10.1016/j.pragma.2006.07.008

Wilson, Deirdre & Carston, Robyn. 2007. A unitary approach to lexical pragmatics: Relevance, inference and ad hoc concepts. In *Pragmatics*, Noel Burton-Roberts (ed.), 230–259. Houndmills: Palgrave. https://doi.org/10.1057/978-1-349-73908-0_12

CHAPTER 2

Ad hoc categorization in linguistic interaction

Caterina Mauri
University of Bologna

The aim of this paper is to describe and explain the role that linguistic interaction plays in category construction and communication, by looking at naturally occurring data of spoken language. First, it will be argued that there is a way of building categories that is inherently interactional and indexical, namely ad hoc categorization. Ad hoc categorization will be defined as a bottom-up exemplar driven process, that is dependent on context for both its construction and its interpretation, and crucially relies on non-exhaustivity and exemplification. After a brief overview of the linguistic strategies that may encode ad hoc categorization, we will concentrate on linguistic interaction, taking the perspective of so-called languaging. It will be shown that categorization is frequently instrumental to intersubjective aims, such as mutual agreement, negotiation, and the general management of the speakers' reciprocal positioning. In turn, it is collaboration between the interlocutors that allows to fine-tune categorization and achieve mutual understanding. Finally, we will focus on the incrementality of ad hoc categorization in interaction along two dimensions, namely, the identification of the category borders and the progressive anchoring of the category to the interlocutors' experience.

Keywords: ad hoc categorization, exemplification, non-exhaustive connectives, general extenders, languaging, indexicality, linguistic interaction

1. Introduction

The aim of this paper is to discuss the role that linguistic interaction plays in category construction and how category construction is achieved in linguistic interaction, by looking at naturally occurring data. In most linguistic approaches, categorization is seen as a primarily cognitive phenomenon, and as a consequence, linguistic structures are regarded as reflecting more or less transparently this pre-linguistic organization of experience. However, when we speak, we actively construct categories

and we typically do it in a cooperative and incremental way with our interlocutors. Therefore, once we take the perspective of speakers, categorization also becomes a crucially linguistic and interactional phenomenon.

We will start by describing the indexicality and context dependency that is inherent to the communication of categories, addressing the procedural nature of a particular type of categorization, that is, ad hoc categorization (Section 2.1). After a brief overview of the linguistic strategies usually employed to convey ad hoc categorization, in Section 3 we will focus on the interactional dimension underlying category construction in discourse, arguing that categories are construed by interlocutors through cooperation and negotiation, with interaction itself being at the same time a path and a goal of categorization. It will be shown indeed that the achievement of mutual agreement is in certain cases the objective to which cooperative category construction is aimed to.

The interactional dimension of ad hoc categorization will become even more evident in Section 4, where it will be argued that the identification of the category borders (Sections 4.1 and 4.2) and the anchoring of the category to the interlocutors' experience (Section 4.3) are achieved incrementally in discourse. Some concluding remarks will then follow in Section 5.

2.   The indexicality of categorization

2.1   Ad hoc categorization and the role of context

In the light of psychological evidence provided since Eleanor Rosch's studies on prototypes (1973, 1975), it is nowadays well established that categorization heavily relies on context. Barsalou (1983, 1991, 2003, 2010) identifies a specific category type that is not only dependent on context for its interpretation, but also for its very conception and construction, namely what he labels *ad hoc* category. Ad hoc categories are indeed goal-driven abstractions, which respond to the need to categorize the world under particular contextual circumstances. They are typically expressed by complex linguistic structures such as "clothes you can buy in a supermarket", and are created on the fly for specific communicative purposes, to be dismissed once they are no longer useful.

Lakoff's theory of categorization (Lakoff 1987), and typological research, such as the study on colors by Berlin and Kay (1969) or the study on spatial relations by Levinson (2003), provide further evidence for a theory of categorization that assigns great importance to contextual, cultural and linguistic variation. Croft and Cruse (2004) propose to consider every category as construed *on-line*, in a

context-dependent way, according to the speakers needs and expectations, leading to a theory in which all categories are to be analyzed as construed in an ad hoc way.

Much of the psychological and cognitive debate on categorization, however, is aimed at understanding how a specific category is interpreted, rather than observing how the process of category construction is achieved and conveyed in linguistic communication. Yet, linguistic data offer an empirical ground where we can observe the choices that speakers make and the strategies they employ to verbalize the process of category construction, monitoring the degree to which they rely on context and shared knowledge. Speakers may indeed decide to label a category through lexical or syntactic means (top-down), they can express the process itself of category construction, through exemplification and listing (bottom-up), or, as is frequently the case, they may opt for both options in alternation.

According to Mauri and Sansò (2018), category construction through reference to individual exemplars allows speakers to refer to abstract concepts by keeping their communication at the level of concrete individuals, employed as triggers of exemplar-driven categorization. They propose to call *ad hoc categorization* the verbalization of a bottom-up, goal-driven, context-dependent categorization process, characterized by the use of one or more examples to refer to a higher-level category. Evidence for both the universality and cross-linguistic variation in the verbalization of bottom-up categorization can be found in typological studies and corpus-based research on specific languages (cf. Ariel & Mauri 2018; Mauri & Sansò 2018, 2020; Barotto & Mauri 2018; Mauri, Goria & Fiorentini 2019; Arcodia & Mauri 2020; Fiorentini & Miola 2020; Barotto 2021, among others).

The semantic core that characterizes linguistic strategies employed to convey ad hoc categorization comprises:

(1) a. one or more *explicit exemplars* of the category
     b. some additional *implicit members*, associated with the explicit exemplars by virtue of a common property or frame that is relevant to the context
     c. a *superordinate category*, which includes both explicit exemplars and implicit further category members.   (cf. Mauri 2017)

Exemplars are processed as arrows pointing towards the higher-level class (i.e. (1c)), which is larger than the set of mentioned exemplars (i.e. (1a)) and includes further potential additions (i.e. (1b)). The crucial step towards the identification of the intended category is the abstraction of the relevant property or frame, which depends on context and on the knowledge shared by the interlocutors, making ad hoc categorization inherently cooperative. Let us consider Example (2):

(2)   A: […] *German prisoners of war used to make toys. And although they weren't allowed to sell them for money*, [*people could give them objects, could give them* [*things like coffee*]$_b$ *and things like that*]$_a$
      B: % *Yes*
      A: % *in exchange*   (BNC Corpus, D8Y)[1]

Example (2) shows two speakers interacting. Speaker A is explaining that German prisoners of war used to build toys and they could not truly sell them, because they could not receive money, but they could receive other goods in exchange. In order to communicate what they could receive, Speaker A starts a process of ad hoc categorization, by exemplifying what people could give them, namely *objects, things like coffee* and *things like that*. By using the expression *and things like that* s/he indicates that the set is non-exhaustive and further potential additions are to be made, thus inviting Speaker B to actively enrich the list and join speaker A in building the relevant category, for which no lexical entry can be easily found. Speaker B promptly accepts the invitation and interrupts the interlocutor, giving the positive feedback *Yes*, to let Speaker A know that ad hoc categorization was successful: the relevant frame was accessed and the category [goods that could be given in exchange in war period]$_a$ was achieved. Interestingly, along the process of category construction, we observe a nested ad hoc categorization process, built through *things like coffee*, once again by introducing a representative exemplar of a larger set. In this case, the relevant property shared by *coffee* and the additional implicit set members is more specific and can be abstracted by opposition with *objects* above, yielding the category [drinks or food that prisoners would like to receive in war period]$_b$.

   In both ad hoc categorization processes, it is the expressions conveying non-exhaustivity (i.e. the exemplifying strategy *things like* and the general extender *and things like that*) that trigger the interpretation of the mentioned items (*objects, coffee*) as exemplars of a larger, higher-level category (cf. Mauri, Goria & Fiorentini 2019 on non-exhaustive listing). The hearer then abstracts the category from context and from the mentioned exemplars, by identifying the property P that is relevant in the specific frame (war period), e.g. 'something difficult to find for prisoners, useful and available in war period'. The identification of the context-specific property P is what allows to discriminate between plausible and implausible category members: for instance, clothes could be easily identified as likely members of [goods that could be given in exchange in war period]$_a$ but a computer is not, despite being an

---

1.   Data cited herein has been extracted from the British National Corpus Online service, managed by Oxford University Computing Services on behalf of the BNC Consortium. All rights in the texts cited are reserved. The symbol % is here employed to indicate an overlap between different turns.

object; tea is a likely member of [drinks or food that german prisoners would like to receive in a war period]$_b$ but Japanese sake is not, despite being a drink.

Access to a shared context is thus what allows to abstract the relevant category being built, and it is also what makes the process of category construction a product of the ongoing interaction, as witnessed by the feedback that Speaker B provides to Speaker A in (2): it is thanks to some shared background that the interlocutors are able to interpret the non-exhaustivity conveyed by *things like* and *things like that*, thus safely drawing on extra-linguistic knowledge to communicate categories. Mauri (2017) and Barotto and Mauri (2018) propose to analyze ad hoc categorization as an indexical phenomenon, whereby the identification of the context-specific value of the property P (shared by the mentioned exemplar(s) and the additional implicit ones) is a process of saturation (cf. Recanati 2004), namely a process whereby a given variable receives one or more values depending on context. Saturation is typically described for classical deictic markers, such as *this*, where reference is made to some entity whose identity can only be retrieved by access to context (cf. Anderson & Keenan 1985). In this case, the saturation of P is necessary to abstract the relevant category: if the identification of the value of P within the relevant frame fails, it is impossible to interpret ad hoc categorization (cf. Mauri & Sansò 2018 for a detailed discussion).

Non-exhaustive reference is necessary to trigger ad hoc categorization, because it introduces further potential additions that have to be inferable from the explicitly mentioned items, in such a way that their identification is only possible once the larger category including both explicit and implicit set members is correctly abstracted. However, the exact identity of the additional members may (and is actually likely to) remain unspecified even for the speaker who intentionally employs a non-exhaustive linguistic expression. Speaker A in (2) does not necessarily have in mind further items beyond objects and coffee, although s/he is aware that the set is not restricted to objects and coffee. What is indeed necessary for the higher-level category to be processed is that the additional category members be identifi*able*. Their identifiability is in turn subordinated to the *identification* of the specific value of P: once the value of the context-relevant Property P has been identified, the hearer is in the condition to discriminate between possible and impossible additional members of the higher-level category (cf. Mauri & Sansò 2018).

2.2   The linguistic expression of ad hoc categorization

Based on what has been argued in the previous section, the linguistic strategies employed for ad hoc categorization can be analyzed as inherently indexical, referring to one or more exemplars and to some further unspecified items characterized by a Property P, which needs to be saturated by accessing context and a shared frame.

The various prosodic, morphological, or syntactic resources, that encode reference to additional elements, work as categorization triggers. That is, by signaling non-exhaustivity, they invite to infer the property P defining the whole category.

In addition to a categorization trigger, these strategies are also characterized by the presence of some overt category member(s), processed as the starting point for abstraction, i.e. as exemplar(s). Indeed, we may say that ad hoc categorization employs exemplification as a road to category abstraction (cf. Section 4). Although we observe some structural and functional variation across languages, all the linguistic constructions encoding ad hoc categorization can be analyzed as being composed by one or more exemplar(s) and some non-exhaustivity marker, working as categorization trigger. According to Mauri and Sansò (2018), the more morphological the strategy, the more the exemplar is likely to be unique and to play a pivotal role in the category construction; the more syntactic the strategy, the more the *online* process of set construction is mirrored in the speaker's listing of exemplars, along what looks like a search for the correct category delimitation.

Syntactic strategies are highly analytical, in some cases even compositional, and may typically be used with phrases and clauses, allowing for as many exemplars as the speaker needs. They include lists and exemplifying constructions. In lists, non-exhaustivity can be expressed by purely prosodic patterns or by explicit elements: if the non-exhaustive element is located at the end of a list, it will be called *general extender* (Overstreet 1999; Cheshire 2007; cf. *or something like that* in (3)); if it is employed to link the list items, it behaves as a *non-exhaustive connective* (Barotto & Mauri forthcoming; Fiorentini & Miola 2020; cf. *-a* in (4)). As pointed out by Kuperschmidt (2018) and Ariel & Mauri (2018), also disjunctive connectives are frequently employed to link exemplars of some higher-level categories.

(3) […] *her mum always cooks a meal in the evening so I, I do something like toasted cheese sandwiches or beans on toast **or something like that** at lunch time* […]
(BNC Corpus)

(4) Mandarin (Sino-Tibetan, Chinese; Zhang 2008: 137; PF = perfect)
   a. *Shu-**a**,   baozhi-**a**,   bai-man-le zhengge shujia.*
       book-and newspaper-and put-full-PF whole   bookshelf
       'Books and newspapers, **among other things**, occupied the whole bookshelf.'
   b. *Tamen tiao-**a**   chang-**a**, huanqing shengli.*
       they   dance-and sing-and celebrate victory
       'They sang, danced, **among other activities**, to celebrate the victory.'
   c. *Yin-(*a) yang-(*a) duili.*
       yin-and yang-and opposite
       'Yin and yang are opposites.'

Exemplification, defined as the process whereby one or more elements are to be interpreted as representative of a broader class (cf. Lo Baido 2018; Barotto 2021),

lies at the core of ad hoc categorization itself, as widely argued in Section 2.1. In this respect, every linguistic strategy encoding ad hoc categorization makes use of exemplification as a bottom-up way towards category construction. However, it is possible to identify a more restricted set of linguistic strategies, such as *for example, let's say, such as* (cf. (5)), etc., for which exemplification is the core meaning, that is, constructions that directly encode the fact that the linguist elements falling under their scope have to be interpreted as exemplars of a larger set. We will restrict the term 'exemplifying construction' to these cases.

(5) *You know, yes, skin tone does often play a role in **things such as** entertainment, politics and business* […]

(COCA Corpus, SPOK: NPR_NewsNotes, Davies 2008)

A very common strategy attested in the world's languages to encode ad hoc categorization is echo-reduplication (Inkelas 2014: 169–171; Stolz 2008: 115ff.), as shown in (6) from Kannada, where a lexical base is reduplicated by replacing the first consonant and vowel of the noun with the sequence *gi-* or *gi:-*:

(6) Kannada (Dravidian, Southern Dravidian; Lidz 2000: 148–149;
1SG = 1st person singular, ACC = accusative, NOM = nominative,
PROH = prohibitive, PST = past, RED = reduplication)
a. *pustaka → pustaka-gistaka*
book   book-RED
'book'   'books and related stuff'
b. *ooda → ooda-giida beeDa*
run   run-RED   PROH
'run'   'Don't run or do related activities.'
c. *nannu baagil-annu much-id-e   giigilannun muchide*
I-NOM door-ACC   close-PST-1SG RED
anta heeLa-beeDa
that   say-PROH
'Don't say that I closed the door or did related activities.'

Ad hoc categorization may also be expressed by means of morphological strategies, such as similative plurals (Daniel & Moravcsik 2013; Mauri & Sansò forthcoming), derivational strategies (collectives, aggregates, Magni 2018) and even compounds (co-compounds, Wälchli 2005: 141ff.). Similative plurals have been described by Mauri & Sansò (forthcoming) as a type of heterogeneous plural, denoting a category of objects sharing similar features with a given exemplar, as in (7) from Manambu:

(7) Manambu (Sepik, Middle Sepik; Aikhenvald 2008: 509;
SIM.PL = similative plural)
bal məwi
pig SIM.PL
'pigs and things like that'

While similative plurals are employed in the same contexts where we find general extenders, that is in open listing, derived collectives and compounds are typically more conventionalized ways to denote ad hoc categories, showing a lower degree of context-dependency. Arcodia and Mauri (2020) analyze what they propose to label *exemplar-based* compounds, i.e. juxtapositions of two exemplars used as strategies to encode the superordinate category encompassing the two exemplars plus other entities connected to them, as in (8):

(8)   Mandarin (Sino-Tibetan, Chinese; Arcodia & Mauri 2020)
    a.   *dāo-qiāng*
         sword-spear
         'weapons'
    b.   *qín-shòu*
         bird-quadruped
         'birds and beasts'
    c.   *gān-gē*
         shield-dagger/axe
         'weapons, war'

All in all, despite the great variation attested, we can safely argue that every language shows indexical strategies to communicate ad hoc categorization, i.e. bottom-up, exemplar-driven, context-dependent category construction. In the next Section, we will focus on the use of such strategies in discourse, based on data of Spoken Italian, aiming to analyze how speakers, involved in a dialogic interaction, cooperate to build and identify categories in discourse.

3.   The languag*ing* perspective: Category construction in interaction

Linguistic interaction has received increasing attention in the last decades, as the place where grammar emerges out of repeated, collaborative use. A term that has had some success in the literature is *languaging*, in opposition to the term *language*, to emphasize the online dimension of the communication process rather than the static dimension of the communication product or tool (cf. Becker 1988; Steffensen 2009, 2015; Thibault 2017; Raimondi 2019, among others). Becker (1991) suggests that there is no such thing as language, but the only thing we can observe is the continuous activity of human communication, which coincides with what he calls 'languaging'. Swain and Watanabe (2013) describe languaging as the "process of making meaning and shaping knowledge and experience through language" (Swain 2006: 98), and argue that the use of the progressive verb *languaging*, instead of the noun *language*, forces a conception of language as a process rather than a reified entity.

Languaging thus refers to the activity performed in speech, which is an ongoing process constantly evolving and developing, thanks to the evolving relation between interacting speakers. The crucial role of human collaboration in languaging is underlined by Raimondi (2019: 19–20), who argues that "the notion of languaging activity is inherently dialogical and radically relational". Building on the biological theory of Maturana (1983) and on the theories of Cowley (2007) and Thibault (2017), which insist on the embodied nature of the languaging activity, Raimondi focuses on dialogicity as being a central aspect of linguistic communication and, in general, of collaborative human activities. According to Raimondi (2019: 24; cf. also Linell 2009), communicating human beings are inherently cooperative and interdependent, therefore each event of individual speech occurs within a discursive framework of dialogue, making dialogicity a core feature not only of human languaging, but also, more in general, of human cooperative interaction

Research developed within conversation analysis and grammaticalization theories reached similar conclusions (cf. Bybee & Hopper 2001; Traugott 2003; Bybee 2015; Traugott & Trousdale 2010). In particular, the role played by dialogicity and use in shaping language has received great attention in studies on so-called constructionalization (Traugott & Trousdale 2013), which focus on the emergence of grammar from recurrent discourse patterns. Other recent approaches have highlighted the online aspects of grammar by taking the consequences of the linearity of speaking in time for syntactic organization into serious consideration (Auer 2009; Auer & Pfänder 2011), by focusing on the specificities of dialogic syntax (Linell 2009; Du Bois 2014), or by identifying the linguistic correlates of spoken modality at all levels of grammar (Voghera 2017).

It is against the background of an approach to linguistic data based on the observation of languaging, that we now aim to analyze ad hoc categorization, considering data of spoken Italian from the VoLIP Corpus and the KIParla Corpus.[2] Linguistic interaction, or *languaging*, is indeed at the same time container and content of categorization processes, with speakers engaging in cooperation and negotiation aimed at the ongoing and everchanging process of reciprocal fine-tuning. A great part of this mutual tuning is determined by reaching a common category construction, exploiting all the tools that discourse provides to manage this online process: if two speakers agree on how a category is to be construed, they agree on the reference, or set of referents, corresponding to the category, and this basically means that they agree on what they are talking about.

---

2.   The KIParla corpus is publicly available at <www.kiparla.it>. It includes spoken data collected in Turin and Bologna in the years 2016–2019 (see Mauri et al. 2019). The VoLIP Corpus is publicly available at <http://www.parlaritaliano.it/index.php/it/volip> and includes data collected in Naples, Rome, Milan, Florence in the early '90s (see Voghera et al. 2014).

Let us consider Example (9) from the KIParla corpus of spoken Italian:

(9)

```
1 A: io ho paura che questa vada a cercare parecchio il pelo nell'uovo
     I'm afraid that this (professor) will split hairs a lot
2 B: dici?
     you think so?
3 A: eh questo mi fa paura
     eh this is what I fear
     cioè sai quelle precise che vogliono sapere tutto cioè i il boccaccio
     anche (.)
     I mean you know those nit-picking who want to know everything I
     mean Boccaccio even,
5    anche il colore delle mutande voglio di-
     even the underwear color I mean
6 B: ah ho capito
     oh I see                                  (KIParla Corpus BOA3001)
```

In Example (9) two students are interacting, A is trying to guide B to identify the type of professor she is talking about, who causes A's feeling of fear. To do this, A starts by naming the aspect that she is afraid of (*the professor will split hairs*, line 1), and to reply to B's doubt (*you think so?*, line 2) she tries to build and communicate the category of professors to which the one at issue belongs.

The process of category construction starts with a reformulation (introduced by *cioè* 'I mean', line 4) and with the search for the interlocutor's feedback (*sai* 'you know', line 4). It then proceeds by labeling the category through the complex relative clause *quelle precise che vogliono sapere tutto* 'those nit-picking (professors) who want to know everything', (line 4). Yet, Speaker A feels that the label is not informative enough, probably because the universal quantifier *everything* is too inclusive and generic, and therefore provides another reformulation (introduced again by *cioè* 'I mean', line 4). This time two highly specific exemplars of what the professor could ask are listed, namely *il boccaccio*[3] and *il colore delle mutande* 'the color of (Boccaccio's) underwear', whereby the second example is a specification of the first one. The choice of these two examples is highly meaningful for B, to the point that he ultimately provides the sought feedback *ho capito* 'I see' (line 6), which confirms mutual understanding. Usually, examples are chosen by virtue of their being prototypical and representative of the category, but here it is clear that A's intention is different: the color of Boccaccio's underwear is an extremely uncommon exemplar, aimed at pushing the borders of the category 'everything' so far as to include the least predictable case, namely non-relevant details that are impossible for a student to learn.

---

3. Giovanni Boccaccio (1313–1375) is one of the most important Italian writers of the 14th century, author of the collection of novellas known as the Decameron.

What we observe in (9) is thus a wave pattern, whereby the speaker first labels the category, which is then reformulated and exemplified in order to make it more accessible for the interlocutor, until he is able to abstract and construe it in the right way. The ad hoc categorization process in (9) is not only highly dependent on context but is also rooted into and led by the dialogical and cooperative interaction of languaging. The interlocutor's feedback is indeed the goal that drives the categorization process, which was triggered by the manifestation of some doubt (*dici?* 'You think so?', line 2), that is, by the risk of potential misunderstanding. Beyond the cognitive dimension of abstraction towards the identification of the category, the languaging perspective indeed highlights the cooperative dimension of conversation, in which ad hoc categorization is instrumental to building shared knowledge and mutual agreement.

Intersubjective moves like the search for agreement and the need for a feedback play a crucial role in categorization processes (cf. Barotto & Lo Baido this volume): the cognitive dimension of category construction cannot indeed be separated from the pragmatic dimension of category verbalization, which is guided by illocutionary aims and discourse needs relating to the speakers' reciprocal positioning.

The interactional dimension of ad hoc categorization clearly emerges from the frequent co-occurrence of ad hoc categorization strategies with expressions that rely on the speakers' relation, such as interactional discourse markers (e.g. *sai* 'you know' and *dici* 'you say?' in (9)), feedback (e.g. *yes* in (1) *ho capito* 'I see' in (9)), or explicit appeals to mutual understanding (e.g. *mhmm* in (16)). When a category is built through a cooperative exchange by means of an incremental online process, once its construction is felt to be felicitous, i.e. when speakers think that a shared category has been reached, this is typically overtly acknowledged. Let us consider (10):

(10)
```
1 A: eh [è vicinissima al centro però i prezzi delle cose tipo supermercati
        cibo eccetera, eh [it's very close to the center but the prices
        of things like food supermarkets etc.,
2      non è esagerato
        is not exaggerated
3      è allegra son tutti giovani,
        it's cheerful everybody is young
4      comodissima coi mezzi]ₐ
        it's very comfortable with transports]ₐ
5      meravigliosa
        wonderful]ₐ
6 B: [ti piace il fatto che sia una zona viva dove c'è movimento]ᵦ
        [you like the fact that it's a lively area where there's movement]ᵦ
7 A: sì io adoro Torino per questo motivo […]
        yes I love Turin for this reason […]          (KIParla Corpus TOD2003)
```

Example (10) shows a conversation in which speaker A enumerates the reasons why a specific area of Turin is wonderful, by means of an exemplification list (list (a), lines 1–4): it's very close to the center, it's cheerful, everybody is young, it's very comfortable with transports. Then, exemplification leaves room to the speaker's evaluation, by means of the superlative *meravigliosa* 'wonderful'. Speaker B abstracts over the examples provided by A, proposing a general formulation that catches why A loves that area, namely 'the fact that it's a lively area where there's movement' ((b), line 6): we observe here the cooperative attitude of B, who aims to categorize over the examples, implicitly asking for a feedback on the abstraction just made. The positive feedback is provided on line 7, where Speaker A confirms the categorization ('yes, I love Turin for this reason').

Barotto & Lo Baido (this volume) argue that exemplification itself can be employed in conversation as a positive feedback, to prove mutual understanding after a process of ad hoc categorization, as in (11). In this case, instead of simply using a backchannel, Speaker B acknowledges that the category has been felicitously construed by proposing a further example as a proof ('optics for the eyes' on line 4).

(11)
```
1 A: dove ci sono gli ambulatori
     where there are clinics
2 B: sì
     yes
3 A: per fare le visite
     to do medical examinations
4 B: sì sì ho capito per esempio come l'ottica per gli occhi
     yes yes I understand for example like optics for the eyes
5 A: sì […]
     yes                                          [LIP Corpus, RC8]
```

Exemplification is here employed to fulfil an interactional need, namely, to communicate the respective alignment and agreement, as becomes evident from the repetition of *sì* 'yes'. Speaker B indeed explicitly says 'I understand' just before providing the example as evidence for this statement.

The cooperative process can easily turn into a negotiation, if the online process of ad hoc categorization takes different directions for the speakers involved in a conversational exchange. Example (12) provides another case in point:

(12)
```
1 A: in cucina ricotta va bene [tipo negli agnolotti, nelle torte salate]ₐ
     ricotta is fine in cooking, [like in agnolotti, in savory pies]ₐ
2    però rigorosamente cotta
     but strictly cooked
```

```
3     ri=cotta
      re-cooked[4]
4 B:  la ricotta è ricotta tu dici […]
      ricotta is re-cooked, you mean
5 B:  e [tipo nei cannoli siciliani]ᵦ?
      and [like in Sicilian cannoli]ᵦ?
6 A:  io non li mangio
      I don't eat them                        [KIParla Corpus TOA3004]
```

In Example (12) Speaker A previously said that she does not like cheese, except for mozzarella and ricotta cheese. Ricotta cheese, however, is argued to be ok only *in cucina* 'in cooking (lit. 'in kitchen', line 1): this category is felt to require further specification to be understood, so Speaker A provides two examples, representative of what the ad hoc category 'ricotta cheese in cooking' refers to ((a) line 1: *negli agnolotti* 'in agnolotti', *nelle torte salate* 'in savory pies'). In line 2, the crucial concept is focused and reformulated: the cheese has to be cooked. At this point, Speaker B feels that he got the point, but the category boundaries may be negotiated, therefore he starts to actively contribute to the categorization process, by proposing an additional element to the exemplification chain that could constitute an exception ((b): *nei cannoli siciliani* 'in Sicilian cannoli', line 5). Sicilian cannoli have indeed raw cheese, but are very popular, so Speaker B employs them here to challenge the actual category borders, i.e. to verify to what extent the category is built around the property of having cooked ricotta. Yet, this example leads to a set that is different from the one that Speaker A had in mind, so an immediate contrastive reaction is provided in line 6, where A says 'I don't like them'. In other words, an intent of negotiation in the identification of the actual category delimitation fails and triggers a reaction that clearly anchors the category borders to Speaker's A tastes, that is, his subjective and individual experience (*I don't eat them*, cf. discussion in Barotto and Lo Baido, this volume).[5]

As argued by Clark and Wilkes-Gibbs (1986), within a conversation speaker and addressee explicitly signal their mutual understanding and alignment, especially concerning the identification of reference and categories. Ad hoc categorization in discourse can thus be triggered not only by the cognitive need to build an abstract concept, but also by the interactional need to cooperate, or negotiate, in the speech event, increasing social cohesion and common ground.

---

4. In Italian *ricotta* literally means re-cooked.

5. I would like to thank Alessandra Barotto for the hintful discussion on this example.

4. Incremental ad hoc categorization: Zooming in and outside categories in discourse

As Examples (9)–(12) have shown, category construction is the result of a cooperative process in which the speakers involved share a common goal, namely mutual understanding. To reach this goal they pursue category co-construction through successive attempts of (re)formulation (cf. (10)) and exemplification (cf. (11)–(12)), until they reach some agreement. In the process of ad hoc categorization, interlocutors are thus able to anchor general statements and abstract concepts to their own experience (cf. Barotto 2021), and this occurs incrementally through the ongoing conversational exchange. In this Section, three phenomena will be discussed that are revealing of how ad hoc categorization is construed incrementally along two dimensions, namely, the identification of the category borders (Sections 4.1 and 4.2) and the progressive anchoring of the category to the interlocutors' experience (Section 4.3). As will become clear from the discussion, the incremental and gradual focusing of the category members and of the category boundaries occurs both on the speaker's side and on hearer's side, because they are mutually and contemporarily involved in ad hoc categorization.

4.1   Identifying the category core and pushing the category borders

When two speakers involved in languaging are cooperating to construe the same category, the choice of the examples is crucial to trigger the right inferential path (cf. Mauri & Sansò 2018). It is indeed the example(s) mentioned in discourse that activates the search for a specific contextually relevant property that ultimately characterizes the frame or category to be abstracted (cf. Section 2.1). Let us consider Example (13), where a category is conveyed in line 6:

(13)
```
1 A: ma infatti scusate, maretta quando torna?
     sorry, but when will maretta (NAME) come back?
2 B: eh mi sa a inizio mese
     eh, I guess at the beginning of next month
3 A: ah proprio?
     oh so (late)?
4 B: eh perché lei il ritorno non l'aveva preso [xxx]
     eh, because she hadn't booked the return trip
5 C: lei aveva detto che si faceva due settimane tipo
     she said she would stay home two weeks or something
6 B: so che [la mamma, operazioni varie, cose] ci sta figurati
     I know that her mum, various surgeries, stuff, it's ok of course
```
(KIParla Corpus BOA3004)

Speaker A wants to know when a friend will come back, Speakers B and C cooperatively answer that she will remain at their parents' house for two weeks (lines 1–5), then B provides a list of examples that are meant to convey the reasons underlying the friend's choice (line 6). The reason is however never mentioned, but only abstracted from the open list 'her mum, various surgeries, stuff', which leads to the ad hoc categorization [complicated situations connected with her family's health]. Speaker B, at the end of the list, provides a subjective evaluation ('it's ok of course', line 6)) about the category just construed.

What are the factors triggering the right abstraction from 'mum' and 'various surgeries', without a general category formulation? First of all we have context: in the ongoing languaging process, context is a continuous source for relevant frames to be activated, which provide the background information necessary to mutual understanding. In (13), the background information is that the girl in question will stay at her parents' house for a longer period than expected, and it is this frame (progressively construed in lines 1–5) that allows for a consistent interpretation of the two examples in line 6. The first element of the list 'mum' drives the interlocutors' focus of attention towards the close family set and issues regarding the mother, while 'various surgeries' triggers the category delimitation around a plurality of medical events, pushing the category borders further enough to include other potential problems associated to surgeries and family. In other words, the two examples refer to mutually relevant elements (medical problems and close family) that provide the borders within which the category has to be construed.

The incremental choice of the category members explicitly mentioned is crucial also in Example (14), where Speaker A is talking about her study experience abroad, arguing that she had to learn how to say small things in a different language and this is something that she had not learnt in class:

(14)
```
1 A: essendo sempre stata abituata comunque a parlare con eh nella mia lingua
     having always been used to speak anyway with eh in my language
2    dover eh mh chiedere anche [per le minime cose]ₐ₁ ad esempio eh mh [a che ora
     when you have to mh ask even [the smallest things]ₐ₁ like for
     instance eh mh [at what time
3    si mangia a cena perché poi dobbiamo uscire]♭ o non lo so eh
     we have dinner because then we have to go out]♭ or I don't know eh
4    [la carta igienica in bagno]꜀ ad esempio, [proprio piccole cose]ₐ₂
     toilet paper in the bathroom for instance, [truly small things]ₐ₂
5 B: mhmh sì
     mhmh yes
6 A: sono comunque e doverle dire in una lingua che non è la tua
     they're anyway and when you have to say them in a language that is not yours
7    sono comunque [cose molto particolari]ₐ₃
     they are anyway [very particular things]ₐ₃
```
                                                    (KIParla Corpus TOD2011)

The category she is interested in is explicitly named through the label 'the smallest things' ((a1) line 2), which is in itself highly generic and calls for more specific descriptions, to construe it in an ad hoc way for the discourse aims. Speaker A therefore undertakes an ad hoc categorization process, mirroring her search for the most relevant examples: the first example she provides is 'at what time we have dinner' ((b), lines 2–3), and the second one is 'toilet paper in the bathroom' ((c), line 4): in what sense can these situations be considered 'smallest things'? They denote situations belonging to the everyday-life frame, the former being more acceptable in a public situation, the latter being more intimate and linked to a familiar context. While example (b) could occur in a textbook and could easily be learnt in a foreign language class, thus constituting a rather prototypical case, example (c) is intentionally chosen among the things that are less likely to be mentioned in class, when studying a foreign language. Example (c) is thus identified with a specific aim, namely, to build the category borders in an unexpected way, so as to include 'truly small things' ((a2, line 4), whereby *proprio* 'truly' focalizes the fact that the example just mentioned is an exceptionally small thing and can *still* be included in the category. The speaker's incremental ad hoc categorization is evident by the two reformulations that conclude the process, namely 'truly small things' ((a2), line 4) and 'very particular things' ((a3), line 7), encapsulating the process of category construction, achieved by means of precise exemplification choices. In line 5 Speaker B acknowledges that the category communication has been felicitous.

4.2   Beyond the borders: Ad hoc categorization of the outside

The languaging perspective allows to widen the observation scope to the entire interactional sequence, and this reveals strategies that involve ad hoc categorization not only within the category being communicated, but also outside its borders. In order to cooperatively build a category, speakers indeed frequently choose to exemplify what is *not* part of it, either by explicit negation or by contrast. This is revealing of the online process through which speakers try to achieve mutual understanding, which is not predetermined from the beginning, but rather evolves through resonance during the interaction (Du Bois 2014) and mirrors the speakers' attempts towards the most effective and accessible way of reaching the hearer's attention and agreement. Let us consider Example (15):

(15)
1 A: *fra gli scrittori americani è proprio* [*uno di quelli che sopporto di meno*]ₐ
      among the American writers he is [one of those that I can't stand the most]ₐ
2   *cioè tu piglia* [*un miller, henry miller*]ᵦ
     I mean, take [a Miller, Henry Miller]ᵦ

```
3      piglia [un un roth per esempio]c
       take [a Roth for example]c
4      [Palahniuk]d qualsiasi cosa
       [Palahniuk]d, anything
5 B:   no, non mi piace
       no, I don't like him
6      a me piace [paul auster]e
       I like [Paul Auster]e
7 A:   paul auster non lo conosco
       Paul Auster, I don't know him                    (KIParla Corpus BOA3006)
```

Speaker A is talking about Hemingway and is arguing that he is 'one the [American authors] that he can't stand the most' ((a), line 1), thus introducing the wider category of authors that he does not like. Yet, instead of directly building this category, the speaker provides a list of authors that he *does* like, thus exemplifying what could *not* be included in the category: Miller ((b), line 2), Roth ((c), line 3), Palahniuk ((d), line 4). In line 5, Speaker B takes the turn and starts negotiating the category being constructed, by first expressing disagreement on example (d) Palahniuk, who would actually fit the 'bad authors' category according to him (because he 'doesn't like him', line 5), then suggesting to replace it by Paul Auster ((e), line 6), which he likes instead. This contribution is however infelicitous, because the new example is not part of the shared background and Speaker A does not know him. It is interesting to note the strategy that Speaker A employs to turn a proper name, inherently specific and identifiable, into an example representative of some larger category: he says *a* Miller ((b), line 2) and *a* Roth ((c), line 3), thus introducing the proper name through an indefinite article (cf. Mauri & Sansò 2019 on ad hoc categorization of proper names).

Example (16) provides another case in point:

```
   (16)
1 A:  la mattina a casa loro si mangiava
      in the morning they used to eat
2     non [biscotti o mh cose diciamo più dolci]a
      not [biscuits or mh let's say sweet stuff]a
3     ma [cose più salate]b quindi ad esempio [toast uova eh]c mh
      but [more salted stuff]b so for instance [toast, eggs eh]c mh
4     insomma [la classica colazione più all'americana]d
      in short [the classic American-style breakfast]d
5 B:  mhmh
      mhmh                                              (KIParla Corpus TOD2011)
```

Speaker A aims to communicate what she used to eat for breakfast during her stay by an Irish family. She starts by construing the category of things that she did *not* eat by means of exemplification ((a) line 2, 'biscuits or sweet stuff'), thus incrementally focusing on what is outside the borders. She then turns to the inside and undertakes an ad hoc categorization process to communicate what she *did* eat (line 3), exploiting the opposition with the 'outside' category just built (a): the

formulation she provides ('more salted things' (b)) is indeed rather general and is symmetric to 'sweet stuff' (a). The actual category members are then exemplified by 'toast, eggs' ((c), line 3), through a zooming-in movement, which leads to the final reformulation 'the classic American-style breakfast' ((d), line 4).

Once again, Speaker B provides a backchannel, allowing Speaker A to continue, knowing that the interaction is working as expected.

4.3   Progressive zoom into the interlocutors' experience

Ad hoc categorization frequently follows a cline that goes from general and abstract statements to highly specific situations intimately involving the interlocutors. The gradual co-construction of a shared category may indeed go hand in hand with a progressive zoom into the specific interlocutors' experience, who incrementally focus on the individual entities or events that match the category being built.

In Example (17), in line 1 the speaker, talking about her experience abroad in Finland, argues that 'it is a completely different situation because people are much more closed-minded' ((a), line 1), thus providing a general frame. Then, to make the point clear, she exemplifies the frame further in lines 4–5 providing a first list of examples representative of how the situation is different (a1): 'they are locked in the house', 'they speak little', 'they also have few places where they can meet'. What we observe is thus a movement going from a general categorization in terms of situation, to a narrower category of [people's behavior], communicated by means of an exemplar-driven process.

(17)

```
1    e' proprio [una situazione diversa perché la gente è molto più chiusa]ₐ
     it is [a completely different situation because people are much more
     closed-minded]ₐ
2    magari anche
     maybe also
3    è anche una conseguenza del clima così rigido che hanno
     it is also a consequence of the harsh climate they have
4    comunque [la gente è molto chiusa, sta tanto chiusa in casa, parlano poco
     anyway [people are very closed, they are so locked in the house, they speak
     little
5    ad esempio hanno anche pochi luoghi in cui eh mh in cui incontrarsi]ₐ₁
     for example they also have few places where they can meet]ₐ₁
6    per es per esempio tipo [i giovani]b
     for example like [young people]b
7    tipo [noi]c
     like [us]c
8    [la sera]d dato che
     [in the evening]d given that
9    tipo [il venerdi' sera]e
     like [on Friday evening]e
```

```
10    dato che la maggior parte dei locali comunque chiudono alle otto di
      sera i bar
      since most of the places still close at eight in the evening, the bars
11    prima di andare tipo in discoteca si radunano tipo [negli androni dei eh
      dei eh
      before going like in the disco they gather like [in the halls of the eh of
      the eh
12    centri commerciali aperti 24 su 24 oppure nei eh all' ingresso dei delle
      shopping centers open around the clock or in eh at the entrance of the
13    banche dove ci sono i bancomat]f
      banks where there are ATMs]f
14    cioe' quindi puoi capire
      I mean so you can understand                    (KIParla Corpus TOD2005)
```

The last example in line 5 ('they also have few places where they can meet') is in focus, as underlined by the use of the additive focus-sensitive adverb *anche* 'also', and allows the speaker to open an exemplification chain, which is embedded in the ad hoc categorization in (a1) and begins a progressive zoom into the speaker's particular experience. The exemplification starts at line 6 and shows three parallel zooming-in tracks. First, 'people' is narrowed down to 'young people' ((b) line 6) and then to 'us' ((c) line 7), strongly anchoring categorization to the speech participants. Second, the general statement is narrowed down to 'evenings' ((d) line 8) and then to 'Friday evenings' ((e) line 9), thus zooming in to access a highly specific and representative example. Finally, the list of examples in lines 11–13 (f) is aimed at elaborating the 'few places where they can meet' (line 5). In line 14 we observe an explicit appeal to the interlocutor's understanding ('I mean so you can understand'), which appears to be the speaker's ultimate aim.

The discourse pattern observed in (17) is summarized in Figure (1):
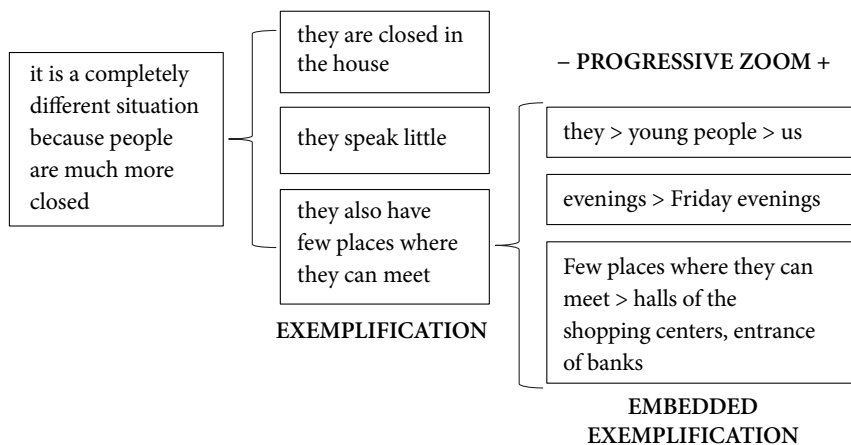
Figure 1. Progressive zoom in ad hoc categorization

An incremental zoom into the speaker's specific experience can be observed also in the ad hoc categorization being construed in Example (18), where the abstract concept of 'living the city in a different way' is progressively focused and communicated.

(18)

1   *e adesso dopo invece appunto due anni sto molto di più in casa* [*vivo la città in modo diverso*]ₐ₁
    and now, however, after just two years I am much more at home, [I live the city in a different way]ₐ₁

2   ***per esempio*** [*giro molto di meno*
    for example [I walk around much less

3   *sto sempre nei soliti posti*]ᵦ
    I'm always in in the usual places]ᵦ

4   ***per esempio*** [*qua a santa cristina*]꜀ *perchè vengo a studiare*
    for example [here in Santa Cristina]꜀ because I come (here) to study

5   *vengo a lezione qua c'è i m i miei corsi sono qui*
    I come (here) to class, here is m my courses are here

6   *e quindi eh vi eh come dire* [*vivo i posti in modo più intimo beh*]ₐ₂
    and so eh there eh how can I say [I live the places in a more
    intimate way well]ₐ₂   (KIParla Corpus BOA3015)

The speaker formulates the abstract frame 'I live the city in a different way' (line 1 (a₁)), which requires some shared categorization to allow mutual understanding. Ad hoc categorization is achieved by stepping down from the abstract formulation to the experience level, choosing examples that are representative of what the speaker means by 'different way' (cf. (b), lines 2 and 3): 'I walk around much less' and 'I'm always in the usual places'.

As we observed for (17), also in this case exemplification incrementally proceeds by zooming into the actual life experience of the speaker: in the last example 'I'm always in the usual places' the speaker embeds a further exemplification level, by naming the specific place she has in mind, namely 'here in Santa Cristina' ((c), line 4). We notice again the use a deictic expression (*qua* 'here' in line 4, cf. 'us' in Example (17)), which strongly anchors the end of the zooming exemplification to the here-and-now of the speech act. Line 6 closes the ad hoc categorization process by providing a reformulation for the category introduced in (a1) (line 1), replacing the term 'city' by 'places' and the term 'different' by 'intimate' ((a₂), line 6), in the light of the progressive focus that guided not only the hearer, but the speaker herself towards the best category label. The reformulation (a₂) is introduced by two discourse markers, *quindi* 'so' and *come dire* 'how can I say', clearly revealing of the speaker's online search for the right phrasing.

The discourse pattern observed in (14) is summarized in Figure (2):
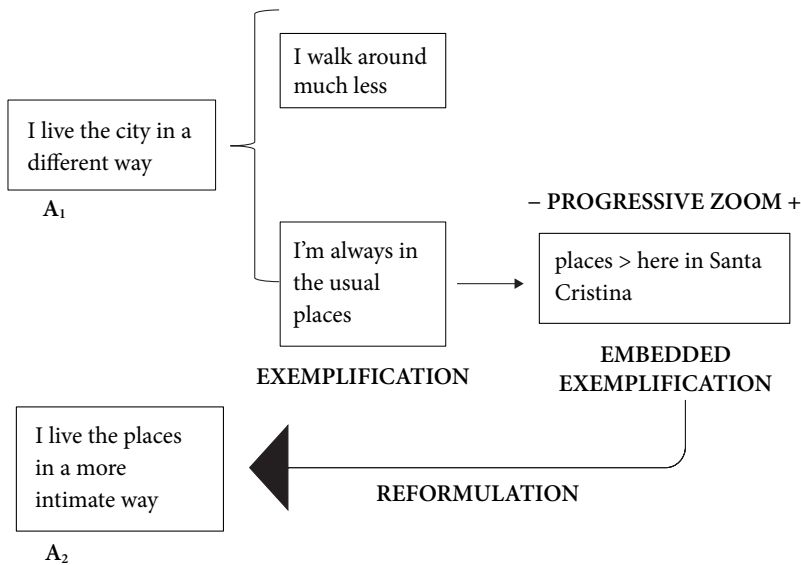
Figure 2. Progressive zoom and reformulation

5. Conclusions

The aim of this paper was to describe the role that linguistic interaction plays in category construction and in its communication, by looking at data of spoken language. We started by identifying a particular way of building categories that is inherently interactional and indexical, namely ad hoc categorization, as opposed to category labeling. We argued that ad hoc categorization can be analyzed as a bottom-up, goal-driven, context-dependent categorization process, characterized by the use of one or more examples to refer to a higher-level category. Ad hoc categorization is dependent on context for both its construction and its interpretation, and crucially relies on non-exhaustivity and exemplification (Section 2).

After a brief overview of the linguistic strategies that may encode ad hoc categorization, we adopted the perspective of languaging (Section 3), that is, the activity performed in speech, which is an ongoing process constantly evolving and developing, thanks to the evolving relation between interacting speakers. By looking at how ad hoc categorization is performed in linguistic interaction, it becomes clear that interaction is not only the place where categorization is communicated, but it

is frequently also the goal itself, triggering category co-construction, negotiation and sharing. In other words, categorization appears to be often instrumental to intersubjective aims, such as mutual agreement and the general management of the speakers' reciprocal positioning. In turn, it is collaboration between the inter-locutors, who actively collaborate in finding examples for the relevant category and who explicitly acknowledge the process going on, that allows to achieve mutual understanding.

In Section 4 we observed the incremental dimension of category construction and delimitation. In this process, speaker and hearer are mutually and contem-porarily involved in the identification of the category members and the category boundaries, recurring to exemplification along a progressive zooming-in move-ment. In particular, ad hoc categorization frequently shows a gradual cline from general category formulations towards the highly specific experience of the speaker, who follows the online flow of reference construction.

All in all, linguistic interaction stands out as playing a crucial role in exem-plar-driven category construction (cf. Clark & Wilkes-Gibbs 1986), because it sets the common ground against which ad hoc categorization occurs, it influences the interlocutors' choices regarding the linguistic strategies to be employed, and it may even act as the real trigger pushing speakers to undertake ad hoc categorization. Last but not least, interaction is also frequently what allows categorization itself to be successful, thanks to processes of cooperation and negotiation that the interloc-utors are constantly engaged in.

References

Aikhenvald, Alexandra Y. 2008. *The Manambu Language of East Sepik, Papua New Guinea*. Oxford: OUP.

Anderson, Stephen R. & Edward Keenan. 1985. Deixis. In T. Shopen (ed.), *Language Typology and Syntactic Fieldwork* vol. III, 259–308. Cambridge: Cambridge University Press.

Arcodia, Giorgio F. & Mauri, Caterina. 2020. Exemplar-based compounds: The case of Chinese. *Language Sciences* 81: 101232. https://doi.org/10.1016/j.langsci.2019.06.002

Ariel, Mira & Mauri, Caterina. 2018. Why use or? *Linguistics* 56(5): 939–993. https://doi.org/10.1515/ling-2018-0020

Auer, Peter. 2009. On-line syntax: Thoughts on the temporality of spoken language. *Language Sciences* 31(1): 1–13. https://doi.org/10.1016/j.langsci.2007.10.004

Auer, Peter & Pfänder, Stefan (eds). 2011. *Constructions: Emerging and Emergent*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110229080

Barotto, Allesandra & Mauri, Caterina. 2018. Constructing lists to construct categories. *Italian Journal of Linguistics* 30: 95–134.

Barotto, Allesandra. 2021. *Exemplification and Categorization: The Case of Japanese*. Berlin: Mouton de Gruyter.

Barotto, Allesandra & Lo Baido, Maria Cristina. 2021. Exemplification in interaction: from reformulation to the creation of common ground. In *Building Categories in Interaction: Linguistic Resources at Work*, Caterina Mauri, Eugenio Goria & Ilaria Fiorentini (eds). Amsterdam: John Benjamins. (This volume)

Barsalou, Lawrence W. 1983. Ad hoc categories. *Memory and Cognition* 11(3): 211–227. https://doi.org/10.3758/BF03196968

Barsalou, Lawrence W. 1991. Deriving categories to achieve goals. In *The Psychology of Learning and Motivation: Advances in Research and Theory*, Gordon H. Bower (ed.), 1–64. San Diego CA: Academic Press.

Barsalou, Lawrence W. 2003. Situated simulation in the human conceptual system. *Language and Cognitive Processes* 18: 513–562. https://doi.org/10.1080/01690960344000026

Barsalou, Lawrence W. 2010. Ad hoc categories. In *The Cambridge Encyclopedia of the Language Sciences*, Patrick Colm Hogan (ed.), 87–88. Cambridge: CUP.

Becker, Alton L. 1988. Language in particular: A lecture. In *Linguistics in Context*, Deborah Tannen (ed.), 17–35. Norwood NJ: Ablex.

Becker, Alton L. 1991. Language and languaging. *Language and Communication* 11(2): 33–35. https://doi.org/10.1016/0271-5309(91)90013-L

Berlin, Brent & Kay, Paul. 1969. *Basic Color Terms: Their Universality and Evolution*. Berkeley CA: University of California Press.

Bybee, Joan L. 2015. *Language Change*. Cambridge: CUP. https://doi.org/10.1017/CBO9781139096768

Bybee, Joan L. & Hopper, Paul (eds). 2001. *Frequency and the Emergence of Linguistic Structure* [Typological Studies in Language 45]. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.45

Cheshire, Jenny. 2007. Discourse variation, grammaticalisation and stuff like that. *Journal of Sociolinguistics* 11: 155–193. https://doi.org/10.1111/j.1467-9841.2007.00317.x

Clark, Herbert H. & Wilkes-Gibbs, Deanna. 1986. Referring as a collaborative process. *Cognition* 22(1): 1–39. https://doi.org/10.1016/0010-0277(86)90010-7

Cowley, Stephen J. 2007. How human infants deal with symbol grounding. *Interaction Studies* 8(1): 83–104. https://doi.org/10.1075/is.8.1.06cow

Croft, William & Cruse, Alan D. 2004. *Cognitive Linguistics*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511803864

Daniel, Michael & Moravcsik, Edith. 2013. The associative plural. In *The World Atlas of Language Structures Online*, Matthew Dryer & Martin Haspelmath (eds), Ch. 36. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/36> (22 March 2021).

Davies, Mark. 2008. The Corpus of Contemporary American English. <www.english-corpora.org/coca/> (22 March 2021).

Du Bois, John W. 2014. Towards a dialogic syntax. *Cognitive Linguistics* 25(3): 359–410. https://doi.org/10.1515/cog-2014-0024

Fiorentini, Ilaria & Miola, Emanuele. 2020. Disjunctive/conjunctive/whatever: The development of Italian barra ('slash') as a non-exhaustive connective. *Language Sciences* 81: 101234. https://doi.org/10.1016/j.langsci.2019.06.004

Inkelas, Sharon. 2014. Non-concatenative derivation. Reduplication. In *The Oxford Handbook of Derivational Morphology*, Rochelle Lieber & Pavol Štekauer (eds), 169–189. Oxford: OUP.

Kuperschmidt, Itai. (2018). Alternative relations and higher level categories. In *Linguistic Strategies for the Construction of Ad Hoc Categories: Synchronic and Diachronic Perspectives*, Caterina Mauri & Andrea Sansò (eds). Special issue of *Folia Linguistica* 52(s39–1): 125–145. https://doi.org/10.1515/flih-2018-0006

Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago IL: The University of Chicago Press. https://doi.org/10.7208/chicago/9780226471013.001.0001

Levinson, Stephen C. 2003. *Space in Language and Cognition: Explorations in Cognitive Diversity*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511613609

Lidz, Jefferey. 2000. Echo Reduplication in Kannada: Implications for a theory of word formation. *University of Pennsylvania Working Papers in Linguistics* 6: 145–166.

Linell, Per. 2009. *Rethinking Language, Mind and World Dialogically. Interactional and Contextual Theories of Human Sense-making*. Charlotte NC: Information Age Publishing.

Lo Baido, Maria Cristina. 2018. Categorization via exemplification: Evidence from Italian. In *Linguistic Strategies for the Construction of Ad Hoc Categories: Synchronic and Diachronic Perspectives*, Caterina Mauri & Andrea Sansò (eds). Special issue of *Folia Linguistica* 52(s39–1): 69–95. https://doi.org/10.1515/flih-2018-0007

Magni, Elisabetta. 2018. Collective suffixes and ad hoc categories: From Latin -ālia to Italian -aglia. In *Linguistic Strategies for the Construction of Ad Hoc Categories: Synchronic and Diachronic Perspectives*, Caterina Mauri & Andrea Sansò (eds). Special issue of *Folia Linguistica* 52(s39–1): 201–224. https://doi.org/10.1515/flih-2018-0008

Maturana, Humberto R. 1983. What is it to see? (¿Qué es ver?). *Archivos de Biología y Medicina Experimentales* 16(3–4): 255–269.

Mauri, Caterina. 2017. Building and interpreting ad hoc categories: A linguistic analysis. In *Formal Models in the Study of Language*, Joanna Blochowiak, Cristina Grisot, Stephanie Durrleman-Tame & Christopher Laenzlinger (eds), 297–326. Berlin: Springer. https://doi.org/10.1007/978-3-319-48832-5_16

Mauri, Caterina, Ballarè, Silvia, Goria, Eugenio, Cerruti, Massimo & Suriano, Francesco. 2019. KIParla corpus: A new resource for spoken Italian. In *Proceedings of the 6th Italian Conference on Computational Linguistics CLiC-it*, Raffaela Bernardi, Roberto Navigli & Giovanni Semeraro (eds). <http://ceur-ws.org/Vol-2481/paper45.pdf> (22 March 2021).

Mauri, Caterina, Goria, Eugenio & Fiorentini, Ilaria. 2019. Non-exhaustive lists in spoken language: A construction grammatical perspective. *Constructions and Frames* 11(2): 290–316. https://doi.org/10.1075/cf.00032.mau

Mauri, Caterina & Sansò, Andrea. 2018. Linguistic strategies for ad hoc categorization: Theoretical assessment and cross-linguistic variation. In *Linguistic Strategies for the Construction of Ad Hoc Categories: Synchronic and Diachronic Perspectives*, Caterina Mauri & Andrea Sansò (eds). Special issue of *Folia Linguistica* 52(s39–1): 1–35. https://doi.org/10.1515/flih-2018-0001

Mauri, Caterina & Sansò, Andrea (eds). 2020. *Ad Hoc Categorization and Language: The Construction of Ad Hoc Categories in Discourse. Special issue of Language Sciences* 81.

Mauri, Caterina & Sansò, Andrea. 2019. Nouns & co. Converging evidence in the analysis of associative plurals. *STUF – Language Typology and Universals* 72: 603–626. https://doi.org/10.1515/stuf-2019-0023

Mauri, Caterina & Sansò, Andrea. Forthcoming. Heterogeneous sets: A diachronic typology of associative and similative plurals. *Linguistic Typology*. https://doi.org/10.1515/lingty-2021-2072

Overstreet, Maryann. 1999. *Whales, Candlelight, and Stuff Like That: General Extenders in English Discourse*. Oxford: OUP.

Raimondi, Vincenzo. 2019. The bio-logic of languaging and its epistemological background. *Language Sciences* 71: 19–26. https://doi.org/10.1016/j.langsci.2018.03.003

Rosch, Eleanor H. 1973. Natural categories. *Cognitive Psychology* 4: 328–350. https://doi.org/10.1016/0010-0285(73)90017-0

Rosch, Eleanor H. 1975. Cognitive reference points. *Cognitive Psychology* 7: 532–547. https://doi.org/10.1016/0010-0285(75)90021-3

Steffensen, Sune Vork. 2009. Language, languaging, and the extended mind hypothesis. *Pragmatics and Cognition* 17: 677–697. https://doi.org/10.1075/pc.17.3.10ste

Steffensen, Sune Vork. 2015. Distributed language and dialogism: Notes on non-locality, sensemaking and interactivity. *Language Sciences* 50: 105–119. https://doi.org/10.1016/j.langsci.2015.01.004

Stolz, Thomas. 2008. Total reduplication vs. echo-word formation in language contact situations. In *Language Contact and Contact Languages* [Hamburg Studies on Multilingualism 7], Peter Siemund & Noemi Kintana (eds), 107–132. Amsterdam: John Benjamins. https://doi.org/10.1075/hsm.7.07sto

Swain, Merrill. 2006. Languaging, agency and collaboration in advanced second language learning. In *Advanced Language Learning: The Contributions of Halliday and Vygotsky*, Heidi Byrnes (ed.), 95–108. London: Continuum.

Swain, Merrill & Watanabe, Yuko. 2013. Languaging: Collaborative dialogue as a source of second language learning. In *The Encyclopedia of Applied Linguistics*, Carol A. Chapelle (ed.), 3218–3225. Oxford: Blackwell.

Thibault, Paul J. 2017. The reflexivity of human languaging and Nigel Love's two orders of language. *Language Sciences* 61: 74–85. https://doi.org/10.1016/j.langsci.2016.09.014

Traugott, Elizabeth Closs. 2003. Constructions in grammaticalization. In *The Handbook of Historical Linguistics*, Brian D. Joseph & Richard D. Janda (eds), 624–647. Oxford: Blackwell. https://doi.org/10.1002/9780470756393.ch20

Traugott, Elizabeth Close & Trousdale, Graeme. 2010. Gradience, gradualness and grammaticalization: How do they intersect? In *Gradience, Gradualness, and Grammaticalization* [Typological Studies in Language 90], Elizabeth Close Traugott & Graeme Trousdale (eds), 19–44. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.90.04tra

Traugott, Elizabeth Close & Trousdale, Graeme. 2013. *Constructionalization and Constructional Changes*. Oxford: OUP. https://doi.org/10.1093/acprof:oso/9780199679898.001.0001

Voghera, Miriam, Iacobini, Claudio, Savy, Renata, Cutugno, Francesco, De Rosa, Aurelio & Alfano, Iolanda. 2014. *VoLIP: A Searchable Italian Spoken Corpus, in Complex Visibles Out There*. In *Proceedings of the Olomouc Linguistics Colloquium: Language Use and Linguistic Structure*, Ludmila Veselovská & Markéta Janebová (eds), 628–640. Olomouc: Palacký University.

Voghera, Miriam. 2017. *Dal parlato alla grammatica*. Roma: Carocci.

Wälchli, Bernhard. 2005. *Co-compounds and Natural Coordination*. Oxford: OUP. https://doi.org/10.1093/acprof:oso/9780199276219.001.0001

Zhang, Niina Ning. 2008. Encoding exhaustivity. *USTWPL* 4: 133–143.

CHAPTER 3

Categories at the interface
of cognition and action

Lawrence W. Barsalou
University of Glasgow

Grounded cognition offers a natural framework for studying categories that lie at the interface of cognition and action. From this perspective, cognition emerges from the coupling of the brain, the modalities, the body, and the environment. Situated action, in particular, links these domains together, as perceived entities and events in the world (e.g., hedonic stimuli, social agents) trigger self-relevant responses (e.g., goals, values, identities, norms), which in turn produce bodily states (e.g., affect, motivation) that initiate actions (e.g., bodily actions, vocalizations) and ultimately produce outcomes, again in the world (e.g., reward, punishment). Ad hoc and goal-derived categories emerge between the internal cognitive states and external physical states that arise during the pursuit of situated action. Although these categories behave similarly in some ways to conventional taxonomic categories (e.g., possessing graded structure), they differ in others, especially in their dependence on context (typically contexts of situated action). On many occasions, these categories are constructed to support current situated action (ad hoc categories), but when they become relevant to situated action across many occasions, they become well-established in memory (goal-derived categories). Across these categories, ideals play central roles in determining graded structure, supporting the goals that drive situated action. Event frames offer a natural means of understanding how ad hoc and goal-derived categories become constructed compositionally, and how they offer an interface between cognition and the world, essentially providing coordinated patterns of values for instantiating frame variables. Interestingly, this frame-based account can be naturally implemented in the simulation mechanisms of perceptual symbol systems, further grounding ad hoc categories in relations between the brain, modalities, body, and world during situated action.

Keywords: abstraction, ad hoc categories, frames, goal-derived categories, grounded cognition, ideals, situated action (goal-directed action), typicality (graded structure)

1.   Introduction

Humans exhibit prolific abilities to generate and use categories. If we assume that every word has a potential category of instances associated with it, then the numerosity and diversity of human categories begins with the tens of thousands of words in a language. If we further assume that words can combine compositionally to form still more categories, then the number of human categories is essentially infinite (e.g., Barsalou 1999; Fodor & Pylyshyn 1988). Of interest in this chapter is the subset of these categories that support agents as they perform situated action to achieve goals in the world, what I will refer to as *ad hoc categories* and *goal-derived categories* – categories that lie at the interface of cognition and action.

Because the perspective of grounded cognition offers a natural framework for understanding categories that support action, I first review this perspective briefly (Section 2). From the grounded perspective, I then review basic properties of ad hoc and goal-derived categories (Section 3), including the factors that determine their internal graded structure (Section 4). I then turn to the compositional construction of ad hoc and goal-derived categories, focusing on the central roles of frames (Section 5). Finally, I explore how these categories might be grounded in the space-time regions of multimodal simulations (Section 6). An emerging theme will be that ad hoc and goal-derived categories constitute an interface between event frames in cognition and goal pursuit in the world. Section 7 concludes with limitations of work to date and potential directions for future work.

Regrettably, for reasons of time and space, I focus on my previous work in this area, omitting many other relevant articles and findings. It would be useful to have a comprehensive integrative review of the literature.

2.   Grounded cognition

Perhaps the best way to understand the perspective of grounded cognition is to consider a significant alternative. The dominant approach to cognition since the Cognitive Revolution in the 1950s assumes that the cognitive system is a module in the brain, separate from other modules for perception, action, emotion, and so forth (Fodor 1975; Pylyshyn 1984). Hurley (2001) referred to this approach as the "sandwich model," given that the module for cognition is sandwiched between modules for perception and action (see Figure 1A). From this perspective, researchers can study and understand cognition as an isolated system without taking perception and action into account. Rather than playing any significant roles in cognition, perception is primarily a means of getting information into the cognitive module, and action is primarily of means of getting information out. A further implication of this perspective – particularly relevant for the study of categories – is that semantics and conceptual processing are viewed as the activation of abstract

A

| perception | → | **COGNITION** | → | action |

Expanding the bases of cognition

B

GROUNDED COGNITION

MODALITIES
External Perception
vision, audition, haptics,
gustation, olfaction
Internal Perception
proprioception,
interoception,
affect, reward,
introspection

BODY
face, limbs, trunk
motor system
endocrine systems
heart, breath, digestion

COGNITION
attention
working memory
long-term memory
knowledge
language
thought

PHYSICAL ENVIRONMENT
Settings
outdoor, indoor
Entities
living things, artifacts,
natural kinds, etc.

SOCIAL ENVIRONMENT
self
agents
groups
social interaction
mirroring
culture

Empirical evidence for grounding

C

MODALITIES
–evidence that the modalities become
active while processing concepts,
language, memories, thoughts
–e.g., color, shape, motion, taste,
sound, touch, interoception

BODY
–evidence for motor activations while
processing perceptions, concepts,
language, memories, thoughts
–e.g., manipulating artifacts,
approach/avoidance actions
–evidence that bodily states influence
cognition and vice versa
–e.g., facial expressions, posture

PHYSICAL ENVIRONMENT
–Gibson (ecological psychology)
–e.g., the environment is an
essential part of perception
and action
–environmental effects on judgment
–e.g., warmth, weight, height, cleanliness
–situational organization of cognition
–e.g., perceptions, concepts,
and words activate background
situations and vice versa

SOCIAL ENVIRONMENT
–evolution of human cognition
–especially language
–bodily mirroring of other agents'
actions, emotions, language, etc.
–e.g., grasping, disgust, syntax
–representing another's mental states
–e.g., theory of mind

D

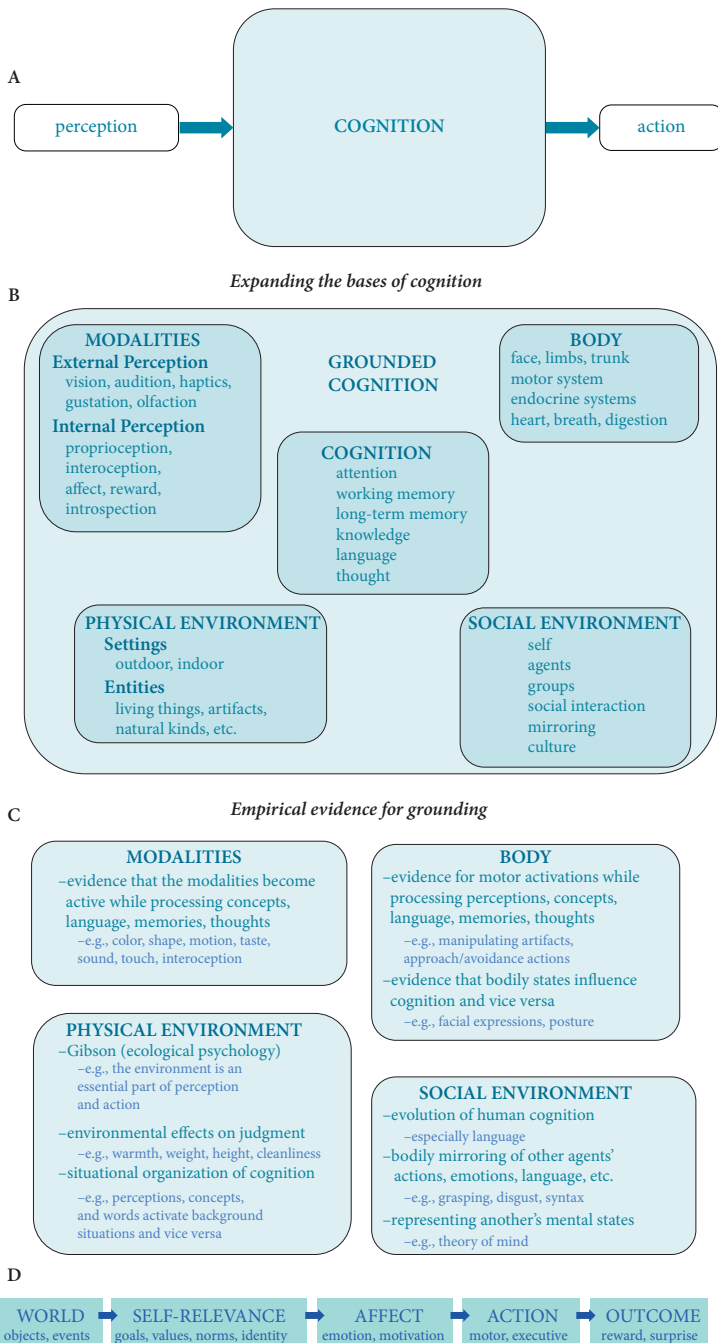| WORLD | → | SELF-RELEVANCE | → | AFFECT | → | ACTION | → | OUTCOME |
| objects, events | | goals, values, norms, identity | | emotion, motivation | | motor, executive | | reward, surprise |

Figure 1. Panel A. The sandwich model of cognition (adapted from Hurley 2001). Panels B and C. Four domains that ground cognition (modalities, body, physical and social environment), and the general kinds of evidence for them. Panel D. The situated action cycle

symbolic structures in a cognitive module, without having any critical dependencies on perception and action (as well as on other potentially important domains, such as emotion, social cognition, etc.)

The perspective of grounded cognition rejects the assumption that cognition is an isolated module in the brain. Instead, the grounded perspective proposes that cognition emerges across diverse domains that become coupled together during situated action (e.g., Barsalou 1999, 2008, 2009; Clark 1998, 2008; Coello & Fischer 2016a, 2016b; De Vega, Glenberg & Graesser 2008; Pecher & Zwaan 2005). No doubt, the brain contains basic cognitive mechanisms associated with attention, memory, language, and thought. Rather than constituting a stand-alone module sufficient for producing cognition, however, these mechanisms depend intrinsically on interaction with mechanisms in other domains for cognition to emerge. The next four sections briefly review four domains that contribute to grounding cognition (summarized in Figures 1B and 1C).

2.1   Grounding in perceptual systems and the construct of simulation

Much research implicates perceptual mechanisms across the modalities in cognition. When cognition represents perceptual properties conceptually, they typically appear to utilize mechanisms in the brain's relevant modality-specific systems. When the color of an object is represented conceptually, for example (e.g., yellow for lemons), the color perception system appears to play central roles (Hsu, Frankland & Thompson-Schill 2012; Martin 2016: 201; Simmons et al. 2007; Wang et al. 2013). When the sound of an object, animal, or event is represented conceptually (e.g., loud for explosion), the auditory system plays central roles (e.g., Hoenig et al. 2011; Kiefer, Sim, Herrnberger, Grothe & Hoenig 2008; Lewis 2010; Trumpp, Kliese, Hoenig, Haarmeier & Kiefer 2013; Yao, Belin & Scheepers 2011, 2012). Similarly, when people represent the taste of a food, taste and reward areas play central roles (e.g., Chen, Papies, & Barsalou 2016; Simmons, Martin, & Barsalou 2005; van der Laan, de Ridder, Viergever, & Smeets 2011). In general, cognition appears to "reuse" perceptual systems for representational purposes when processing many types of semantic information (e.g., Anderson 2010 2016; Barsalou 2016b) but not all (e.g., Barsalou 2016a; Barsalou, Dutriaux & Scheepers 2018).

Reenacting perceptual states to represent information conceptually is often referred to as *simulation* (Barsalou 1999, 2008, 2009, 2016b). While the brain is engaged in perceiving an entity or event, memory systems establish a record of the brain state active, including perceptual systems. When, for example, someone eats an apple, states of the visual, gustatory, olfactory, auditory, and tactile systems become superimposed together on memory as an associative pattern, accumulating

across interactions with the category of apples to establish an entrenched network that represents the category (Barsalou 2017). On later occasions, when cues associated with apples are encountered (e.g., seeing an apple, being in a snack situation, hearing the word "apple"), this network becomes active and reenacts the kind of state that the brain would be in if it were eating an apple (Papies & Barsalou 2015).

Simulations need not be complete, conscious, or veridical, often being vague, unconscious, and biased instead; they can potentially be implemented at varying levels of granularity and vividness (Barsalou 1999). Simulations are also highly task dependent and context-specific (Lebois, Wilson-Mendenhall & Barsalou 2015), and are not the only way conceptual information is represented (Barsalou 2016b). Regardless of these factors, simulations – in general – provide diverse inferences about a category that can be used for a wide variety of functions (Barsalou 2008, 2016d). As described later (Section 6.1), simulations may play central roles in the representation of ad hoc and goal-derived categories.

2.2   Grounding in the motor system and body

Much research similarly implicates the motor system and body in cognition (e.g., Glenberg 1997; Lewis 2006; Pulvermüller 1999, 2005; Pulvermüller, Moseley, Egorova, Shebani & Boulenger 2014). When people represent the semantics of action words, for example, they utilize the brain's motor system to simulate motor properties and bodily states, often very rapidly (e.g., Shtyrov, Butorina, Nikolaeva & Stroganova 2014). Furthermore, engaging the action system affects conceptual processing, as when configuring the face into a smile produces liking (e.g., Barsalou, Niedenthal, Barbey & Ruppert 2003; Niedenthal, Barsalou, Winkielman, Krauth-Gruber & Ric 2005). Similar to how cognition reuses perceptual systems for representational purposes, it reuses the motor and other bodily systems in simulations that represent motor and body semantics.

2.3   Grounding cognition in the physical environment

Since Gibson's (1966, 1979) classic arguments about the nature of perception, researchers have believed that much is to be gained from studying the coupling of perceptual systems with the physical environment. Because perception evolved in the physical environment, and because the two are coupled continually during situated action, it is impossible and misleading to study perceptual systems in isolation.

Researchers adopting the grounded perspective often make similar assumptions about cognition, namely, that it makes little sense to study cognition independently of the environment. Although cognition can become decoupled from the

environment during mind wandering and other forms of thought, cognition is typically coupled with the environment as agents pursue diverse forms of situated action in the world. As a consequence, the cognitive system establishes situational patterns in memory to guide goal-directed action effectively in familiar situations (e.g., Aydede & Robbins 2009; Barsalou 2016c, 2016d; Brooks 1991; Clark 1998, 2008). Because habitual behavior in familiar situations conditions cognitive patterns in the brain robustly, it makes sense to study cognition when it's actually coupled with these situations (analogous to Gibson's arguments for studying perception coupled with the environment). Research with the visual world paradigm makes similar arguments (e.g., Huettig, Rommers & Meyer 2011; Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy 1995). From this perspective, full-blown cognition emerges when coupled with the physical situations it has learned to process in the support of situated action (Barsalou et al. 2018).

2.4   Grounding cognition in the social environment

Theories that aim to explain the evolution of human cognition often propose that significantly enhanced social cognition was critical (e.g., Donald 1993; Tomasello 2009). Although language clearly distinguishes humans from other animals, abilities associated with social cognition underlie language, making it possible, while further contributing to the unique and powerful cognition that humans exhibit (Tomasello, Kruger & Ratner 1993). In particular, humans are unique in establishing extensive joint attention, adopting the perspective of others, representing other minds, following social norms, performing social mirroring at multiple levels, establishing self-identities, establishing cultural identities, and so forth.

Because social interaction plays such central roles in human cognition, it follows that cognition is grounded in the social environment, analogous to how it is grounded in the physical environment. Social cognition primarily emerges as humans are coupled with relevant social agents and situations while engaged in social interaction. Attempting to understand social cognitive processes without understanding their coupling with the social environment is likely to limit and distort the conclusions reached. Only when these processes are assessed as they operate in the social environment is their full extent likely to be revealed (cf. Barsalou et al., 2018).

2.5   Cognition as emergent

Putting the previous sections together, cognition is not a module that can be studied successfully in isolation. Instead, it is better viewed as a set of processes that emerge when core cognitive mechanisms are coupled with the modalities, the body, the

physical environment, and the social environment (e.g., Barsalou, Breazeal & Smith 2007; Barsalou, et al. 2018; Clark 1998, 2008; Hutchins 1995). Cognition emerges when the brain, body, and environment are coupled together.

Sometimes this approach is referred to as "embodied cognition." One problem with this description, however, is that it only refers to one form of grounding (the body), failing to capture other important forms, namely, perceptual systems, the physical environment, and the social environment. For this reason, "grounded cognition" describes this approach more comprehensively, covering all relevant forms of grounding (also see 4E cognition; Newen, Bruin & Gallagher 2018).

2.6   The situated action cycle

As we have seen, a central theme of grounded cognition is that most basically (although not exclusively) cognition supports goal-directed action (e.g., Clark 1998, 2008; Glenberg 1997). Rather than simply being an information processor that implements symbol manipulation, cognition most fundamentally enables agents to interact effectively in physical and social environments. Indeed, one could argue that cognition evolved to implement the unusually powerful forms of action that humans exhibit (Barsalou 2016a).

What I will call the *situated action cycle* offers one way of conceptualizing the relation between cognition and action (see Figure 1D). The basic idea is follows: As entities and events are encountered in the environment, they activate (cue) cognitive states of self-relevance, including goals, values, norms, identities, and so forth. As these cognitive states become active, they in turn elicit relevant affective and motivational states in the body, which in turn may induce a wide variety of actions in the motor system (from eye movements, to hand movements, to full body movements) and in the executive system (e.g., inhibition, reappraisal, planning, problem solving). Finally, the actions performed change the environment, leading to diverse outcomes including reward, surprise, and so forth. Once the environment has changed, the situated action cycle iterates, as new cognitive and bodily states lead to new actions and outcomes.

Interestingly, the situated action cycle links together the various domains of grounded cognition in Figures 1B and 1C: As events occur in the *physical* and *social environment*, the *modalities* perceive them and activate *cognitive mechanisms* that produce states *of self-relevance* (with *cognition* often using the *modalities* to simulate relevant predictions and related information). In turn, these cognitions produce *bodily states* that induce *actions*, which finally lead to outcomes in the *environment*. Given the central role of the situated action cycle across everyday activities, it's no accident that cognition emerges from all the domains in Figures 1B and 1C.

Various forms of the situated action cycle have played central roles in theories of psychology and cognitive science for decades. Behaviorist theories of conditioning, for example, focused on how cues, bodily states, actions, and outcomes become organized to implement learning and habit acquisition (while typically omitting cognitive and affective states; e.g., Bouton & Todd 2014; Domjan 2014). Early theories in cognitive science added cognitive and affective states to the Behaviorist version of the situated action cycle to explain plans, goal-directed behavior, and problem solving (e.g., Miller, Galanter & Pribram 1960; Newell & Simon 1972). Since then, theories of narrative and text processing have incorporated the basic form of the situated action cycle into how people understand and produce language about diverse forms of goal-directed events, including stories, autobiographical memories, and so forth (e.g., Baerger & McAdams 1999; Edson Escalas 2004; Reese et al. 2011; Stein & Hernandez 2007). As we will see in the remainder of this chapter, the situated action cycle offers a natural framework for understanding ad hoc and goal-derived categories.

3.   Ad hoc and goal-derived categories

The traditional study of categories in cognitive psychology and cognitive science has tended to focus on common taxonomic categories such as *animals, plants, tools, clothing, furniture*, and so forth (e.g., Hampton 1979; Murphy 2002; Rosch & Mervis 1975; Rosch, Mervis, Gray, Johnson & Boyes-Braem 1976; Smith & Medin 1981). Interestingly, and perhaps tellingly, a wide variety of less conventional categories populate cognition as well, such as *things to pack in a suitcase, ways to get from San Francisco to New York, foods not to eat on a diet*, and *things to take from one's house during a fire*. At first blush, this latter kind of category might appear to simply be a quaint idiosyncrasy of human cognition. A closer look, however, reveals that these categories are ubiquitous. As we will see, they play central roles in situated action and share important properties with common taxonomic categories.

Barsalou (1983) initially dubbed these as "ad hoc categories," given that they appear to be constructed spontaneously in an ad hoc manner when needed to achieve goals (during the situated action cycle). Of interest in this early assessment of ad hoc categories was how they differed from common taxonomic categories, and also how they might be similar. The next several sections document important properties of ad hoc categories and how they relate to the properties of taxonomic categories, beginning with dissimilarities before turning to similarities.

3.1   The correlational structure of the environment

One important difference is that ad hoc categories violate "the correlational structure of the environment," something to which common taxonomic categories often conform (Rosch & Mervis 1975; Rosch et al. 1976). "Correlational structure" refers to clusters of features that different taxonomic categories circumscribe, such as *feathers, wings, nests*, and *flying* for *birds*, contrasted with *scales, gills, fins*, and *swimming* for *fish*. As Rosch and her colleagues proposed, common taxonomic categories capture and convey the correlated patterns of features in the environment, especially at the basic (generic) level (for a definitive review, see Malt 1995). In contrast, ad hoc categories typically organize exemplars that don't exhibit similar features while instead exhibiting considerable dissimilarity, cutting across correlational structure and thereby violating it. *Things to take on a vacation*, for example, might include *shirt, apple, toothbrush, iPad*, and *snorkel*. Rather than emerging from a correlated set of properties, this category instead emerges from the context of planning a vacation. This category might not appear to share any significant overlapping features (especially physical). As described later (Sections 5.1, 5.2), however, these kinds of categories appear to be defined by abstractions that integrate frame attributes with optimizations and constraints relevant to performing situated action.

3.2   Contexts of situated action

Another important difference is that the salience of common taxonomic categories is often largely independent of context, whereas the salience of ad hoc categories typically appears to depend on it. Barsalou (1982, 1983) showed, for example, that when receiving *apple, orange, banana*, and *peach*, people immediately recognize them as belonging to the category *fruit*, without requiring any supporting context to do so (see Lebois, Wilson-Mendenhall & Barsalou 2015, for an alternative perspective). In contrast, when receiving *move to the remote reaches of Wyoming, sail around the world, go to Mexico*, and *become a drunk in Detroit*, people have no idea what category they instantiate. Interestingly, however, when participants receive a relevant context about situated action, such as, "Roy was in big trouble. The Mafia had a contract out on him. He knew he couldn't continue living in Las Vegas or he'd be dead in a week," they immediately generate the category, *ways to escape being killed by the Mafia* (Barsalou 1982, 1983).

  The importance of context illustrates the potential relevance of grounded cognition for understanding ad hoc categories: These categories often only appear to become salient in a relevant context of situated action (Section 2.6). Once a relevant environmental setting has been established, along with its self-relevance

for an agent, useful categories in this context become apparent. The ability to construct such categories prolifically is no small computational achievement (imagine building computer intelligence that could do this). It is truly impressive how readily people construct new categories to support the situated action they're pursuing. Chrysikou (2006) found that the ability to construct ad hoc categories plays a central role in successful goal pursuit.

3.3   Establishment in memory

Barsalou (1983) also demonstrated that ad hoc categories can be much less established in memory than taxonomic categories. For example, participants generated members of taxonomic categories much more consistently than they did for ad hoc categories, suggesting that only taxonomic categories were represented by well-established networks of instances in memory (i.e., which produced consistent retrieval). Similarly, when participants were presented with a set of instances from a category and later asked to recall them, they tended to recall them better for taxonomic categories than for ad hoc categories. Together these findings suggested that ad hoc categories are constructed temporarily when relevant and do not become established well in memory unless processed repeatedly on later occasions. This issue will be explored further when goal-derived categories are introduced shortly (Section 3.5).

3.4   Typicality gradients

Perhaps the most surprising initial result about ad hoc categories was that they not only exhibit typicality gradients, but exhibit ones that tend to be as stable as those found in taxonomic categories (Barsalou 1983). When the intraclass correlation is used to measure the average agreement between participants in their typicality judgments for the members of a category, the values are comparable for both category types (typically values in the range of .30 to .60, depending on the study). As Table 1 illustrates (from Barsalou 1991), ad hoc categories unexpectedly exhibit typicality gradients that are comparable in stability to those in taxonomic categories.

One explanation of this equivalence is that a common process of constructing prototypes is responsible for typicality gradients in both category types (Barsalou 1983, 1987, 1989, 1991, 1993). For both taxonomic and ad hoc categories, people construct prototypes to represent them. Whereas the prototype for *birds* might be a *sparrow*, the prototype for *foods not to eat on a diet* might be a *bacon cheeseburger* (alternatively, instead of the prototypes being actual instances, they could simply be *features* of these instances). Once the prototype for a category is established, the

Table 1. Average measures of stability for prototype structure*

Experiment	Between-subject agreement		Within-subject agreement		Contextual shift	
	Common taxonomic	Goal-derived	Common taxonomic	Goal-derived	Common taxonomic	Goal-derived
Barsalou (1983)						
Experiment 2 ratings	.50	.56	–	–	–	–
Experiment 2 rankings	.57	.54	–	–	–	–
Barsalou and Sewell (1984)						
Experiment 1a	.33	.36	–	–	.58	.64
Experiment 1b	.41	.40	–	–	.70	.86
Experiment 1c	.46[a]	.39[a]	–	–	.28[f]	.48[f]
Barsalou (1985)						
Experiment 1	.45[b]	.32[b]	–	–	–	–
Barsalou (1986)						
Experiment 1	.60	.49	.82[d]	.76[d]	.97	.88
Barsalou et al. (1986)						
Experiment 1	.57[c]	.44[c]	.81[c]	.76[c]	.04	.37
Experiment 3	.47	.40	.74	.74	−.20	−.08
Experiment 4	.49	.44	.85	.84	−.39	−.37
Experiment 7	.48	.42	.81	.82	.99	.64
Average	**.48**	**.42**	**.81**	**.78**	**.28**	**.40**

* Pairs of means indexed by the same superscript differ reliably at $p < .05$.
Reprinted from Table 1 in Barsalou (1991)

typicality of the category's members varies as a function of how similar they are to the prototype. As individual birds become increasingly similar to *sparrow* (or to its features), they become increasingly typical of *birds*; as individual foods become increasingly similar to *bacon cheeseburger*, they become increasingly typical of *foods not to eat on a diet*.

An additional possibility is that the prototype construction process dynamically depends on context (Barsalou 1983, 1987, 1989, 1991, 1993). Consistent with the perspective of grounded cognition, prototypes for categories are constructed to support situated action in the current context. As a result, prototypes vary widely both between and within individuals as the goals and action currently relevant vary. This assumption explains the considerable variability in typicality gradients that occurs not only between different individuals (again, agreement in the .30 to .60 range) but within the same individual over time (agreement in the .70 to .85 range). It is further consistent with strong effects of context on what's currently typical for a

category. When participants are asked to judge typicality from the perspective of the average US citizen, for example, they construct typicality gradients for categories that differ substantially from those constructed when asked to adopt the perspective of the average Chinese citizen. People's ability to dynamically reconfigure a category is consistent with the proposal that categories are typically conceptualized in the context of the situated action cycle, with current goals and actions determining what's prototypical. Rather than exhibiting rigid typicality gradients, categories exhibit typicality dynamically in a manner that supports current situated action.

3.5   Goal-derived categories

Barsalou (1985) noted that many categories violating the correlational structure of the environment nevertheless appear to be well established in memory. In particular, when people repeatedly perform the same situated action over and over again, the categories constructed to support it are likely to become well established. When individuals travel frequently, for example, *things to pack in a suitcase* might become well established, as might *foods not to eat on a diet* for people who diet regularly. Certainly, these categories may vary somewhat from occasion to occasion, reflecting the particular constraints of each situation. For a particular traveler, however, the category of *things to pack in a suitcase* may contain a relatively habitual set of exemplars that instantiate the category regularly, or that are at least usually considered when instantiating it.

Based on this observation, Barsalou (1985) introduced the construct of *goal-derived categories*, namely, categories constructed to achieve goals that violate the correlational structure of the environment, regardless of how well established they are in memory. Within this definition, ad hoc categories constitute the subset of goal-derived categories that are not well established. Rather than being retrieved from memory, they are constructed temporarily on-the-fly to meet current processing demands associated with a course of situated action not pursued previously.

To the extent, however, that the same course of situated action is performed habitually, what was once an ad hoc category becomes increasingly established in memory to streamline situated action. Rather than having to construct relevant categories anew on each occasion, well-established categories are retrieved that facilitate processing (see Sections 5.3, 5.4). Not only does situated action become faster, it may also become more optimal as the learning process converges on prototypes for the categories that maximize goal achievement (see Sections 4.2, 5.2).

4.   Determinants of graded structure in ad hoc and goal-derived categories

As we have seen, ad hoc categories exhibit typicality gradients, even when not well established in memory (Barsalou 1983), perhaps reflecting the dynamic construction of context-dependent prototypes. Furthermore, the stability of the typicality gradients in ad hoc categories is comparable to their stability in taxonomic categories (Table 1). An important question, however, is whether the determinants of typicality gradients are common for both category types. Essentially, this boils down to the question of whether prototypes for the two category types are similar in content (assuming that comparison to prototypes determines typicality; Section 3.4).

4.1   Family resemblance and central tendency

Barsalou (1985) investigated this issue (Barsalou 1991, provided a subsequent review). At the time, it was widely believed that typicality reflects *family resemblance*, based on classic work by Rosch and Mervis (1975) and Hampton (1979). As a category member increasingly resembles other category members, it becomes increasingly typical. A roughly equivalent way of describing this proposal is that as a category member becomes increasingly similar to the *central tendency* of its category, it becomes increasingly typical (where average similarity to a category's members – family resemblance – is roughly the same as similarity to a category's central tendency; Barsalou 1985, 1990). Thus, sparrow is a typical bird because it is more similar to other birds on the average than are most other birds (or more similar to the central tendency of birds as represented by their most common features).

Although previous work had shown clearly that similarity to central tendency is an important determinant of typicality for taxonomic categories (Hampton 1979; Rosch & Mervis 1975), it wasn't clear that central tendency also determined typicality in goal-derived categories. Because these categories violate the correlational structure of the environment, the central tendency of correlational structure may be irrelevant. Furthermore, because goal-derived categories support situated action, other kinds of information that support successful situated action may be important instead.

4.2   Ideals

In this spirit, Barsalou (1985) proposed that two new factors might be important for determining typicality in goal-derived categories: (1) ideals, and (2) frequency of instantiation. Ideals are features that exemplars should have to optimally achieve goals associated with their category during situated action. In *foods not to eat on*

a diet, for example, *minimal calories* is an ideal that category members should approximate to be good category members. Note that *minimal categories* is *not* the central tendency, which would be the *average* number of calories that *foods to eat on a diet* have. Most importantly, category members become increasingly typical as they approximate the ideal of minimal calories, rather than as they approximate the central tendency.

4.3   Frequency of instantiation

Another potential determinant of typicality in goal-derived categories is how frequently specific members are used to instantiate their respective categories during situated action. Perhaps the more often an exemplar from a goal-derived category is selected to support situated action, the more typical it becomes (i.e., because it becomes increasingly habitual). Because certain members are used frequently, they become typical.

4.4   Contrasting central tendency, ideals, and frequency of instantiation

Barsalou (1985) assessed central tendency (CT), ideals (I), and frequency of instantiation (FOI) as potential determinants of typicality in samples of taxonomic and goal-derived categories. Whereas central tendency was predicted to be the best predictor of typicality in taxonomic categories (because it captures correlational structure), ideals and frequency of instantiation were predicted to be the best predictors in goal-derived categories (because they support optimal and habitual situated action, respectively).

The results of an initial quasi-experiment (correlational in nature) strongly supported these predictions, with Table 2 presenting the results. For each category, the second-order partial correlations relevant to predicting its typicality are shown, where each predictor of interest (CT, I, or FOI) was correlated with typicality (i.e., exemplar goodness, EG), with the other two predictors partialled out. Central tendency was only associated with typicality in taxonomic categories, not in goal-derived categories (average partial correlations of .71 vs. .05). For goal-derived categories, typicality was associated with both ideals and frequency of instantiation (average partial correlations of .51 and .44). Surprisingly, however, ideals and frequency of instantiation were also important for taxonomic categories (average partial correlations of .36 and .45). This unexpected finding suggests that taxonomic categories also become organized to support goal-pursuit during situated action – not simply to represent the correlational structure of the environment. Vallee-Tourangeau, Anthony, and Austin (1998) further demonstrated the importance of situated action in organizing *both* taxonomic and ad hoc categories.

Table 2. Second-Order partial correlations by categories from experiment 1

Correlation	EG-CT	EG-I	EG-FOI
Goal-derived categories			
Birthday presents	.42	.53	.80
(how happy people are to receive it)			
Camping equipment	.15	−.12	.66
(importance to survival)			
Transportation for getting from San Francisco to New York	−.51	.56	.40
(how fast it gets people there)			
Personality characteristics in people that prevent someone	−.06	.78	.45
from being friends with them			
(how much people dislike it)			
Things to do for weekend entertainment	.34	.43	.08
(how much people enjoy doing it)			
Foods not to eat on a diet	.31	.53	.62
(how many calories it has)			
Clothes to wear in the snow	−.22	.64	.34
(how warm it keeps people)			
Picnic activities	−.28	.17	.83
(how much fun people think it is)			
Things to take from one's home during a fire	.29	.47	.41
(how valuable people think it is)			
Common taxonomic categories			
Vehicles	.86	.63	.53
(how efficient a type of transportation it is)			
Clothing	.71	.81	−.10
(how necessary it is to wear it)			
Birds	.75	.42	.78
(how much people like it)			
Weapons	.59	.91	.68
(how effective it is)			
Vegetables	.69	−.02	.29
(how much people like it)			
Sports	.74	.53	.11
(how much people enjoy it)			
Fruit	.71	.34	.49
(how much people like it)			
Furniture	.84	.03	.14
(how necessary it is to have)			
Tools	.49	.37	.29
(how important it is to have)			

Note. Ideal dimensions are in parentheses. EG is exemplar goodness, CT is central tendency, FOI is frequency of instantiation, and 1 is ideals.
Reprinted from Table 3 in Barsalou (1985)

Together, these results suggest that all categories become organized to support situated action. A subsequent experiment with artificial categories in Barsalou (1985) demonstrated that central tendency, ideals, and frequency of instantiation causally determine typicality gradients (i.e., aren't simply correlated with them).

Much subsequent work, using more sophisticated and ambitious methods, has confirmed these initial results (e.g., Borkenau 1990; Chaplin, John & Goldberg 1988; Loken & Ward 1990; Lynch, Coley & Medin 2000; Read, Jones & Miller 1990; Voorspoels, Storms & Vanpaemel 2013; Voorspoels, Vanpaemel & Storms 2010, 2011). Perhaps most surprisingly, ideals tend to emerge as not only important determinants of typicality, but often as the *most* important determinants, even for taxonomic categories. This robust finding again highlights the importance of situated action in cognition. Because ideals optimally support situated action, their centrality in determining typicality strongly implicates the pervasive influence of situated action on cognition.

5.   Deriving categories from frames to achieve goals

As we have seen, goal-derived categories are coherent categories that exhibit many of the same properties as taxonomic categories, but with important differences as well. An emerging theme is that situated action plays central roles in creating and organizing all categories, not just goal-derived categories. On the one hand, goal-derived categories are constructed to support goal achievement; on the other, categories in general become organized around ideals and frequency of instantiation to optimize goal pursuit during situated action.

Moving beyond these initial results, the issue arises as to how goal-derived categories originate in the cognitive system. In addressing this issue, Barsalou (1991) first distinguished between two modes of establishing categories. Most basically, categories can be induced from *exemplar learning*. As the members of a category are encountered, memories of them become superimposed on memory (perhaps producing abstractions in the process). In this bottom-up form of relatively receptive learning, knowledge about the world develops. To a large extent, this may be how taxonomic categories are acquired (e.g., Barsalou 1990, 2012, 2016b; Barsalou & Hale 1993; Goldstone, Kersten & Carvalho 2018; Malt 1995; McRae & Jones 2013; Murphy 2002; Smith & Medin 1981). As people encounter various kinds of plants and animals, for example, they induce categories at multiple taxonomic levels about them in a relatively bottom-up manner (although various types of top-down constraint surely contribute; e.g., Carey 2009; Malt 1995; Murphy & Medin 1985).

Alternatively, humans can readily construct new categories via top-down compositional processes typically under the control of language (e.g., Barsalou 2017;

Hinzen, Machery & Werning 2012; Winter & Hampton 2017). In this mode of category construction, language is likely to play central roles indexing combined concepts and specifying their relational structure. We next turn to Barsalou's (1991) account of how people use compositional mechanisms to produce ad hoc and goal-derived categories.

5.1   A general procedure for deriving an ad hoc category

Based on protocol analyses of people planning events (e.g., vacations), Barsalou (1991) developed an account of how ad hoc categories are newly constructed (summarized in Table 3). According to this account, an agent has currently activated a frame used to represent an event being planned. Following standard linguistic theory (e.g., Fillmore 1985), a frame is an interrelated set of attributes that become bound to a set of values in a specific situation. Barsalou (1992) presents the specific account of frames adopted here. Barsalou and Hale (1993) provide further discussion of this account, and Barsalou (1999) describes how it can be implemented in simulation mechanisms. Figure 2 illustrates an example of an event frame from a participant planning a vacation, showing only attributes of the frame mentioned, not other frame components including values, relations, and constraints (where attributes can be nested in sub-frames as shown).

Table 3. A general procedure for deriving an ad hoc category

1. Select a frame.
2. Select an attribute in a frame.
3. Identify optimizations that bear on the attribute.
4. Identify constraints that bear on the attribute.
5. Combine the attribute with the optimizations and constraints that bear on it to form a category description.
6. Search for exemplars that satisfy the category description.
7. Order exemplars according to how well they satisfy the category description, i.e.. prototype structure.
8. Store information about the category.

Reprinted from Table VI in Barsalou (1991).

As participants' protocols revealed, planning an event typically involved instantiating attributes in the relevant event frame with one or more specific values. Planning a vacation, for example, typically required instantiating the attribute for *locations* with one or more specific values, such as *beach*, which were often focused hierarchically (e.g., *beach, Florida, Panhandle, Navarre Beach*). Notably, a potential set of values for a frame attribute constitutes a goal-derived category, such as *locations to visit on a vacation*. Fully planning an event, such as a vacation, typically

requires instantiating many attribute values like this in advance, although some may be instantiated on the fly during the event itself, and some may be inadvertently passed over, perhaps disrupting the event when discovered later. When planning a vacation, for example, agents must establish values for *departure time, duration, companions, lodging, transportation, cost, activities, preparations*, and so forth. In the process, many goal-derived categories may become active to support the process of instantiating relevant frame attributes with suitable values. Language appears to play central roles in specifying these categories, as describing *locations to visit on a vacation* illustrates.
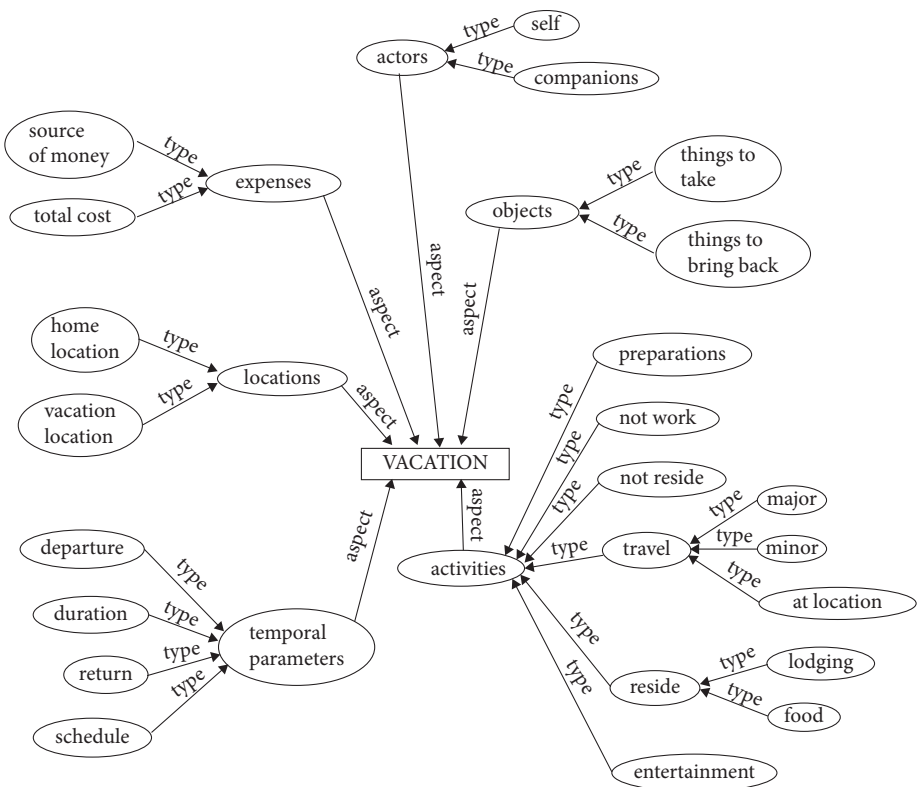

Figure 2. Attributes in a partial frame for vacation.
Reprinted from Figure 1 in Barsalou (1991)

5.2   Specifying an abstraction for an ad hoc category: Optimizations and constraints

Prior to instantiating a frame attribute with specific values, planners in Barsalou (1991) often conceptualized an abstraction that described potential values that the attribute could take (e.g., an abstraction that described potential values of *locations to visit on vacation*). Typically, these abstractions incorporated two kinds of information: *optimizations* (illustrated in Table 4), and *constraints* (illustrated in Table 5). Again, language processes appeared to play central roles in combining optimizations and constraints to establish increasingly specified frame attributes. The processes of optimization and constraint are addressed next in turn.

Table 4. Examples of optimizations from participants planning a vacation

Type of goal	Goal-derived category	Optimization
Achievement		
Knowledge	Entertainment	Maximize educational Value
Enjoyment	Location	Minimize crowdedness
Preservation		
Abandoned system	Departure	Minimize work disruption
Comfort	Vacation location	Optimize temperature
Health	Actors	Maximize immunizations
Personal security	Things to take	Maximize emergency phone numbers
Resource		
Money	Transportation	Minimize cost
Time	Things to take	Minimize time to pack
Meta-planning		
Knowledge	Vacation location	Maximize amount of knowledge
Adaptability	Schedule	Maximize flexibility of schedule
Preparations	Lodging	Maximize timing of reservations

Reprinted from Table V in Barsalou (1991)

While instantiating event frame attributes during event planning, participants generally attempted to *optimize* various goals. When planning a vacation, for example, people often attempted to optimize goals associated with money, health, security, and enjoyment. As Table 4 illustrates for vacation, participants sometimes tried to optimize the warmth of a location, minimize the cost of transportation, maximize the flexibility of vacation activities, and so forth. Essentially, this optimization process establishes *ideals* that determine typicality gradients in ad hoc and goal-derived

Table 5. Examples of constraints from participants planning vacations that utilised the partial frame for vacation in Figure 1. Note that verbal descriptions of the constraints are not direct quotes from the protocols. Instead they are redescriptions that capture the constraint directly and often incorporate surrounding context. In addition, these examples of primitive relations were sometimes extracted from constraint chains. Key: EX, exemplars; LOC, location; OBJ, object. Adapted from Figure 2 in Barsalou (1991)

Requires

Possible companions must be able to take off from work at the time I can go.
   time (departure (ex = my_departure_time)) -R-> actor (companion (work (vacation_departure (ex = my_departure_time))))
The amount of luggage I take depends on how I travel.
   activity (travel (major (ex = X (max_luggage = Y)))) -R-> objects (things to_take (amount ≤ Y))

Disallows

If my girlfriend goes with me (this requires romance with her), there can be no romances with strangers.
   actor (self (goal = activity (entertainment (ex = romance (companion = girlfriend)))) & companion (ex = girlfriend)) -D-> activity (entertainment (ex = romance (companion = stranger)))
If the vacation location is far (this requires long travel), I cannot drive.
   location (vacation (ex = X (distance = far))) -D-> activity (travel (major (ex = car (rate = slow))))

Enables

I can go on vacation when I have saved enough money.
   expenses (source = self (time = X, savings = Y > expenses (totalcost = Z))) -E-> time (departure (ex = X))
Being at the vacation location will enable visiting friends who live there, assuming they're home.
   time (schedule (loc = X, time = Y)) -E-> activity (entertainment (ex = visit (time = X, loc = Y, obj = friends (loc = x, time = Y))

Prevents

If I'm going to be flying, then I won't have my car at the vacation location.
   activity (travel (major (ex = fly))) -P-> activity (travel (at_loc (ex = my car)))
If I'm only going to be there a short time, this makes renting a boat unfeasible.
   time (duration (ex = X)) -P-> activity (entertainment (ex = rented_boat (required duration = Y > X)))

Leaves

Because I will be taking my car (this prevents distant vacation locations), I must go some place close to home.
   activity (travel (major (ex = my_car))) -L-> location (vacation (dlstance_from_home = close))
Because I want to spend little money on accommodations (this prevents going at the peak season), I can only afford going in the off season.
   location (vacation (accomodations (cost = low))) -L-> time (departure (ex = off_season))

Cooccurs

The amount of money I have available depends on the time of year.
   expenses (source (ex = self (available_money = X) <-C-> time (departure (ex = Y))
The climate a person wants to escape to depends on their current climate.
   actor (companion (ex = X (like (obj = climate (ex = Y))))) <-C-> actor (companion (ex = X (climate (ex = Z)))

categories (Sections 4.2, 4.4). As attributes of event frames are instantiated, back-ground goals being optimized establish attractive features – ideals – of potential attribute values. Once ideals become established for an attribute, they guide the generation and selection of the values considered, structuring these values around an ideal-based typicality gradient. As a consequence, values that best support these ideals become selected as instantiations of the frame attribute. Once again, we see the importance of situated action in category formation and use, where the categories in this case are instantiations of potential frame attributes relevant to planning a specific event.

As a frame's attributes become increasingly established over the course of plan-ning, they increasingly *constrain* the instantiation of other attribute values instan-tiated subsequently. If, for example, a participant decides to vacation in December, this constrains vacation activities to be winter sports and other winter activities (unless the location is instantiated in the southern hemisphere, thereby constrain-ing vacation activities to be summer activities). Table 5 illustrates the diverse sorts of constraints that participants mentioned while planning vacations, where the underlying constraint relations typically took the form of *requires, disallows, ena-bles, prevents, leaves*, and *cooccurs*. Vacationing with a particular companion who works, for example, requires that the companion can get off work. Once again, we see the importance of situated action in category formation and use, where the potential category of instantiations for a frame attribute is constrained by the entire course of situated action being planned. Every possible phase of the situated action cycle, from the setting to possible outcomes, potentially constrains every category of attribute instantiations.

Finally, as optimizations and constraints are initially established for a frame attribute before instantiating it, they combine (via compositional mechanisms) to form an abstraction that covers the potential category of relevant instances. When instantiating the *location* attribute, for example, optimizations for *nearby* and *not too crowded* might be combined with constraints for *in July* (time) and *hiking* (ac-tivity) to form the abstraction, *nearby locations with hiking trails that aren't too crowded in July*. A new ad hoc category emerges from the process of combining a frame attribute with relevant optimizations and constraints. Once the abstraction has been created, it can then be used to guide the process of selecting possible instantiations. Members of the category must first satisfy the constraints, and then become organized along a typicality gradient defined by relevant ideals (originating in the optimizations). Once a potential category exists and has been structured in this manner, it supports selecting instances that will not only satisfy constraints on goal achievement but attempt to optimize it.

5.3   Roles of goal-derived categories in expertise

When an ad hoc category is first created, it is likely to become established in memory, at least weakly. If, however, the same kind of event is planned repeatedly, the category may become increasingly entrenched in memory. If someone regularly planned the same kind of vacation every summer, a category like *nearby locations with hiking trails that aren't too crowded in July* could become well established. Moreover, many other categories associated with the event might similarly become well established, such as *places to stay, things to take, preparations to make*, and so forth. All these increasingly established categories, together with their ideal-based typicality gradients, create expertise, streamlining the planning process on subsequent occasions, hopefully leading to increasingly optimal event outcomes. What were once ad hoc categories become well-established goal-derived categories, with these categories constituting fundamental mechanisms of planning expertise (for this specific kind of event).

Barsalou (1991) further proposed that planning expertise can produce complex *fields* of categories *within* individual frame attributes. To the extent that a particular individual plans many different versions of the same event, different subcategories of possible values develop that support different versions of the event, each subcategory associated with different optimizations and constraints. A travel agent, for example, might develop complex fields of categories for attributes of the vacation frame as a result of frequently planning vacations of many kinds. The *location* attribute, for example, would become associated with many sets of locations that support different kinds of vacations. In a given planning context, the relevant subset would be indexed by first establishing relevant optimizations and constraints for the specific vacation being planned, which then activate the relevant set of attribute values for *vacation location*. With increasing expertise, the ease and precision of activating optimal categories for all relevant frame attributes may make planning relatively easy and effective. In this manner, the continual use of relevant goal-derived categories plays a central role in developing expertise at performing a specific kind of situated action.

5.4   Habitual goal-derived categories and their adaptation

It follows from the above account that every adult has an extensive set of well-established goal-derived categories that support everyday activities they perform habitually. For example, people generally establish goal-derived categories that support the clothes they wear each day (at different times of year; for different occasions), the foods they tend to eat (both at home and eating out), the entertainment

and health activities pursued regularly (during the week and on weekends), the social activities pursued in social networks, and so forth. For any daily activity to become habitual and routinized, it must develop relevant goal-derived categories that efficiently and optimally instantiate relevant frame attributes. As situations for performing the activity arise, the relevant goal-derived categories become active quickly, guiding frame instantiation such that the activity can be performed as effectively and optimally as possible. On deciding what to have for breakfast at home, goal-derived categories for relevant foods, preparation methods, utensils, and so forth become active immediately to support the activity (even before consuming sufficient caffeine).

Rather than being applied rigidly, the goal-derived categories that support habitual activities are often adapted to current individual and situational constraints. When having breakfast, for example, if one is in a hurry or out of a typical breakfast food, selecting instantiations from supporting goal-derived categories is likely to proceed differently than usual. Dual-process frameworks offer a natural cognitive architecture for thinking about how habitual processing and regulatory processing combine when habits are adapted to current contextual constraints (e.g., Chaiken & Trope 1999; Sherman, Gawronski & Trope 2014). Better understanding how goal-derived categories are processed from this perspective offers an important direction for future work.

5.5   Lexicalization of goal-derived categories

As Barsalou (1991) noted, a surprising number of goal-derived categories are lexicalized. For culturally normative activities, lexemes often develop for attributes of the underlying event frames to designate important goal-derived categories. Consider the activity of *eating* and lexicalizations of categories associated with its important semantic roles, such as *diner* (agent), *food* (object), *utensil* (instrument), *eatery* (location), and *breakfast* (time). Similarly, for the activity of *farming*, lexicalizations develop for *farmer* (agent), *crop* (object), and *farm* (location); for *purchasing*, lexicalizations develop for *buyer* (agent), *seller* (indirect object), *merchandise* (object), and *payment* (instrument). Interestingly, some common taxonomic categories appear to be goal-derived categories, such as *clothing* and *tools*, where *clothing* can be viewed as the category of *wear* objects, and *tools* can be viewed as the category of *work* instruments.

Lexicalized goal-derived categories can become contextualized just like their non-lexicalized counterparts. By adding optimizations and constraints to lexicalized categories, abstractions become tailored to current contexts of situated action (Section 5.2). *Food*, for example, can be contextualized to *tasty food to be consumed*

on a diet, and *clothing* can be contextualized to *warm clothing to be worn while hiking in winter*. Barsalou (1991) explores a variety of issues associated lexicalizing and contextualizing goal-derived categories, including their joint use with taxonomic categories during situated action.

6.   Grounding ad hoc and goal-derived categories

We have seen multiple ways that situated action permeates not only goal-derived categories, but taxonomic categories as well. Ideals structure members of both category types. Ad hoc categories depend critically on having a context of situated action present. All types of categories can be optimized and constrained to produce subsets of category members that support specific courses of situated action in particular contexts. All types of categories can become associated with fields of subcategories that provide expertise as the same category becomes embedded in different situational contexts.

6.1   Instantiating space-time regions in simulations

This next section explores another potential connection of goal-derived categories to grounded cognition, illustrating how these categories could emerge naturally from processing simulations that support situated action (Barsalou 1999; also see Barsalou 2003). Although no empirical evidence exists for this grounded account of ad hoc categories, the proposal here is potentially useful because it illustrates how goal-derived categories could originate in simulation – perhaps the most basic cognitive mechanism associated with grounded cognition.

Figure 3 illustrates Barsalou's (1999) proposal. Imagine that an agent wants to change a light bulb that has gone out and must stand on something to reach it. Further imagine that the agent constructs a simulation of changing the light bulb that includes reaching up to unscrew it, followed by inserting a new bulb (Figure 3A). Finally, imagine that the agent doesn't have a specific object in mind to stand on, but simply realizes that an appropriate object will be needed. Because the agent possesses a rough idea of what could be stood upon, the approximate size and position of a potential object is included in the simulation within a relevant space-time region (i.e., a 3D spatial region of the simulation that evolves over some period of time as the simulation runs). Nevertheless, the simulation for changing the light bulb is incomplete because the details of this space-time region have not been specified fully. Although the simulation represents changing the bulb with some precision, it only simulates the object to be stood upon in a sketchy manner (see Barsalou 1999, Section 2.2 for discussion of simulation sketchiness).
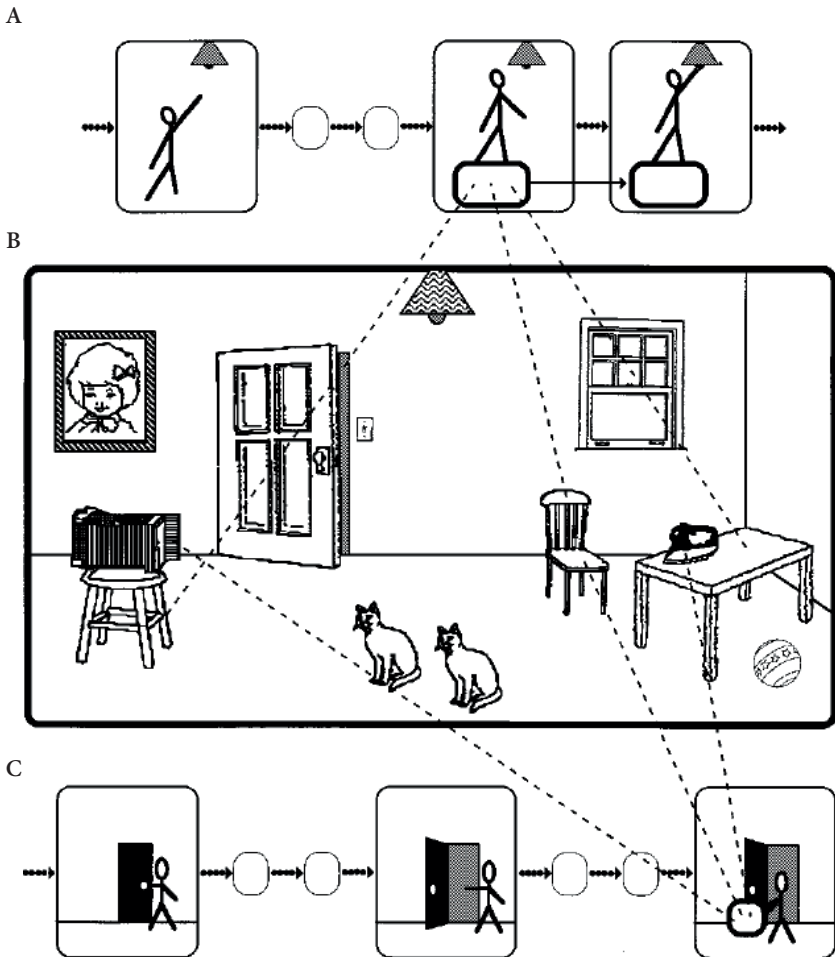
Figure 3. Within the simulation framework of perceptual symbol systems, accounting for the ad hoc categories of *things to stand on to change a light bulb* (A) and *things that could hold a door open* (C), which construe common entities in the same scene differently (B). Reprinted from Figure 9 in Barsalou (1999)

Further imagine, however, that the agent attempts to instantiate this sketchy space-time region of the simulation by simulating potentially relevant objects as instantiations *within it*. Doing so essentially instantiates the instrument attribute of the frame for changing a light bulb by generating relevant values of the goal-derived hoc category, *things to stand on to change a light bulb*. Rather than performing this instantiation process symbolically as in classic frame theories, however, the instantiation is implemented as simulating different possible objects in the relevant space-time region of an action simulation.

Prior to actually simulating objects in this space-time region, the agent may first try to specify relevant optimizations (e.g., *sturdy* and *safe*) and constraints (e.g., *should be located in the home*). Interestingly, both can be implemented via simulation as well. Sturdy and safe can be simulated as specific outcomes of standing on potential instrument objects in the space-time region and comparing their relative amounts of support and subsequent injury. *In the home* can be simulated as searching within a circumscribed region of space.

Using this abstraction for the category, the agent might then examine the environment for possible members. As Figure 3B illustrates, the agent in this particular example finds a stool, a chair, and table that all satisfy the constraint of being in the home. The agent can then simulate using each of these objects to stand on and change the light bulb in the space-time region for the instrument, thereby making the original simulation in Figure 3A more complete. In the process, the simulations of different objects can be compared with respect to the relevant optimizations, perhaps inferring that the chair would be sturdiest and safest. The chair could then be used to actually change the light bulb as situated action proceeds. Figure 3C further illustrates how the same chair could be differently construed in the ad hoc category of *things that could be used to hold a door open*.

This example illustrates how ad hoc and goal-derived categories can be implemented within a simulation architecture. From this perspective, a given ad hoc or goal-derived category contains a disjunctive set of simulations that instantiate the same space-time region of a larger simulation to support a particular form of situated action. By virtue of their relation to other space-time regions in a simulation, each particular space-time region plays a specific semantic role, such as *agent, theme, instrument*, and so forth, implicitly implementing a frame attribute. Although these roles and attributes may remain implicit within the simulation's overall structure, they may often become explicit when described linguistically (e.g., space-time regions for *food* and *utensil* in simulations of *eating*). Additionally, abstractions can develop for these regions (attributes) that include both optimizations and constraints, which again may be represented implicitly in simulations, as well as explicitly in language. Finally, a disjunctive set of simulations that effectively instantiate a space-time region constitutes a goal-derived category to support the situated action being simulated. Language is likely to play important roles throughout the category construction process. In this manner, the simulation architecture naturally and powerfully offers a way of understanding and conceptualizing goal-derived categories from the perspective of grounded cognition.

6.2   The interface between cognition and action

On the one hand, we have seen how people can establish goal-derived categories in memory to support situated action performed habitually (Sections 3.5, 5.3). On the other, we have seen how goal-derived categories can be implemented in space-time regions of simulations for situated action (Section 6.1). As we will see next, these two principles can be combined with the situated action cycle itself (Section 2.6) to develop a proposal of how goal-derived categories provide an interface between cognition and action (Barsalou 2003; see Barsalou 1991, for an earlier "pre-simulation" account).

As we saw for Figure 1D, the situated action cycle begins with the perception of entities and events in the environment, which activate relevant cognitive states of self-relevance associated with goals, values, norms, identities, and so forth. In turn, these cognitions induce bodily states for affect and motivation that produce diverse actions at many levels of behavior, with these actions ultimately causing outcomes in the world, which then initiate another situated action cycle.

The account of goal-derived categories developed in preceding sections can be integrated with the situated action cycle as follows. As cognitive states of self-relevance become active, they activate habitual forms of situated action (referred to as *situated conceptualizations* in related articles; e.g., Barsalou 2003, 2016d, 2016c; Lebois, Wilson-Mendenhall, Simmons, Barrett & Barsalou 2018). As the situated conceptualization for a habitual form of situated action becomes active, it reinstates itself as a *simulation* in relevant brains areas active previously active when performing the relevant situated action (Sections 2.1, 2.2). In the process, relevant objects and events in the environment are simulated, along with actions to be performed, followed by possible outcomes. Most importantly, each simulated object, event, and action occupies a space-time region in the simulation that, more generally, can be instantiated with a well-established goal-derived category having a typicality gradient (Section 6.1). As the simulation unfolds, it is likely to instantiate each of these space-time regions with a simulation of the most typical member, as defined by frequency and ideals. These simulated inferences about likely category members provide inferences about what's likely to be occurring in the world as the situated action is performed (Barsalou 2009; Barsalou et al. 2007). Relevant objects and events are anticipated; potential actions are planned; likely outcomes are predicted. In other words, these simulations attempt to link cognition with the world. Once actual (as opposed to simulated) actions begin to be executed, these simulations run and are used to monitor actions performed in the situated action cycle.

If the habitual activity goes as planned, the overall simulation – including anticipated members of goal-derived categories – creates an interface between the original states of self-relevance that initiated the situated action and the actual actions

and outcomes that follow in the environment. What's simulated in anticipation of situated action maps successfully onto what is actually perceived and done as the event unfolds. Default simulations in the space-time regions for goal-derived categories create accurate anticipations and guide appropriate actions to streamline performance of situated action, providing a successful interface between cognition and action. Cognition and the world become aligned through instantiations of relevant goal-derived categories.

Various factors can decrease the effectiveness of this interface. If the physical environment contains objects and events that aren't anticipated via initial simulations, the respective defaults of the relevant space-time regions must be modified with simulations of less typical category members down the typicality gradient. Similarly, if the agent's goals and the situational constraints are somewhat atypical, again default simulations may not optimally anticipate what needs to be done, therefore requiring revision. Under these conditions, alternative members of relevant goal-derived categories may become active and simulated in their respective space-time regions. If necessary, more laborious search may be conducted in memory or in the environment to find appropriate members. The interface needs to be tuned before cognition and the world are aligned.

In general, situated action cannot run to completion unless all relevant space-time regions for relevant goal-derived categories are instantiated with effective simulations. To the extent that the agent has expertise in the domain, rich fields for all the relevant goal-derived categories may quickly and efficiently index relevant instantiations across most situations (Section 5.3). Only occasionally may significant work be required to find instantiations that bring simulations and the world into alignment. Regardless, a requirement for successful situated action is establishing an interface that aligns cognition and the world, where this interface revolves around simulated instantiations of goal-derived categories.

7.   Outstanding issues

Many proposals made in this chapter about ad hoc and goal-derived categories have rested on relatively preliminary (and sometimes no) evidence, except perhaps for the determinants of typicality gradients, where much literature exists. Clearly, these categories appear to be interesting and important for effective human cognition. Thus, developing more sophisticated accounts, together with empirical evidence that do these accounts justice, has the potential to significantly increase our understanding of human cognition, especially as it plays out in situated action. The following sections highlight a few of the potential proposals made here that could be developed further.

7.1   The concept composition process that creates ad hoc categories

Section 5.1 proposed that ad hoc categories originally develop to instantiate attributes in frames that organize situated action. As agents begin to instantiate these categories, they integrate optimizations and constraints with attributes to create abstractions, which then guide the process of finding and selecting category members. Because this proposal rests on a preliminary set of protocol data, much more ambitious assessments could be made of both the representations and processes assumed in this account of how goal-derived categories originate compositionally. Establishing the roles of language in the composition process will be essential.

7.2   The development of expertise

Several proposals were made about expertise, with none of them supported by empirical data. Section 3.5 suggested that as ad hoc categories are used repeatedly, they become well established in memory. Section 5.3 suggested that as expertise develops, these categories become increasingly well established and optimized (in terms of their graded structures) to streamline expert performance. Section 5.3 further proposed that experts develop fields of attribute values for a goal-derived category, which support efficient performance across different contexts of situated action. Given the potential importance of such processes to expertise, especially in applied areas, examining and developing these proposals has potential for significant impact.

7.3   Use and adaptation of goal-derived categories in habitual behaviors at the interface of cognition and action

Sections 5.4 and 6.2 suggested that goal-derived categories support everyday habits by providing an interface between cognition and action during habit performance. Section 5.4 further suggested that top-down regulation of these habitual categories could play central roles in adapting habits to individual and situational constraints that vary across contexts. When eating daily meals, for example, well-established goal-derived categories associated with frame attributes for eating anticipate foods to be eaten, along with how to prepare and consume them. As contextual factors vary, however, regulatory processes adapt habitually generated categories to best accommodate the current situation (with dual-process theory offering a useful way to think about the adaptation process).

Assessing goal-derived categories in everyday habits may have considerable potential for both understanding health behaviors (e.g., eating) and changing them.

Because these categories appear to operate at the interface of cognition and action, changing their activation during cognitive processing has significant potential for changing subsequent action in the world that emanates from their specific instantiations in anticipatory simulations. Changing the foods that someone anticipates eating at breakfast, for example, could move one's morning food intake in a healthier direction. For these reasons, developing assessments of goal-derived categories in health habits, and further developing interventions to change them, seem like important future goals.

Another possibility in this area would be to use goal-derived categories for assessing individual differences in health domains. By establishing the foods that an individual eats habitually, and then characterizing these foods on various dimensions (e.g., healthiness, tastiness), profiles of individual eating behavior could be established. Similar kinds of profiles based on goal-derived categories could be similarly be established for many other kinds of domains, such as stress, physical activity, sustainability, and so forth.

7.4   Grounding goal-derived categories in space-time regions of simulations

No direct evidence supports the proposal in Section 6.1 that ad hoc and goal-derived categories are implemented as disjunctive sets of possible instantiations of space-time regions (although Wu and Barsalou 2009, offer indirectly related evidence). Nevertheless, this seems like a particularly provocative and potentially significant proposal. From the perspective of grounded cognition, this account offers a powerful implicit mechanism for producing categories (i.e., via variations in what instantiates a space-time region of a simulation). It also establishes a natural interface between cognition and action, as simulations map onto entities and events in the world. Again, establishing the roles of language in this process will be essential.

7.5   Establishing the neural bases of goal-derived categories

As far as I know, we have virtually no understanding yet of how goal-derived categories are implemented in the brain. Presumably frontal control areas are important when ad hoc categories are constructed initially, with these areas falling away as goal-derived categories become well established in memory. If these categories are implemented as instantiations of space-time regions, then a given type of goal-derived category should be implemented in relevant modalities. Goal-derived categories of objects should be implemented in the ventral stream; goal-derived categories of actions should be implemented in the motor system; goal-derived

categories of locations should be implemented in the parahippocampal place area; and so forth. Establishing the neural resources used to implement goal-derived categories should offer useful constraints on cognitive-level accounts of the relevant representations and processes.

7.6   Establishing relations between cognitive and linguistic mechanisms

Perhaps the most exciting new work on ad hoc and goal-derived categories is being performed in linguistics (e.g., Barotto & Mauri 2018; Mauri 2017; Mauri & Sansò 2018). Powerful new techniques are being developed to detect and collect ad hoc categories, goal-derived categories, and related constructions found ubiquitously in linguistic corpora. The resulting data offer important new sources of insight into the structure and use of these categories that can be readily integrated with what we already know. Just the specific categories observed in this new work are provocative in illustrating the kinds of categories that people entertain on the fly during everyday activities. Rather than cherry picking ad hoc and goal-derived categories for specific research purposes, searching linguistic corpora systematically is likely to provide much more representative sampling.

Perhaps more significantly, much potential exists for integrating insights from linguistic work with insights from psychological work to better understand important processes such as abstraction and concept composition (Section 5.1). Another exciting possibility is developing linguistically-informed measures that establish individual differences and behavior change in health habits (Section 7.3). Finally, there is much to learn about how people use language to coordinate cognition through the categories they use during social interaction. Indeed, much of the recent data collected in the linguistic literature illustrates social coordination exquisitely. As is often the case, language provides a unique window onto cognition that offers much potential for developing and testing hypotheses about its underlying mechanisms.

Acknowledgements

References

Anderson, Michael L. 2010. Neural reuse: A fundamental organizational principle of the brain. *Behavioral and Brain Sciences* 33: 245–266. https://doi.org/10.1017/S0140525X10000853

Anderson, Michel L. 2016. Précis of after phrenology: Neural reuse and the interactive brain. *Behavioral and Brain Sciences* 39: 1–22. https://doi.org/10.1017/S0140525X15000631

Aydede, Murat & Robbins, Philip. 2009. *The Cambridge Handbook of Situated Cognition*. Cambridge: CUP.

Baerger, Dana Royce & McAdams, Dan P. 1999. Life story coherence and its relation to psychological well-being. *Narrative Inquiry* 9: 69–96. https://doi.org/10.1075/ni.9.1.05bae

Barotto, Allessandra & Mauri, Caterina. 2018. Constructing lists to construct categories. *Italian Journal of Linguistics* 30: 95–134.

Barsalou, Lawrence W. 1982. Context-independent and context-dependent information in concepts. *Memory & Cognition* 10: 82–93. https://doi.org/10.3758/BF03197629

Barsalou, Lawrence W. 1983. Ad hoc categories. *Memory & Cognition* 11: 211–227. https://doi.org/10.3758/BF03196968

Barsalou, Lawrence W. 1985. Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 11: 629–654.

Barsalou, Lawrence W. 1987. The instability of graded structure: Implications for the nature of concepts. In *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*, Ulric Neisser (ed.), 101–140. Cambridge: CUP.

Barsalou, Lawrence W. 1989. Intraconcept similarity and its implications for interconcept similarity. In *Similarity and Analogical Reasoning*, Stella Vosniadou & Andrew Ortony (eds), 76–121. Cambridge: CUP. https://doi.org/10.1017/CBO9780511529863.006

Barsalou, Lawrence W. 1990. On the indistinguishability of exemplar memory and abstraction in category representation. In *Content and Process Specificity in the Effects of Prior Experiences: Advances in Social Cognition* [Advances in Social Cognition 3], Thomas K. Srull & Robert S. Wyer , Jr. (eds), 61–88. Hillsdale NJ: Lawrence Erlbaum Associates.

Barsalou, Lawrence W. 1991. Deriving categories to achieve goals. In *Psychology of Learning and Motivation. Advances in Research and Technology*, Vol. 27, Gordon H. Bower (ed.), 1–64. New York NY: Academic Press. https://doi.org/10.1016/S0079-7421(08)60120-6

Barsalou, Lawrence W. 1992. Frames, concepts, and conceptual fields. In *Frames, Fields, and Contrasts: New Essays in Semantic and Lexical Organization*, Adrienne Lehrer & Eva F. Kittay (eds), 21–74. Hillsdale NJ: Lawrence Erlbaum Associates.

Barsalou, Lawrence W. 1993. Flexibility, structure, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. In *Theories of Memory*, Alan F. Collins, Susan E. Gathercole, Martin A. Conway & Peter E. Morris (eds), 29–101. Hillsdale NJ: Lawrence Erlbaum Associates.

Barsalou, Lawrence W. 1999. Perceptual symbol systems. *Behavioral and Brain Sciences* 22: 577–660. https://doi.org/10.1017/S0140525X99002149

Barsalou, Lawrence W. 2003. Situated simulation in the human conceptual system. *Language and Cognitive Processes* 18: 513–562. https://doi.org/10.1080/01690960344000026

Barsalou, Lawrence W. 2008. Grounded cognition. *Annual Review of Psychology* 59: 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Barsalou, Lawrence W. 2009. Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364: 1281–1289. https://doi.org/10.1098/rstb.2008.0319

Barsalou, Lawrence W. 2012. The human conceptual system. In *The Cambridge Handbook of Psycholinguistics*, Michael Spivey, Ken McRae & Marc F. Joanisse (eds), 239–258. Cambridge: CUP. https://doi.org/10.1017/CBO9781139029377.013

Barsalou, Lawrence W. 2016a. Can cognition be reduced to action? Processes that mediate stimuli and responses make human action possible. In *Where's the Action? The Pragmatic Turn in Cognitive Science* [Strüngmann Forum Reports 18], Andreas K. Engel, Karl J. Friston & Danica Kragic (eds), 81–96. Cambridge MA: The MIT Press. https://doi.org/10.7551/mitpress/9780262034326.003.0005

Barsalou, Lawrence W. 2016b. On staying grounded and avoiding quixotic dead ends. *Psychonomic Bulletin & Review* 23: 1122–1142. https://doi.org/10.3758/s13423-016-1028-3

Barsalou, Lawrence W. 2016c. Situated conceptualization offers a theoretical account of social priming. *Current Opinion in Psychology* 12: 6–11. https://doi.org/10.1016/j.copsyc.2016.04.009

Barsalou, Lawrence W. 2016d. Situated conceptualization: Theory and applications. In *Perceptual and Emotional Embodiment* [Foundations of Embodied Cognition 1], Yann Coello & Martin H. Fischer (eds), 11–37. Hove: Psychology Press.

Barsalou, Lawrence W. 2017. Cognitively plausible theories of concept composition. In *Compositionality and Concepts in Linguistics and Psychology*, James A. Hampton & Yoad Winter (eds), 9–30. Dordrecht: Springer. https://doi.org/10.1007/978-3-319-45977-6_2

Barsalou, L. W., & Sewell, D. R. (1984). Constructing representations of categories from different points of view. *Emory Cognition Project Technical Report #2*, Emory University.

Barsalou, Lawrence W., Breazeal, Cynthia & Smith, Linda B. 2007. Cognition as coordinated non-cognition. *Cognitive Processing* 8: 79–91. https://doi.org/10.1007/s10339-007-0163-1

Barsalou, Lawrence W. & Hale, Christopher R. 1993. Components of conceptual representation. From feature lists to recursive frames. In *Categories and Concepts: Theoretical Views and Inductive Data Analysis*, Ivan Van Mechelen, James A. Hampton, Ryszard Michalski & Peter Theuns (eds), 97–144. San Diego CA: Academic Press.

Barsalou, Lawrence W., Niedenthal, Paula M., Barbey, Aron K. & Ruppert, Jennifer A. 2003. Social embodiment. In *Psychology of Learning and Motivation*, Vol. 43, Brian H. Ross, 43–92. New York NY: Academic Press.

Barsalou, Lawrence W., Dutriaux Léo & Scheepers, Christoph. 2018. Moving beyond the distinction between concrete and abstract concepts. *Philosophical Transactions of the Royal Society B: Biological Sciences* 373: 20170144. https://doi.org/10.1098/rstb.2017.0144

Borkenau, Peter. 1990. Traits as ideal-based and goal-derived social categories. *Journal of Personality and Social Psychology* 58: 381–396. https://doi.org/10.1037/0022-3514.58.3.381

Bouton, Mark E. & Todd, Travis P. 2014. A fundamental role for context in instrumental learning and extinction. *Behavioural Processes* 104: 13–19. https://doi.org/10.1016/j.beproc.2014.02.012

Brooks, Rodney A. 1991. Intelligence without representation. *Artificial Intelligence* 47: 139–159. https://doi.org/10.1016/0004-3702(91)90053-M

Carey, Susan. 2009. *The Origin of Concepts*. Oxford: OUP. https://doi.org/10.1093/acprof:oso/9780195367638.001.0001

Chaiken, Shelly & Trope, Yaacov. 1999. *Dual-process Theories in Social Psychology*. New York NY: Guilford Press.

Chaplin, William F., John, Oliver P. & Goldberg, Lewis R. 1988. Conceptions of states and traits: Dimensional attributes with ideals as prototypes. *Journal of Personality and Social Psychology* 54: 541–557. https://doi.org/10.1037/0022-3514.54.4.541

Chen, Jing, Papies, Esther K. & Barsalou, Lawrence W. 2016. A core eating network and its modulations underlie diverse eating phenomena. *Brain and Cognition* 110: 20–42. https://doi.org/10.1016/j.bandc.2016.04.004

Chrysikou, Evangelia G. 2006. When shoes become hammers: Goal-derived categorization training enhances problem-solving performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32: 935–942.

Clark, Andy. 1998. *Being There: Putting Brain, Body, and World Together Again*. Cambridge MA: The MIT Press, A Bradford Book.

Clark, Andy. 2008. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford: OUP. https://doi.org/10.1093/acprof:oso/9780195333213.001.0001

Coello, Yann & Fischer, Martin H. (eds). 2016a. *Perceptual and Emotional Embodiment* [Foundations of Embodied Cognition 1]. London: Routledge.

Coello, Yann & Fischer, Martin H. (eds). 2016b. *Conceptual and Interactive Embodiment* [Foundations of Embodied Cognition 2]. London: Routledge.

De Vega, Manuel, Glenberg, Arthur M. & Graesser, Arthur C. (eds). 2008. *Symbols, Embodiment, and Meaning*. Oxford: OUP. https://doi.org/10.1093/acprof:oso/9780199217274.001.0001

Domjan, Michael. 2014. *The Principles of Learning and Behavior*. Independence KY: Cengage Learning.

Donald, Merlin. 1993. Precis of Origins of the modern mind: Three stages in the evolution of culture and cognition. *Behavioral and Brain Sciences* 16: 737–748. https://doi.org/10.1017/S0140525X00032647

Edson Escalas, Jennifer. 2004. Narrative processing: Building consumer connections to brands. *Journal of Consumer Psychology* 14: 168–180. https://doi.org/10.1207/s15327663jcp1401&2_19

Fillmore, Charles J. 1985. Frames and the semantics of understanding. *Quaderni Di Semantica* 6: 222–254.

Fodor, Jerry A. 1975. *The Language of Thought*. Cambridge MA: Harvard University Press.

Fodor, Jerry A. & Pylyshyn, Zenon W. 1988. Connectionism and cognitive architecture: A critical analysis. *Cognition* 28: 3–71. https://doi.org/10.1016/0010-0277(88)90031-5

Gibson, James J. 1966. *The Senses Considered as Perceptual Systems*. Boston MA: Houghton Mifflin.

Gibson, James J. 1979. *The Ecological Approach to Visual Perception*. Boston MA: Houghton Mifflin.

Glenberg, Arthur M. 1997. What memory is for: Creating meaning in the service of action. *Behavioral and Brain Sciences* 20: 41–50. https://doi.org/10.1017/S0140525X97470012

Goldstone, Robert L., Kersten, Alan & Carvalho, Paulo F. 2018. Categorization and concepts. In *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*, John T. Wixted (ed.), 1–43. New York NY: Wiley & Sons. https://doi.org/10.1002/9781119170174.epcn308

Hampton, James A. 1979. Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior* 18: 441–461. https://doi.org/10.1016/S0022-5371(79)90246-9

Hinzen, Wolfram, Machery, Edouard & Werning, Markus (eds). 2012. *The Oxford Handbook of Compositionality*. Oxford: OUP. https://doi.org/10.1093/oxfordhb/9780199541072.001.0001

Hoenig, Klaus, Müller, Cornelia, Herrnberger, Bärbel, Sim, Eun-Jin, Spitzer, Manfred, Ehret, Günter & Kiefer, Markus. 2011. Neuroplasticity of semantic representations for musical instruments in professional musicians. *NeuroImage* 56: 1714–1725. https://doi.org/10.1016/j.neuroimage.2011.02.065

Hsu, Nina S., Frankland, Steven M., & Thompson-Schill, Sharon L. 2012. Chromaticity of color perception and object color knowledge. *Neuropsychologia* 50: 327–333. https://doi.org/10.1016/j.neuropsychologia.2011.12.003

Huettig, Falk, Rommers, Joost & Meyer, Antje S. 2011. Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica* 137: 151–171. https://doi.org/10.1016/j.actpsy.2010.11.003

Hurley, Susan. 2001. Perception and action: Alternative views. *Synthese* 129: 3–40. https://doi.org/10.1023/A:1012643006930

Hutchins, Edwin. 1995. *Cognition in the Wild*. Cambridge MA: The MIT Press.

Kiefer, Markus, Sim, Eun-Jin, Herrnberger, Bärbel, Grothe, Jo & Hoenig, Klaus. 2008. The sound of concepts: Four markers for a link between auditory and conceptual brain systems. *The Journal of Neuroscience* 28: 12224–12230. https://doi.org/10.1523/JNEUROSCI.3579-08.2008

Lebois, Lauren A. M., Wilson-Mendenhall, Christine D., Simmons, W. Kyle, Barrett, Lisa F. & Barsalou, Lawrence W. 2018. Learning situated emotions. *Neuropsychologia* 145: 106637. https://doi.org/10.1016/j.neuropsychologia.2018.01.008

Lebois, Lauren A. M., Wilson-Mendenhall, Christine D. & Barsalou, Lawrence W. 2015. Are automatic conceptual cores the gold standard of semantic processing? The context-dependence of spatial meaning in grounded congruency effects. *Cognitive Science* 39, 1764–1801. https://doi.org/10.1111/cogs.12174

Lewis, James W. 2006. Cortical networks related to human use of tools. *The Neuroscientist* 12: 211–231. https://doi.org/10.1177/1073858406288327

Lewis, James W. 2010. Audio-visual perception of everyday natural objects – Hemodynamic studies in humans. In *Multisensory Object Perception in the Primate Brain*, Jochen Kaiser & Marcus Johannes Naumer (eds), 155–190. New York NY: Springer. https://doi.org/10.1007/978-1-4419-5615-6_10

Loken, Barbara & Ward, James. 1990. Alternative approaches to understanding the determinants of typicality. *Journal of Consumer Research* 17: 111–126. https://doi.org/10.1086/208542

Lynch, Elizabeth B., Coley, John D. & Medin, Douglas L. 2000. Tall is typical: Central tendency, ideal dimensions, and graded category structure among tree experts and novices. *Memory & Cognition* 28: 41–50. https://doi.org/10.3758/BF03211575

Malt, Barbara C. 1995. Category coherence in cross-cultural perspective. *Cognitive Psychology* 29: 85–148. https://doi.org/10.1006/cogp.1995.1013

Martin, Alex. 2016. GRAPES—Grounding representations in action, perception, and emotion systems: How object properties and categories are represented in the human brain. *Psychonomic Bulletin & Review* 23: 979–990. https://doi.org/10.3758/s13423-015-0842-3

Mauri, Caterina. 2017. Building and interpreting ad hoc categories: A linguistic analysis. In *Formal Models in the Study of Language*, Joanna Blochowiak, Cristina Grisot, Stephanie Durrleman-Tame & Christopher Laenzlinger (eds), 297–326. Berlin: Springer. https://doi.org/10.1007/978-3-319-48832-5_16

Mauri, Caterina & Sansò, Andrea. 2018. Linguistic strategies for ad hoc categorization: Theoretical assessment and cross-linguistic variation. In *Linguistic Strategies for the Construction of Ad Hoc Categories: Synchronic and Diachronic Perspectives*, Caterina Mauri & Andrea Sansò (eds). Special issue of *Folia Linguistica* 52(s39-1): 1–35. https://doi.org/10.1515/flih-2018-0001

McRae, Ken & Jones, Michael N. 2013. Semantic memory. In *The Oxford Handbook of Cognitive Psychology*, Daniel Reisberg (ed.), 206–219. Oxford: OUP.

Miller, George A., Galanter, Eugene & Pribram, Karl H. 1960. *Plans and the Structure of Behavior*. New York NY: Holt Reinhart & Winston. https://doi.org/10.1037/10039-000

Murphy, Gregory L. 2002. *The Big Book of Concepts*. Cambridge MA: The MIT Press. https://doi.org/10.7551/mitpress/1602.001.0001

Murphy, Gregory & Medin, Douglas L. 1985. The role of theories in conceptual coherence. *Psychological Review* 92: 289–316. https://doi.org/10.1037/0033-295X.92.3.289

Newell, Alan & Simon, Herbert A. 1972. *Human Problem Solving*. Englewood Cliffs NJ: Prentice-Hall.

Newen, Albert, Bruin, Leon D. & Gallagher, Shaun (eds). 2018. *The Oxford Handbook of 4E Cognition*. Oxford: OUP. https://doi.org/10.1093/oxfordhb/9780198735410.001.0001

Niedenthal, Paula M., Barsalou, Lawrence W., Winkielman, Piotr, Krauth-Gruber, Silvia & Ric, François. 2005. Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review* 9: 184–211. https://doi.org/10.1207/s15327957pspr0903_1

Papies, Esther K. & Barsalou, Lawrence W. 2015. Grounding desire and motivated behavior: A theoretical framework and review of empirical evidence. In *The Psychology of Desire*, Wilhelm Hofmann & Loran F. Nordgren (eds), 36–60. New York NY: Guilford Press.

Pecher, Diane & Zwaan, Rolf A. (eds). 2005. *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511499968

Pulvermüller, Friedemann. 1999. Words in the brain's language. *Behavioral and Brain Sciences* 22: 253–279. https://doi.org/10.1017/S0140525X9900182X

Pulvermüller, Friedemann. 2005. Brain mechanisms linking language and action. *Nature Reviews Neuroscience* 6: 576–582. https://doi.org/10.1038/nrn1706

Pulvermüller, Friedeman, Moseley, Rachel L., Egorova, Natalia, Shebani, Zubaida & Boulenger, Véronique. 2014. Motor cognition–motor semantics: Action perception theory of cognition and communication. *Neuropsychologia* 55: 71–84. https://doi.org/10.1016/j.neuropsychologia.2013.12.002

Pylyshyn, Zenon W. 1984. *Computation and Cognition. Toward a Foundation of Cognitive Science*. Cambridge MA: The MIT Press, A Bradford Book.

Read, Stephen J., Jones, David K. & Miller, Lynn C. 1990. Traits as goal-based categories: The importance of goals in the coherence of dispositional categories. *Journal of Personality and Social Psychology* 58: 1048–1061. https://doi.org/10.1037/0022-3514.58.6.1048

Reese, Elaine, Haden, Catherine A., Baker-Ward, Lynne, Bauer, Patricia, Fivush, Robyn & Ornstein, Peter A. 2011. Coherence of personal narratives across the lifespan: A multidimensional model and coding method. *Journal of Cognition and Development* 12: 424–462. https://doi.org/10.1080/15248372.2011.587854

Rosch, Eleanor & Mervis, Carolyn B. 1975. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology* 7: 573–605. https://doi.org/10.1016/0010-0285(75)90024-9

Rosch, Eleanor, Mervis, Carolyn B., Gray, Wayne D., Johnson, David M. & Boyes-Braem, Penny. 1976. Basic objects in natural categories. *Cognitive Psychology* 8: 382–439. https://doi.org/10.1016/0010-0285(76)90013-X

Sherman, Jeffrey W., Gawronski, Bertram & Trope, Yaacov. 2014. *Dual-process Theories of the Social Mind*. New York NY: Guilford Publications.

Shtyrov, Yury, Butorina, Anna, Nikolaeva, Anastasia & Stroganova, Tatiana. 2014. Automatic ultrarapid activation and inhibition of cortical motor systems in spoken word comprehension. *Proceedings of the National Academy of Sciences* E1918–E1923. https://doi.org/10.1073/pnas.1323158111

Simmons, W. Kyle, Martin, Alex & Barsalou, Lawrence W. 2005. Pictures of appetizing foods activate gustatory cortices for taste and reward. *Cerebral Cortex* 15: 1602–1608. https://doi.org/10.1093/cercor/bhi038

Simmons, W. Kyle, Ramjee, Vimal, Beauchamp, Michael S., McRae, Ken, Martin, Alex & Barsalou, Lawrence W. 2007. A common neural substrate for perceiving and knowing about color. *Neuropsychologia* 45: 2802–2810. https://doi.org/10.1016/j.neuropsychologia.2007.05.002

Smith, Edward E. & Medin, Douglas L. 1981. *Categories and Concepts*. Cambridge MA: Harvard University Press. https://doi.org/10.4159/harvard.9780674866270

Stein, Nancy L., & Hernandez, Marc W. 2007. Assessing understanding and appraisal during emotional experience. In *Handbook of Emotion Elicitation and Assessment*, Jim A. Coan & John J. B. Allen (eds), 298–317. Oxford: OUP.

Tanenhaus, Michael K., Spivey-Knowlton, Michael J., Eberhard, Kathleen M. & Sedivy, Julie C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science* 268: 1632–1634. https://doi.org/10.1126/science.7777863

Tomasello, Michael. 2009. *The Cultural Origins of Human Cognition*. Cambridge MA: Harvard University Press. https://doi.org/10.2307/j.ctvjsf4jc

Tomasello, Michael, Kruger, Anne C. & Ratner, Hilary H. 1993. Cultural learning. *Behavioral and Brain Sciences* 16: 495–511. https://doi.org/10.1017/S0140525X0003123X

Trumpp, Natalie M., Kliese, Daniel, Hoenig, Klaus, Haarmeier, Thomas & Kiefer, Markus. 2013. Losing the sound of concepts: Damage to auditory association cortex impairs the processing of sound-related concepts. *Cortex* 49: 474–486. https://doi.org/10.1016/j.cortex.2012.02.002

Vallee-Tourangeau, Frederic, Anthony, Susan H. & Austin, Neville G. 1998. Strategies for generating multiple instances of common and ad hoc categories. *Memory* 6: 555–592. https://doi.org/10.1080/741943085

van der Laan, Laura Nynke, de Ridder, Denise, T. D., Viergever, Max A. & Smeets, Paul A. M. 2011. The first taste is always with the eyes: A meta-analysis on the neural correlates of processing visual food cues. *NeuroImage* 55: 296–303. https://doi.org/10.1016/j.neuroimage.2010.11.055

Voorspoels, Wouter, Storms, Gert & Vanpaemel, Wolf. 2013. Idealness and similarity in goal-derived categories: A computational examination. *Memory & Cognition* 41: 312–327. https://doi.org/10.3758/s13421-012-0252-y

Voorspoels, Wouter, Vanpaemel, Wolf & Storms, Gert. 2010. Ideal representations in a similarity space. *COGNITION IN FLUX* 32: 2290–2295. Mahwah NJ: Lawrence Erlbaum Associates.

Voorspoels, Wouter, Vanpaemel, Wolf, & Storms, Gert. 2011. A formal ideal-based account of typicality. *Psychonomic Bulletin & Review* 18: 1006–1014. https://doi.org/10.3758/s13423-011-0122-9

Wang, Xiaoying, Han, Zaizhu, He, Yong, Caramazza, Afonso, Song, Luping & Bi, Yanchao. 2013. Where color rests: Spontaneous brain activity of bilateral fusiform and lingual regions predicts object color knowledge performance. *NeuroImage* 76: 252–263. https://doi.org/10.1016/j.neuroimage.2013.03.010

Winter, Yoad & Hampton, James A. (eds). 2017. *Compositionality and concepts in Linguistics and Psychology*. Berlin: Springer.

Wu, Ling-ling & Barsalou, Lawrence W. 2009. Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica* 132: 173–189. https://doi.org/10.1016/j.actpsy.2009.02.002

Yao, Bo, Belin, Pascal & Scheepers, Christoph. 2011. Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex. *Journal of Cognitive Neuroscience* 23: 3146–3152. https://doi.org/10.1162/jocn_a_00022

Yao, Bo, Belin, Pascal & Scheepers, Christoph. 2012. Brain 'talks over' boring quotes: Top-down activation of voice-selective areas while listening to monotonous direct speech quotations. *NeuroImage* 60: 1832–1842. https://doi.org/10.1016/j.neuroimage.2012.01.111

CHAPTER 4

Category-building lists between grammar and interaction

A constructionist view

Eugenio Goria and Francesca Masini
University of Turin / University of Bologna

Lists are one of the most common devices that are used in interaction to refer to a category. Yet, there are only few studies that analyze the relationship between lists and categorization. Our paper aims at advancing our knowledge of this relationship, and of lists in general. From a theoretical point of view, we discuss the benefits of integrating the Construction Grammar approach to lists adopted in Masini et al. (2018) with some of the basic assumptions of Interactional Linguistics. From an empirical point of view, we offer a qualitative analysis of lists based on data from two corpora of spoken Italian: the LIP corpus (De Mauro et al. 1993) and the KIParla corpus (Mauri et al. 2019a). In particular, we discuss a case study on the use of the Italian discourse marker *insomma* within list constructions: while it serves as a reformulation marker in most of its uses, *insomma* also proves to be used (more marginally) as a category introducer within category-building lists. Our findings provide useful insight to ultimately bridge the gap between denotation lists as a reference-oriented phenomenon and other types of mechanisms that are relevant at the discourse level, including conversational repair.

Keywords: list, categorization, construction grammar, interactional linguistics, spoken Italian, reformulation, discourse marker

1.   Introduction

In the present contribution, we focus on a device for creating categories – especially contextually relevant categories or "ad hoc categories" (Barsalou 1983; Mauri this volume) – that lies in between grammar and discourse, namely "lists" as defined in Masini et al. (2018). Under this view, the object "list" is a highly abstract linguistic pattern that encompasses a number of more specific phenomena that are

traditionally ascribed to different domains of analysis and thus treated separately, such as coordination, repetition, reduplication, co-compounding, reformulation and disfluency. In this sense, the term is reminiscent of the notion of "retraction" as used by Auer and Pfänder (2007).

Besides illustrating the nature of lists and their relevance for categorization (Sections 1.1 and 1.2), we discuss their formal and functional properties as "constructions" in the sense of Construction Grammar (Section 2.1) and analyze their interactional properties in two corpora of Spoken Italian, the LIP corpus (De Mauro et al. 1993) and the KIParla corpus (Mauri et al. 2019a) (Section 2.2). We ultimately embrace a view that aims at bridging the gap between these two perspectives, that is, between grammar and interaction or discourse (Section 2.3). This bridging perspective is best exemplified by a case-study on the Italian polyfunctional discourse particle *insomma* (Section 3.1), which, besides its most common use as a reformulating and concluding marker (see Waltereit 2006), proves to be used also as a category introducer within category-building lists (Section 3.2). Corpus data show that the latter use is likely an extension of the former, which proves the connection between the formulation level and the denotation level, as well as the need for a new dedicated position within the list structure, which was previously disregarded in the literature on lists (Section 3.3).

1.1   What is a list?

According to Masini et al. (2018: 50), lists are characterized by the "syntagmatic concatenation of two or more units of the same type (i.e. potentially paradigmatically connected) that are on a par with each other, thus filling one and the same slot within the larger construction they are part of". This loose definition captures structurally different kinds of lists: most notably, the definition does not specify the nature of the conjuncts, which may be linguistic units of variable size and complexity that entertain a paradigmatic relation. These linguistic units therefore represent the linear, or syntagmatic, realization of two or more paradigmatic alternatives (see Blanche-Benveniste 1990). In order to capture the variability of lists, Masini et al. (2018) propose the abstract and flexible structure in Figure 1, that represents the 'skeleton' onto which the listing phenomena are mapped.

The minimal list is made of two conjuncts. The other components may or may not be expressed: of course, we may have more than two conjuncts; we may or may not have connectives that keep the conjuncts together; we may have a "list completer", like general extenders (e.g. *and the like, etcetera*), which indicate "additional members of a list, set, or category [and combine] with a named exemplar (or exemplars)" (Overstreet 1999: 11); then, around the list we may have a "projection

PRO-C	projection component
LI	list introducer

<p style="text-align:center">INSERTIONS</p>

X_1	**conjunct 1**
CO	coordinator / connective
X_2	conjunct 2
CO	coordinator / connective
X_3	conjunct 3
…	…
CO	coordinator / connective
X_{LAST}	**conjunct last**
LC	list completer

POST-C	post-detailing component

Figure 1. List skeleton

component", i.e. a "more-to-come" element that is then detailed or expanded by the list, and a "post-detailing component", which completes "the structure around the list and at the same time tying the list back to the ongoing topic or activity" (Selting 2007: 523). Finally, insertions of different kinds (discourse markers, hesitations, etc.) may interrupt the list, especially lists in spoken discourse.[1] Let us take the English example[2] in (1): this would map onto the list structure in Figure 1 as illustrated in Figure 2.

---

1.   The prosody of the list is left implicit in the skeleton in Figure 1, which focuses on the morphosyntax of list structures (which is common to lists in spoken and written language). However, prosody is obviously one of the key features of lists in spoken language, as also emphasized by Masini et al. (2018), who mention the difference between prosodically open and prosodically closed lists (Selting 2007); see also Matalon (this volume) and Section 3.1.

2.   The examples used for this article were extracted from a number of English and Italian corpora. For English we used the Santa Barbara Corpus of Spoken American English (Du Bois et al. 2000–2005; SBC henceforth) and the enTenTen15 corpus (Masaryk University, NLP Centre, 2011) accessed through the SketchEngine interface (Kilgarriff et al. 2014). For Italian we used the LIP corpus (De Mauro et al. 1993) and the KIParla corpus (Mauri et al. 2019a); the latter is alternatively referred to as KIP, in order to refer only to the first module of this resource, namely the one involving academic talk. This has been available online since September 2019. See Section 3 for a more detailed description of these two resources. The examples are reported in different formats according to their purpose in the discussion. In some cases, we use an orthographic transcription, in other cases, we need to represent specific features of spoken language, as well as the temporal organization of specific conversational activities. For this reason, we present the latter examples following the conventions given in Jefferson (2004), that are widespread in Conversation Analysis. In order to protect the identity of the participants, the original names have been replaced with invented ones.

(1)   *it's great to fill up on seasonal bounties that may include **fresh peaches, melons, apples, pears and the like*** [enTenTen15]

			it's great to fill up on
	PRO-C	projection component	*seasonal bounties*
	LI	list introducer	*that may include*
	X_1	**conjunct 1**	***fresh peaches***
	CO	coordinator / connective	–
	X_2	conjunct 2	***melons***
	CO	coordinator / connective	–
	X_3	conjunct 3	***apples***
	…	…	–
	CO	coordinator / connective	–
	X_{LAST}	**conjunct last**	***pears***
	LC	list completer	***and the like***
	POST-C	post-detailing component	–

Figure 2. Example (1) mapped onto the list skeleton in Figure 1

As we can see, some "positions" of the skeleton are filled by linguistic material, whereas other positions are left unexpressed.

Consider that the skeleton in Figure 1 is a pre-theoretical apparatus that does not define the nature of the conjuncts, nor the constraints and/or requirements a specific (kind of) list may be subject to. Indeed, lists may have not only variable size and complexity, but also different degrees of conventionalization. Example (1) is an ephemeral creation of the speaker in a given context, namely a specific instance of a coordination which is not retained in memory. This is even more true of discourse-related patterns (also ascribable to lists as defined here) such as reformulations and repairs:

(2)   *A new cast of **thirty, oops I mean seventeen** year olds* [enTenTen13]

However, some lists are lexically fixed and stored as a stable part of our grammar: take for instance irreversible binomials, like (3) (Masini 2006), or co-compounds, like (4) (Wälchli 2005).

(3)   Italian   *coltello e forchetta* lit. knife and fork 'cutlery'
(4)   Chuvash *sĕt-śu*   lit. milk-butter 'dairy products'

Obviously, the latter types of lists are subject to different restrictions and constraints with respect to lists that are productively created in discourse (see also Masini & Arcodia 2018). The attempt here described to keep all these phenomena together,

despite the great variation we face, has the advantage of highlighting the similarities of form and function between these different manifestations at a more abstract level, which might unveil interesting connections and regularities in both intra- and inter-linguistic terms (see also Section 2.1).

Given the loose definition of "list" provided above, and the consequent wide range of formal variation, we also expect a variety of functions performed by lists, depending on their exact manifestation, and this is indeed the case.

Functionally speaking, Masini et al. (2018: 64) regard the list as a device that performs some kind of semantic operation over its conjuncts. Depending on the kind of semantic operation performed, one gets different kinds of lists, which can be conceptually classified into two major classes: formulation and denotation lists.

Formulation lists operate over utterances, hence their function can be regarded as metalinguistic from a discourse perspective: the conjuncts of the list correspond to different attempts of formulating one and the same concept; these attempts are structurally equivalent to each other in that context and correspond to the different types of repair that may occur in conversation. Consider (2) above: what the speaker actually does, after proposing a first formulation, is scanning a paradigm for possible better alternatives and finally choosing the most appropriate. Hesitations and disfluency phenomena, as in (5) below, also rely on a formulation listing pattern in our view (see also Auer & Pfänder 2007).

(5)   PATTY: *and she didn't care,.. the- to b- – to – she didn't care about emancipation*.   [SBC023, Du Bois et al. (2005)]

Denotation lists create a complex denotation by exploiting the denotative meanings of the conjuncts. This may occur in a basically compositional way, in those cases where the intended meaning results from the sum of the single denotational meanings of the conjuncts; however, in a vast majority of cases, the felicitous interpretation of a denotation list strongly depends on inferential reasoning. Let us compare (6) with (1), two similar lists leading to different interpretations:

(6)   *I like fresh peaches, melons, apples and pears*   [constructed example]

In (6), inference plays a minor role, if anything, while in (1) above, even if the same conjuncts are used, the addressee is called to a much greater inferential effort. The general extender *and the like* in (1) hints at other possible items that are not explicitly mentioned, but evoked; it is on their identification, and on the selection of the appropriate semantic operation, that the interpretation of the list depends. Note that this process is not straightforward: the list in (1) might easily refer to 'fruit' in general, but instead refers to a more specific category, namely 'seasonal bounties', as revealed by the projection component.

Denotation lists may have different functions that range from the typical values of syntactic coordination, namely addition (like in (1)), contrast and alternativity (see Haspelmath 2007 and Mauri 2008), to less compositional functions such as generalization, intensification, approximation, and of course categorization (cf. Masini et al. 2018). The latter is the focus of the present article and is analyzed in detail in the following subsection.

1.2   Lists and categorization

Based on the previous discussion, every instance of a denotation list corresponds, in a way, to an act of categorization, as all lists pragmatically presuppose that all their listed items belong to the same set (Barotto & Mauri 2018). However, in our model we distinguish between lists that implicitly rely on some presupposed category and lists whose primary function is to convey at the content level some specific way of categorizing reality in a specific context. For the latter type we use the term "categorizing" or "category-building" lists.

Like other types, such lists may have different size and complexity (see e.g. the list in (7), containing a high number of conjuncts), different degrees of convention-alization (compare, again, the list in (1) with those in (3)–(4)), and may be constructed monologically or dialogically (e.g. (8), where both speakers contribute to the expression of the category 'saving endangered animals'; see also Section 2.2.3).

(7)   *These are the stuff of physics.* **Chimps and dogs and bats and cockroaches and people and worms and dandelions and bacteria and galactic aliens** *are the stuff of biology.*   [Google]

(8)   ROY: *saving the whale,*
                    or saving uh … the .. polar bea[*r,*
        PETE: [**Right… Pandas**],
        ROY: *or making sure there's enough*] *grizzly bears,*
                    that's fine.   [SBC: 003, Du Bois et al. (2000–2005)]

In most cases, categorizing lists are semantically non-exhaustive, which means that only some members of the category that is being constructed are explicitly verbalized. The participants are thus invited to make inferences about the identity of possible further members belonging to the same set. The list in (7), for instance, contains a number of entities studied by biologists. However, we hardly interpret this list as exhaustive: we expect racoons too (just to name one more member of the category 'forms of life'), to be *the stuff of biology*, even though they are not mentioned explicitly in the list.

Non-exhaustive lists, hence, evoke more than what is just said. Following Mauri (this volume), they express an eminently pragmatic function in that they invite

inferencing concerning the property upon which the set has to be built: items occurring in this structure are presented as exemplars of a broader set, whose extension remains unspecified. Crucially, this type of list does not necessarily include elements that are inherently, i.e. semantically, related to each other. On the contrary, non-exhaustive lists are often dependent, for their felicitous interpretation, on the context in which they are used. For this reason, if we adopt Barsalou's distinction between 'common' and 'ad hoc' categories (see Barsalou this volume and previous works), we can argue that this type of lists, more specifically, builds categories in an ad hoc fashion, that is, relying on contextual information and on contextualization cues provided by the speaker in order to direct their interlocutor towards the intended interpretation. Take for instance (9):

(9)   *I need **flour, milk, yeast and so on**.*   [Mauri (2017: 302)]

This list requires contextual information to be interpreted: 'ingredients for a cake' seems to be a quite possible option, but, depending on context and on the information shared by the speakers, other (even extravagant) options might turn out to be the intended ones (e.g. 'the stuff mum always buys on Tuesday afternoon'). Therefore, we can conclude that lists of this type represent an important resource through which speakers verbalize their ways of categorizing reality based on specific contexts, as opposed to using a pre-established set of categories stored in cognition (see Edwards 1991 and Goria 2020). This aspect will be analyzed in detail in Section 2.

2.   Lists between grammar and discourse

In this section we illustrate two different views on lists, which in our opinion are largely compatible and can benefit from each other. The first perspective comes from Construction Grammar (Section 2.1), the second from Interactional Linguistics (Section 2.2). Their fruitful interaction is advocated for in Section 2.3.

2.1   Lists as constructions

Previous accounts have proposed to treat (at least a subset of) list structures as "constructions" in the sense of Construction Grammar (cf., among others, Fillmore et al. 1988; Goldberg 1995, 2006; Croft 2001; Östman & Fried 2005; Hoffmann & Trousdale 2013; Hilpert 2019), that is, as conventionalized form-meaning pairings (cf. Masini & Pietrandrea 2010; Masini et al. 2018; Bonvino et al. 2010).

A constructional analysis of lists relies on the observation that some types of lists – especially those with a non-compositional meaning – display unique and

consistent correspondences between form and function, a situation which lends itself to be analysed in constructional terms.

Masini et al. (2018) propose to regard various kinds of denotation lists, including categorizing ones (see Section 1.2), as constructions licensed (via an instance inheritance link) by a maximally abstract list construction with the very schematic formal and functional properties described in (10).[3]

(10)   ABSTRACT LIST CONSTRUCTION
       Form: ([PRO-C]) ([LI]) {X_1 | (^CO) (X_2) | (^CO) (X_3) | … | (^CO) X_{LAST} | (LC)} ([POST-C])
       Function: 'function f over the set of Xs + presupposition p = common categorization underlying Xs'

Let us start from the 'function' side. As we can see, the meaning of the abstract list construction corresponds to an underspecified function f over the set of conjuncts, which are presupposed to belong to an underlying common category (the presupposition being a stable part in the functional side of the list construction). This general meaning becomes more specific in the daughter constructions licensed by the abstract list construction. For what concerns the categorizing list construction, on which we concentrate in the present chapter, its function f is to create a superordinate, higher-level, category starting from the enumeration of some exemplars of that category. Hence, categorization here is not implicit, as is the case with other types of lists, but rather it is present at the content level and represents the primary communicative goal.

As for the formal side of list constructions, what is crucial is that, besides being schematic, they are also "flexible". This allows to do justice to both the variety of forms lists may display and, at the same time, their unity in terms of core properties. For instance, categorizing lists are typically associated with some properties, such as the fact that conjuncts should be co-hyponyms.[4] So, this property will hold for all

---

3.   In this notation, braces '{ }' delimit the list; the pipe sign '|' separates the conjuncts; round brackets '()' indicate optionality; the circumflex accent '^' marks list markers, namely coordinators/connectives and list completers; square brackets [] enclose the list surroundings, namely projection components (including list introducers) and post-detailing components. Remember that insertions may intervene at virtually any point of the list structure.

4.   An anonymous reviewer suggests that the conjuncts of categorizing lists may also be meronyms. This is probably true, as an Italian irreversible binomial like *giacca e cravatta* (lit. jacket and tie) 'formal suit' seems to suggest. However, the data in our possession so far point to co-hyponymy as the core relation in this type of lists. Moreover, we believe a deeper investigation is needed to understand if lists of meronyms are really equal to lists of co-hyponyms. They are, in the sense that both verbalize lower-level items in order to refer to a broader concept. However, the fact that the lower-level items entertain a different relation with this broader concept in the

categorizing lists, independently of their actual formal realization. See for instance the following examples from Italian:

(11) a. *coltello e forchetta*   lit. knife and fork   'cutlery'
     b. *colazione pranzo e cena* lit. breakfast lunch and dinner 'main meals'

(12) a. *se potessi comprarmi una casa a torino me la comprerei in vanchiglia senza dubbio perche' // eh e' vicinissima al centro però i prezzi delle cose tipo **supermercati cibo eccetera** // non è esagerato*   [KIP, TOD2003]
     'if I could buy a house in Torino I would surely buy it in Vanchiglia, because it is very close to the center but the prices of things like supermarkets, food etcetera, is [sic] not overpriced'.
     b. *tutto quello che riguarda l'apparato genitale_ **disfunzioni malformazioni eccetera** fanno sempre parte cosi' di un campo_ su cui c'e' molta_ eh reticenza a parlare*   [LIP, MC12]
     'everything that concerns the genital apparatus, malfunctions, malformations etcetera are always part of a field of which there is much reluctance to speak'.

In (11) we find the already mentioned irreversible binomial *coltello e forchetta* (cf. (3), Section 1.1) and an irreversible trinomial, i.e. fixed expressions which are stored in our mental lexicon and can be analyzed as lexically specified constructions. Examples in (12), instead, illustrate two lists in spoken Italian, which are obviously not retained in memory, but are rather ephemeral creations that serve a specific purpose: in the first case *supermercati cibo eccetera* identifies the category 'staple necessity', whereas in the second case *disfunzioni malformazioni eccetera* identifies the category 'health problems'. These two lists are created by picking two (supposedly representative or relevant) members of these categories, which are however not mandatory nor fixed: in both cases, we might use another order of the same conjuncts (*cibo supermercati eccetera*), or different lexical items (*alimentari roba da mangiare eccetera*, lit. groceries things to eat etcetera), or more lexical items (*disfunzioni malformazioni difetti eccetera*, lit. malfunctions malformations defects etcetera), and we would still retain the same categorizing effect. Despite these differences, namely the different constraints to which they are subject, the

---

two cases ('type of' vs. 'part of') may have consequences on interpretation. Whereas hyponyms occur as exemplars of a higher-level category, meronyms are not, being rather distinctive features of the intended concept. In this respect, lists of meronyms are reminiscent of what Masini and Arcodia (2018) call "frame-naming" lists, namely lists made of frame-related conjuncts (not necessarily lexico-semantically related) that depict a frame by just mentioning its most salient parts with (what they call) an "impressionistic" technique (e.g., *Chris era il tipico giovanottone inglese tutto pub, sport e fidanzata* 'Chris was the typical English young man devoted to pub, sport and girlfriend (lit. all pub, sport and girlfriend)').

expressions in (11) and (12) still share some core properties which identify them as categorizing lists.

In order to account for this variation, Masini et al. (2018) propose to regard the pattern introduced in Figure 1 above (Section 2.1) as the 'form' of the abstract list construction, which in (10) is displayed as "linearized": this pattern can in principle accommodate simple and complex lists (including those produced in spoken language, which might contain interruptions, etc.), "fixed" lists as well as freely created lists. The abstract list construction then develops its own constructional network, via instance inheritance links, being instantiated by more specified (but still schematic) list constructions (like the categorizing list construction), which in turn may be instantiated either by even more specified, and lexically fixed, list constructions (like (11)) or by constructs to be created in spontaneous speech (like (12)).

Some online-created lists display recurrent characteristics that might lead us to consider them as instances of separate (more specified, albeit still schematic) constructions. For example, a subtype of non-exhaustive, categorizing list in Italian involves the use of semantically weak forms, especially generic nouns, in the X_{LAST} position. Consider for Example (13):

(13)   [p]erche' comunque sia la professoressa rossi che la professoressa verdi (.) sia la mia scuola della lingua dei segni qua a bologna mi possono dare una mano quindi comunque, **materiali bibliografie co:se**,   [KIP, BOA1009]
'for anyway both professor Rossi and professor Verdi, and also my sign language school here in Bologna can help me, so, anyway, **materials, bibliographies, stuff**.'

Here the speaker produces a list in order to provide examples of the help they can get for a research work. In this case, the X_{LAST} element of the list is a generic noun (*cose* 'things'), which basically figures as a dummy element devoid of a specific denotational meaning in the given context: its function is only to signal that the list is unfinished, and hence non-exhaustive. It is worth observing that in such cases non-exhaustivity is not an inherent feature of the noun *cose* (*versus* general extenders like *eccetera*[5] in (12) above), but rather emerges from the whole pattern in which this element is embedded.[6] Therefore, we are in the presence of a subtype

---

5.   On Italian *eccetera* and its diachronic development, see Fiorentini & Magni (this volume).

6.   An anonymous reviewer observes that *cose* 'things' is not on a par with preceding conjuncts: it rather has an anaphoric function with respect to them, being some sort of "grammaticalized" version of more complex general extenders of the *(and) things like that* type. Although this might turn out to be the case, there are reasons to believe that *cose* is not a proper general extender (yet), as can be evinced, for example, from its distributional behavior: *cose* is morphosyntactically coherent with the preceding conjuncts, i.e. it occurs with lists of plural bare nouns and very rarely with lists involving other parts of speech, while full-fledged general extenders do not appear to have this constraint.

of list construction that is characterized by greater specification in terms of both meaning (i.e. the expression of non-exhaustivity) and form, since one of its syntactic slots (X_{LAST}) is bound to be filled by a dummy element like *cose* (or other items, according to the nature of the exemplars involved; see Mauri et al. 2019b for further examples and discussion).

To conclude, the flexible nature of the list pattern, with components that are obligatorily realized (as mentioned above, the "minimal list" is made of two conjuncts) and many others that are optional (and whose presence may give rise to a number of subconstructions with different properties), is one of the main advantages of this approach, which unifies under one and the same abstract mechanism expressions that are traditionally treated as belonging to different levels of analysis. In their contribution, Masini et al. (2018) propose a constructional analysis of denotation lists only, although they advocate for a similar account for some types of formulation lists: whereas disfluency lists can hardly be regarded as constructions (namely, they are list patterns, but not "list constructions" in the proper sense), reformulation lists are more likely to be analyzable in constructionist terms, as also argued by Kahane and Pietrandrea (2012) and Bonvino et al. (2010).

2.2   Lists in interaction

The construction-based account of lists outlined in the previous section provides a powerful tool for the description of lists, especially because, as argued before, it allows us to provide an accurate representation of what is shared by *all* lists in terms of form and function. Under this view, the representation of the list pattern is permanently stored in grammar and used for different purposes. However, under this account, it remains uncertain how to identify the boundaries between denotation lists and other related constructions occurring in spoken discourse, in particular with reformulation. While Masini et al. (2018) adopt a discrete distinction between denotation lists and formulation lists, the latter corresponding to various types of reformulation, the analysis of spontaneously occurring instances of lists seems to undermine such a radical view. In this section, we adopt a different perspective, asking what we can learn by observing the behavior of lists in unplanned spoken interaction. That is, after identifying lists as syntactic forms, we will look at how these forms surface out in interaction, or how they are adapted to local contexts and local conversational needs. For this purpose, we adopt some of the categories of Interactional Linguistics (see Selting & Couper-Kuhlen 2001). After making an attempt to reconcile the description of lists given so far with the perspective of spoken syntax, relying on Auer's (2009) discussion of *retraction*, we concentrate on the temporal organization of lists and on the possibility for multiple speakers to participate in the production of a list.

2.2.1   *Lists and spoken syntax*

One of the most relevant generalizations captured by Masini et al.'s (2018) model of listing consists in its ability to include under the same descriptive label ('list') two different phenomena that are often regarded as unrelated and belonging to different levels of linguistic analysis: lists sensu stricto and reformulation. When speaking of 'lists' in a non-technical sense, one normally refers to *denotation lists*, that is, lists whose function is to perform semantic and pragmatic operations on the referential meaning of the elements included within this pattern. The other side of the coin is represented by *formulation lists*, which, on the contrary, operate on the illocutionary level: these structures are normally classified as instances of reformulation, intended in a broad sense as the replacement of an older utterance, or part of it, with an alternative. However, despite capturing the similarity between these mechanisms (both regarded as 'lists' in a broad, descriptive sense), Masini et al. (2018) do not offer a constructionist analysis of formulation lists (although they advocate for it) and therefore do not include formulation lists in the constructional network developed for denotation lists. So, they leave the question open whether speakers have at some level a unique representation for denotation lists and reformulation.

A different view on this issue is offered by the theoretical framework of Interactional Linguistics, mainly due to its consideration of language structure as rooted in conversation, and therefore, as a context-bound and temporally-bound activity (Selting & Couper-Kuhlen 2001; Deppermann & Günthner 2015). Particularly, Auer's (2000, 2009, 2015) notion of on-line syntax provides in our view the proper analytic tools for a unified account of denotation lists and reformulation. By following this orientation, we assume that language is produced and processed in a temporally organized way: thanks to this inherent feature of spoken communication, speakers are able to project (i. e. anticipate) what is coming next in an emergent and still unfinished linguistic unit, as well as to relate new linguistic units to previously completed syntactic gestalts (Auer 2015). One of the basic operations of spoken syntax is *retraction* (Auer 2009), which is defined as the ability to recycle a previously activated syntactic slot and use it more than once. As argued in Auer & Pfänder (2007), retraction is the common denominator shared by denotation lists and other phenomena that are typical of spoken language and virtually absent in written productions, such as various types of repair (Schegloff 1979; Pfeiffer 2017), reduplication and hesitations.

We will now provide examples that show how the notion of retraction can come at hand to help reconciling denotation lists with other phenomena occurring in spoken language. Let us first consider Example (14):

(14)   KIP, BOC1006

```
01  Silvio   diciamo che probabilmente non=e::h,
             let us say that probably
02           non ci puo' essere un metodo che va bene per tutti,
             there cannot be a method that goes for everybody
03           perche' appunto:,
             because indeed
04           diverse: s=diversissime possono essere le situazioni che
             si incontrano nella prassi dell'insegnamento,
             different, very different can be the situations that are
             encountered in the praxis of teaching
05           =diversi sono gli attori che vi prendono parte,
             different are the actors taking part in it
06           le esigenze, i bisogni,
             the needs, the necessities
07           e:::: insomma tutta una serie insomma di fattori da
             tenere:: in conto,
             in sum, a wide range of factors to take into account
```

This is an excerpt from an oral exam of an undergraduate student. At line 3, Silvio starts producing a complex unit in order to provide a justification for the statement at line 2 (*there cannot be a method that goes for everybody*). If we look at the sequential organization of his turn, we can see that after selecting a syntactic form at line 4, the speaker constructs his turn incrementally: at line 5 he re-uses the previously activated syntactic structure to add a new element (*actors*) in the NP slot; at line 6 he chooses another strategy and simply adds two more referents in the same NP slot. The syntactic parallelism established between these three adjacent units has the effect of activating a semantic parallelism between the referents involved, which are constructed as being on a par with each other. This is made explicit at line 7 with the production of a post-detailing component in Selting's (2007) terms: by providing a general label, *factors to take into account*, Silvio invites his interlocutor to go back to the previously formulated units and retrospectively construct them as part of a broader set; see Figure 3.

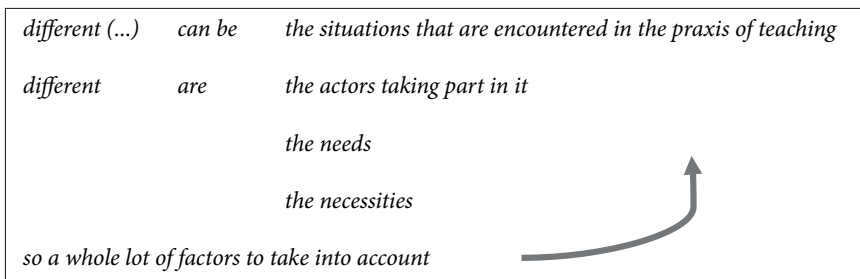|		
different (...)	*can be*	*the situations that are encountered in the praxis of teaching*
different	*are*	*the actors taking part in it*
		the needs
		the necessities
so a whole lot of factors to take into account		

Figure 3. Schematic representation of Example (14)

In the excerpt reported in (14), what would be labelled as a list leading to a category of 'factors that a teacher must take into account' comes about as a sequence of partially unrelated increments, whose interpretation builds on the replication of the same syntactic pattern. The analysis of such an example is particularly helpful to bridge the gap between denotation lists and conversational repair: when the temporal dimension is taken into account, both structures appear as strategies to elaborate on a previously realized linguistic unit. In some cases, retraction is used to add new elements, in other cases it allows the speaker (or another participant) to replace some previously added unit and propose a semantically or pragmatically more adequate alternative. It is particularly worth noting that post-detailing components such as the one produced in (14) are indeed instances of reformulation, in that they propose a better definition for something that was previously formulated as a list of subsequent increments. Therefore, there is no need, in principle, to distinguish between denotation and formulation lists from the perspective of interaction: the two are produced by the same operation of spoken syntax.

2.2.2   *The temporal organization of lists*

According to one of the main tenets of the conversation analytic framework "the units which speakers use to construct turns-at-talk house actions" (Barth Weingarten & Couper-Kuhlen 2011: 264). A complex pattern such as the list can thus be better imagined as the combination of several separated and partly independent moves that incrementally lead, through inferential reasoning, to the construction of reference to individuals or sets. In this interpretation, lists are extremely flexible tools for the purpose of categorization because they exploit the possibility to activate structural parallelisms across linguistic units (see also the notion of 'resonance' in Du Bois 2014) in order to establish a relation between the referents that are used as exemplars. For a better understanding of the relation between lists and categorization, we will now look at how lists are organized along the temporal dimension in order to convey reference to a particular category by analyzing what we call, after Selting (2007), the pre- and post-detailing components.

2.2.2.1   *Pre-detailing*

A pre-detailing component is generally represented by a semantically vague expression by means of which a speaker makes a first attempt to refer to a category. Such vague expressions, being semantically insufficient for the purpose of categorization, have the effect of projecting one or more exemplars, to narrow down the denotation of the category that the speaker is trying to build. Perhaps the simplest example of pre-detailing in a list is found in exhaustive lists where the speaker overtly specifies the number of exemplars that will be included in a given set. This is straightforward in Example (15):

(15)   KIP, TOA3006

```
01  Romina   ma io, volevo provare::: a parte i tre big,
             but I wanted to try, except from the three big ones
02           che sono roma, milano, [e genova,   ]
             which are Roma, Milano and Genova
03  Mara                             [milano. sì.] sì sì.
             Milano, yes, yes, yes.
04  Romina   ci sono::: altre scuole,=che sono i teatri stabili alla fine.
             there are other schools, which are the Teatro Stabile in the end
```

In (15) Romina is talking to Mara about private academies that she is planning to apply for after graduation. At line 1 she produces a pre-detailing component, *the three big ones*, that projects an upcoming list which is overtly constructed as having three members. These are made explicit in the list at line 2, *Milano, Roma* and *Genova*. This example also shows clearly that it is misleading to assume a direct relation between categorization and non-exhaustive lists: in spite of having a finite number of members, the list contained in this example leads nonetheless towards the ad hoc construction of a category. Even if the construction here does not invite to infer possible other members, as is the case with non-exhaustive lists (Barotto & Mauri 2018), it leads all the same to the construction of a set that is only relevant in a very specific context and grounded in the specific activity performed by speakers. An example of pre-detailing in a non-exhaustive list is found in (16):

(16)   unpublished data

```
01  Ugo   ma ci sono anche documenti che provengono direttamente da
          quel passato più lontano
          but there are also documents that come directly from that
          distant past
02        (0,39)
03        ad esempio (.) i dati dell'archeologi::a,
          for example the data provided by archaeology
04        (0,18)
05        le iscrizioni di cui:: parlere:mo,
          the inscriptions we will talk about
06        (0,09)
07        la numismatica le::: mh:: le monete,
          numismatics, coins
```

The example is the excerpt of a lesson in ancient history. At line 1, Ugo introduces in a pre-detailing component a general category of *documents that come directly from that distant past*. Such a general formulation anticipates the introduction of a number of exemplars of this category.

2.2.2.2 *Post-detailing*

Post-detailing is a backward-oriented move that has the function of going back to an already produced piece of interaction and make explicit its interpretation as a list. This is more explicitly related to categorization in that, by using a post-detailing component, participants overtly recognize the previous units as part of a list and provide a label for the category that is constructed. Consider Example (17) below:

```
(17)   KIP, TOA3005
01  Mara   io vorrei imparare (.) a fare gli gnocchi da mia nonna,
           I would like to learn how to make gnocchi from my grandma
02          =gli gnocchi, e::: le tagliatelle i ravioli da mia nonna,
            gnocchi and tagliatelle, ravioli from my grandma
03          la pasta fresca che lei la fa (.) benissimo,
            fresh pasta, which she is very good at preparing
```

Here the post-detailing component (*la pasta fresca* 'fresh pasta') has an almost metalinguistic value, in that it is anaphorically related to the previous list of exemplars and has the function to provide a label for the set evoked by the whole list, with similarities with the phenomenon of anaphoric encapsulation (Conte 1996). The interpretation of the list in this case is rather straightforward in that there is little reference to context or shared knowledge, therefore the function of post-detailing could be to fill a momentaneous lexical gap. A more significant example of the backward-oriented function of post-detailing is found in (18):

```
(18)   KIP, TOA3006
01  Mara   cioè gli stavo dicendo io adesso voglio prendermi:::
           I mean I was saying to him, right now I want to take for
           myself
02          almeno:::
            at least
03          non ti dico un mese che forse è troppo. però un paio di
            settimane che::
            I don't want to say a month, which is maybe too much. But
            at least a couple of weeks when
04          mi guardo tutti i film che non ho visto,
            I watch all the movies that I haven't seen
05          mi leggo tutti i libri che non ho le:tto,
            I read all the book that I haven't read
06          mi guardo tutte le serie che non ho guarda:to, sai,
            I watch all the TV shows that I haven't seen you know
07  Romina [sì sì sì sì]
           yes yes yes yes
08          [faccio un po' di::: no?]
            I do a little of, you know
09          (0.36)
10          un po'di ripiglio,
            a little recover
```

In this example, Mara produces a highly cohesive list by replicating three times the same syntactic pattern (lines 4–6). Reference to a shared common ground is also indicated by the use of the discourse marker *sai* (see also Fox et al.'s 2002 discussion of *you know* in English) at the end of line 6. This is interpreted by Romina as a possible point of completion, as she acknowledges the felicitous construction of a category by repeating *sì* 'yes', which is in this case used as a discourse marker operating on the illocutionary level. At lines 8–10, Mara produces a post-detailing component in order to make explicit the category she is referring to, that includes *watching movies, reading books* and *watching TV shows*. After a few hesitations she uses the noun *ripiglio*, produced as an occasional conversion from the verb *ripigliare* 'recover'.

2.2.3   Dialogicity in list production

While written language offers an inherently monological view on the phenomenon of categorization, an analysis of this phenomenon in spoken language must necessarily take into account the dialogic nature of spontaneous speech. It is a widely acknowledged fact within the interactional linguistic framework that every construction may be co-operatively produced by the different participants of a conversation, so that its syntactic form will be split between different turns, with several implications for our understanding of the temporal organization of language processing and production. This is the case of the so-called co-constructions (Lerner 1991; Pekarek-Doehler 2011). Not unlike other linguistic structures, lists may be constructed either monologically, as in all the examples discussed so far, or dialogically, allowing for contributions from different participants, as anticipated in Section 1.2. This has important implications for an account of categorization because it reveals that creating a category in interaction is a socially meaningful activity that serves some purpose in a given context; therefore, all the participants may engage in this task. Bringing this reasoning to an extreme, co-construction in the production of lists is the reflex of co-operation between the participants in constructing reference to a specific category of entities.

Example (19) is particularly representative of how different speakers can co-operatively engage in the production of lists:

(19) KIP, BOA3004

```
01  Fernanda   e::, un matto, tipo=eh::, ha capito che noi p-, io e la
                mia amica parlavamo in italiano::,
                some crazy guy, like realized that my friend and I
                were speaking in Italian
02              =e m'ha detto stronzo (.) vaffanculo!
                and he said to me "stronzo", "vaffanculo"
03              ed è sceso dall'au-, dalla: s-bahn.
                and then he got down from the S-Bahn
```

```
04   Flavia     eh,
                eh
05   Fernanda   e ci son rimasta tipo co[sì.]
                and I was like this
06   Livio                         [eh:,] la terza, la terza
                parola di sol-, in genere è berlusconi.
                the third word is normally Berlusconi
07   Fernanda   no, no.      [stron]zo, vaffan[culo.]
                no, no (he said) "stronzo, vaffanculo"
08   Flavia      no. anche ma[fia.]
09                                      [mafia]
                e[::::]
                no, there's "mafia" as well, mafia and
10   Livio      [mafia.]sì.
                "mafia" yes
11   Flavia     sì.
                yes
12   Fernanda   mafia mandolino:, pizza.
                mafia, mandolin, pizza
13   Flavia     beh [pizza,] (.)[pasta:. spaghetti,]
                well, pizza, pasta, spaghetti
14   Fernanda      [spaghetti.]
                    spaghetti
15   Livio        [a me::,] berlusconi
                             [un sacco. spes]so mi
                son sentito dir[e.]
                to me, Berlusconi a lot. I often heard that being
                said to me
16   Fernanda               [pe]nsa te.
                Can you imagine?!
```

The example was extracted from the spontaneous speech section of the KIParla corpus and is part of a conversation between four students chatting during lunch break. The conversation does not have a fixed topic. In this passage, Fernanda is talking about an experience she had while travelling in Austria. At line 2, she reports a sequence of Italian insults received from a local in the underground, merely due to the fact that he recognized that they were speaking in Italian. The utterance has the prosodic contour associated to exclamative sentences, which is characterized by falling intonation, and is therefore a complete unit. This sequence of insults reported by Fernanda is also an exhaustive list that in principle does not allow further completions, as her intention is to report the exact two words that she heard in that occasion. This intention is, however, misinterpreted by Livio: in fact, he metalinguistically "re-opens" the list by asserting the existence of a third element belonging to the same set and corresponding to *Berlusconi*, the name of a well-known Italian politician. At line 7, Fernanda rejects this contribution by repeating the two exact words from her previous quotation; again, falling intonation is a cue indicating that the list is semantically complete, and thus exhaustive.

Flavia, however, accepts Livio's expansion and adds in turn another member to the emerging list. What happens here from the perspective of categorization? And why? At line 6, Livio misinterprets the list produced by Fernanda and provides a new interpretation to an already completed unit. By adding *Berlusconi* as a third element, he actually performs a shift from the strictly closed set of "insults I received on that particular occasion" constructed by his interlocutor, to an open set of "Italian expressions typically known abroad". Thus, a previously completed list is recycled by another participant with a different meaning. The new emerging list is also constructed as non-exhaustive and is now open for contributions by all the participants. The creation of this new category is felicitous, as can be seen from the fact that Fernanda accepts the shift to the new category by providing herself some exemplars. The creation of this category is thus carried out cooperatively by all three participants.

2.3   Interim summary: Bridging the gap between grammar and interaction

To conclude this section, we would like to make the point that the two approaches presented here – lists as constructions, namely as part of our grammatical knowledge, and lists as interactional devices – are not necessarily competing views but two sides of the same coin that should rather complement each other.

Whereas Construction Grammar has traditionally focused on grammatical description, with special reference to sentence-level syntax, more recently some of its practitioners have been paying increasing attention to discourse-related and interactional phenomena. As Fried (2010) reminds us, the idea of integrating grammatical description and the communicative dimension was already envisaged in early works on Frame Semantics by Charles Fillmore, who identified two kinds of frames: "cognitive frames" and "interactional frames". Whereas cognitive frames "can be thought of as motivating the categories speakers wish to bring into play when describing situations that might be independent of the actual speech situation", interactional frames have "to do with how we conceptualize what is going on between the speaker and the hearer", i.e. the actual communication situation (Fillmore 1982: 117; quoted in Fried 2010: 125). However, the thorny issue of the relationship between grammar and interaction (and spoken language in general) has long remained in the background and has entered the construction grammarians' agenda only in recent times. As Fried (2010: 126) puts it, this new orientation seeks to test the hypothesis that "certain knowledge of the socio-pragmatic patterns in which linguistic expressions are used constitutes a fundamental part of the speakers' communicative competence and plays a role in shaping grammatical organization as well", much in line with what usage-based (constructionist) approaches speak for

(Bybee 2006, 2010, 2013; Diessel 2015, 2017). The notions of "discourse pattern" (Östman 1999, 2005), for instance, goes in this direction.

Hilpert (2019: Chapter 9), in discussing the role of constructions in spoken language, compares Hopper's theory of Emergent Grammar – according to which "structure, or regularity, comes out of discourse and is shaped by discourse" (Hopper 1987: 142) – with constructionist models and claims that the two are mostly compatible, the main difference lying in the extent to which abstraction is allowed: whereas Hopper is reluctant to posit schematic structures, abstraction has a key role even in more usage-oriented constructionist models. Hilpert concludes that emergent phenomena are nonetheless linked to abstract mental representations by virtue of "sedimentation" (Linell 2005): "[r]ecurrent linguistic routines become sedimented as grammatical constructions", although "[d]istinguishing between emergent structures and sedimented constructions is not a trivial task" (Hilpert 2019: Chapter 9). In this respect, also Fischer (2010), by pointing out that Construction Grammar is fit for handling interactional features, argues that both the interactional and the cognitive dimensions are necessary, with generalizations being part of the picture.

Finally, recent proposals such as *Interactional Construction Grammar* (Imo 2015) advocate for the need to integrate Construction Grammar with the body of knowledge on spoken language developed within approaches such as Interactional Linguistics (Section 2.2.1) and Conversation Analysis. The latter had already been recognized as compatible with Construction Grammar by Östman (2005), one of the first constructionist works tackling the issue of spoken discourse. The present study falls quite naturally into this line of research, to which we hope to contribute.

3.   Between categorization and reformulation: The case of *insomma*

3.1   Discourse markers and lists

Not unlike other linguistic expressions, the felicitous production and processing of lists builds on contextualization cues, intended as a set of linguistic resources through which "speakers signal and listeners interpret what the activity is, how semantic content is to be understood and how each sentence relates to what precedes or follows" (Gumperz 1982: 131). Prosody, for example, plays an eminent role in the production and processing of lists, as it is used to frame single elements into a broader list pattern spanning over several sub-units (see Jefferson 1990; Selting 2007), or to signal a locally relevant interpretation of a list pattern. In this regard, Matalon (this volume) argues that particular prosodic patterns in spoken Hebrew

are used when the speaker, by producing a list, is taking counter-stance and is therefore listing arguments and constructing a category for the purpose of supporting their claims.

Discourse markers (DMs), intended here in a broad sense as those items whose primary function is to provide "instructions from the speaker to the hearer on how to integrate their host unit into a coherent mental representation of the discourse" (Moosegard-Hansen, 1998: 358), are another resource that may give cues on how a list should be interpreted. Several works in the last years have illustrated that DMs are in many ways related to categorization and to the linguistic resources by which it is conveyed: DMs may lead the interpretation of lexical expressions through lexical adjustment (e.g. Wilson & Carston 2007), convey intentional vagueness (e.g. Ghezzi & Andorno 2014; Voghera 2013) or signal the exemplar value of a given referential expression, establishing a paradigmatic relation with other implicit or explicitly verbalized alternatives (e.g. Barotto 2018; Lo Baido 2018; Fiorentini 2018a; b). For what concerns list constructions, a great deal of recent research has been devoted to elements – such as the English *etcetera, and things like that, and stuff, and all* – that are used in various languages to mark the conclusion of an open list and at the same time to signal the exemplar value of the previous elements, thus inviting the construction of a category based on these exemplars. They have been referred to as 'generalized list completers' (Jefferson 1990), 'vague category identifiers' (Channell 1994) or, much more frequently, as 'general extenders' (Overstreet 1999).[7] However, little attention has been given so far to the contribution given to the production and interpretation of lists by other types of DMs.

If we consider the abstract representation of the list construction by Masini et al. (2018) given in Figure 1 (Section 1.1), general extenders systematically occur in the LIST COMPLETER slot. This means that general extenders have sequential properties within an emergent list of exemplars, in that they signal the conclusion of a list and at the same time they invite the construction of a category based on the previously mentioned exemplars. Other types of DMs may occur as INSERTIONS within the list (expressing hesitations, metalinguistic or modal comments, etc.). Finally, DMs may occur in the pre-detailing or post-detailing positions that surround the list, overtly signalling the relationship between their host unit and the whole emergent list construction. Consider the following example:

---

7. For recent discussion on general extenders in Italian, see Fiorentini (2018a).

(20)   KIP, TOD2011

```
01  Natalia   è stato bello anche capire:: quali sono gli usi e costumi
              degli irlandesi,
              it was nice to get an idea also of what are the customs
              and traditions of the Irish
02            ehm::: eh e: poterli vivere a pieno,
              and to be able to have a full experience of them
03            e: quindi:: osservarli:: da: da vicino,
              and so observe them closely
04            e:::h abitando nella stessa casa con loro.
              living in the same house with them
05            =quindi ad esempio:: la mattina:: a casa loro: si:
              mangiava::
              so for example in the morning at their place we used
              to have
06            non biscotti:, o::=mh
              not biscuits or
07            cose diciamo piu' dolci, ma:::
              let's say sweeter things, but
08            cose piu' salate quindi ad esempio: toast, uova:
              saltier things so for example toasts, eggs,
09            e::h=m::h insomma la classica:: colazione:: piu'
              all'americana,
              in sum, the classic more American-style breakfast
```

In this passage, which is extracted from an interview given by a student to one of her
fellow students acting as interviewer, Natalia is telling about a summer school she
attended in the past, thanks to which she could live in the house of an Irish family.
DMs contribute at various points to the sequential organization of the narrative,
mostly in association with instances of retraction (see Section 2.2.1). At line 3, a
cluster formed by a conjunction and a DM (*e quindi* 'and so') signals that its host
unit is the concluding element of a three-parted list starting at line 1. After its
completion, this list is used as a general formulation that is further expanded by
another unit that starts at line 5, where Natalia gives an example of the differences
between Italian and Irish breakfasts. The unit is introduced by two DMs, *quindi* 'so'
and *ad esempio* 'for example', whose joint effect is to link the upcoming unit with its
previous context, presenting it as a consequence of the previous formulation and at
the same time expliciting its exemplar value. More interestingly, the same cluster oc-
curs at line 8 as the introducer of a list that is nested within a broader structure (see
Figure 4): the list is preceded by a general formulation, *saltier things*, functioning
as the pre-detailing component; then, the cluster *quindi ad esempio* introduces the
list itself, formed by two exemplars; finally, the list is followed by a post-detailing
component where Natalia provides a label for the category she is trying to convey
(namely *more American-style breakfast*), which is introduced by the DM *insomma*
'in short, in conclusion, so'. The whole syntactic complex developed in this passage
can be represented schematically as in Figure 4.

It was nice to	get an idea also of what are the customs and traditions of the Irish		
	and to be able to have a full experience of them		
	and so observe them closely, living in the same house with them		
	so for example in the morning at their place we used to have		
		not biscuits or	
		let's say sweeter things, but	
PRE DET		saltier things,	
X₁		**so for example** toasts,	
X₂		eggs	
POST DET		in sum, the classic more American-style breakfast	

Figure 4. Schematic representation of Example (20)

The representation given in Figure 4 demonstrates how the construction of a list is nested within other conversational practices that speakers use and combine in order to structure their turns. DMs have a crucial role in structuring this monological passage, as they provide cohesion between its units thanks to their ability to signal the value of their host unit in the context of the ongoing discourse, either anaphorically or cataphorically.

In the remaining part of this section we discuss a case study concerning one of the DMs contained in (20), namely *insomma*. In Section 3.2, we provide a corpus-based analysis of the cases in which *insomma* occurs within categorizing lists, and we defend the hypothesis that it has developed the function of introducing the post-detailing component. The theoretical implications of this proposal, including a refinement of the list skeleton in Figure 1 and of the constructional network developed by the abstract list construction (Section 1.1), are discussed in Section 3.3.

3.2   Insomma and lists: Data

Italian *insomma* is a good case in point for illustrating how DMs may contribute to the production and interpretation of categorizing lists in interaction. In contemporary Italian, *insomma* is a polyfunctional item with a variety of uses: it may be used as a DM, especially with a reformulating, conclusive function ('in sum /

in short / in conclusion / so'), or as a more general adverb meaning 'so-so', or yet as an exclamation meaning (among other things) 'for goodness sake! / come on!' (see De Mauro n.d.). Given its polyfunctionality, we may expect to find this item in different positions and functions, and this is indeed the case. See for instance (21), where *insomma*, together with the hesitation marker *eh*, serves as an insertion that is used to gain some time to elaborate the rest of the list. In (22), instead, *insomma* serves as a reformulation marker in-between the two formulations (*dello stato* and *dei servizi dello stato*), whereas in (23) it occurs after the second (and last) formulation (*perdeva forze*).

(21) *perche'_ loro avevan vissuto da_**insomma** da molto tempo sotto la democrazia e quindi avevano anche delle delle liberta' e anche **insomma** eh e potevano e_ avevano delle liberta'*   (LIP, FC6)
      'for they had been living since, INSOMMA, for a very long time under democracy and so they also had some, some liberties and also INSOMMA eh, they could eh, they had some liberties'.

(22) *quella che e' stata una politica_ dello stato **insomma** dei servizi dello stato*
                                                                      (LIP, RE9)
      'what has been a policy of the State, INSOMMA of the State's services'

(23) *allora vuol dire che l'organizzazione eh aveva eh stava perdendo stava cedendo perdeva forze **insomma***   (LIP, NC12)
      'so it means that the organization eh had eh was losing, was failing, was losing its strength INSOMMA'

In various accounts, *insomma* is described as an element used for delimiting discourse units and for signaling repair, often with paraphrastic reformulation value (e.g. Bazzanella 1995; Fiorentini 2018a; b). Waltereit (2006) also focusses on the sequential properties of *insomma* and argues that it may be either forward-oriented, in the cases where it marks the host clause "as a near paraphrase, as a formulation alternative for the preceding portion of discourse", or backward oriented, in those cases when it marks its host unit as "a closing statement for the preceding portion of discourse" (Waltereit 2006: 65). These are exemplified respectively in Examples (24) and (25):

(24) *lei mi aveva detto nel caso poi facciamo un'analisi direttamente all'orale **insomma** in classe*
      'You told me, if it's the case then we perform an analysis directly at the oral (i.e. exam), INSOMMA, in class'   (KIP, BOA1004)

(25) *non so se avete studiato filosofia, hegel, la dialettica della storia. **insomma** comunque c'è un'idea di progresso.*
      'I don't know if you studied philosophy, Hegel, the dialectics of history. INSOMMA, anyway there is an idea of progress.'   (KIP, TOD1014)

Little attention has been given so far to the role played by *insomma* specifically within list patterns, as is the case in Example (20), which is partially repeated as (26):

(26) *cose più salate, quindi ad esempio toast, uova, **insomma** la classica colazione più all'americana*
'saltier things so for example toasts, eggs, INSOMMA, the classic more American-style breakfast'   (KIP, TOD2011)

In cases like this, *insomma*'s function is not just to introduce a reformulation, but rather to introduce the post-detailing component of a categorizing list, and, hence, to label the relevant higher-level category intended by the speaker. Crucially, this function can be considered at the intersection between the repair function and the sequential backward-oriented function identified by Waltereit in non-list contexts. The post-detailing component is indeed a type of repair (see also Section 2.2.1), in that it provides a general label for a previously formulated list of elements, and it is also a closing move that signals the end of a list pattern and resolves the task of categorization by providing a label for the category.

In what follows we describe this (apparently emerging) function of *insomma* as an introducer of post-detailing component within categorizing (hence denotation) lists, emphasizing the link between these cases and its (major) role as a reformulation marker in formulation lists, which are formally very similar.

For the purpose of the present study, we performed a search on two corpora of spoken Italian, the KIParla corpus (Goria & Mauri 2018; Mauri et al. 2019a) and the LIP/VoLIP corpus (De Mauro et al. 1993; Voghera et al. 2014).[8] We extracted all the occurrences of *insomma*, then focusing on those in which *insomma* occurs in a pattern of retraction (see Section 2.2.1) with a form "X *insomma* Y",[9] like in the previous examples. This allowed us to single out, among all the possible uses of the scrutinized DM, instances of *insomma* that introduce either a reformulation or a post-detailing component. The total instances of *insomma* are 978 in the LIP corpus and 577 in the KIParla corpus, and in both cases the vast majority of this amount is represented by pragmatic uses that are not relevant here. Patterns of

---

8. KIParla is a recently built corpus that collects around 70 hours of various types of spoken interaction occurring in the academic context, including conversation within the students' peer group, lessons, office hours, exams and semi-structured interviews. Furthermore, it offers a wide range of metadata concerning both the participants and the situation. LIP/VoLIP is a collection of recordings from the Eighties and Nineties from a broader range of settings, belonging to 5 macro-types, namely: face-to-face conversations, telephone conversations, interviews and debates, monologues, radio and TV broadcasts.

9. We also considered that *insomma* does not have a fixed position within its host unit and considered thus also cases where the same relation within two conjuncts, either at sentence level or at discourse level, may be instantiated by patterns like "X, Y *insomma*", as in Example (23) above.

retraction – including instances of reformulations and categorizing lists – represent in both corpora less than 10% of the occurrences, hence a minority of cases. Given our current purposes, we carried out a strictly qualitative analysis of this subset of data, relying on the observation of the patterns of usage of *insomma* that are relevant for the study of categorization, as defined above.

Corpus data show that, when *insomma* occurs within a pattern of retraction, it may signal a repair (see Example (23) above), or introduce the post-detailing component in a list, as in Example (20). The latter function is particularly clear in (27):

(27) *quando vivi da solo o vivi col partner da soli **insomma** senza altri coinq- inquilini estranei*   (KIP, TOD2012)
      'when you live alone, or with your partner alone, INSOMMA, without other stranger flatmates'

Here, the speaker produces a list of alternatives connected by the conjunction *o* 'or' (see Ariel & Mauri 2018 for similar cases) and then uses *insomma* to introduce a closing move where she expresses explicitly the higher-level category she is referring to. It must also be noted that the intended denotative meaning of the list, without considering the move introduced by *insomma*, is by no means encoded compositionally by the two exemplars, which refer to two almost antithetic situations (living alone and living alone with one's partner). From a pragmatic perspective, the speaker here is not only suggesting that the two situations she mentioned are to be read as parts of a broader set of situations, but also inviting to narrow (Wilson & Carston 2007) the reading of 'living alone' to a particular interpretation. The post-detailing component introduced by *insomma* solves thus a potential conflict between the semantics of the two referential expressions and indicates what is the intended context-bound reading of the list: '(situations where you live) without stranger flatmates'.

Even if, as argued in Section 1.1, post-detailing is not a necessary component of lists, Selting contends that in her spoken German data "participants <…> orient to the completeness of the structure" (2007: 496) and generally wait for their interlocutor to produce a post-detailing component before turn-taking. In this respect, Italian data do not show such a clear-cut tendency and in several cases post-detailing is not produced: therefore, we can assume that when speakers do produce a post-detailing component, this must reflect some more specific activity related to categorization and have some particular semantic or pragmatic value that is not present elsewhere.

Our analysis allows us to identify, in particular, two relevant contexts. The first one is represented by cases where the speaker needs to solve a potential clash between syntactically heterogeneous elements included in the same list. Consider Example (28):

(28)   KIP, TOD1014

```
01  Susanna   ma guardare una sfilata, e' la cosa piu' noiosa che ci
                possa essere, perche',
```
*but watching a fashion show is the most boring thing
there can ever be, because*
```
02              >cioè< as- un' m::h infilata di lo:ok,
```
I mean,[10] a row of looks
```
03              e:h con la stessa punto di vi:sta
```
with the same point of view
```
04              il drone sta sempre li',
```
the drone always stays still,
```
05              la camera:, cioè la slide che fa, tum tum avanti e
                indietro,
```
*the camera, I mean, the slide goes vroom vroom
back and forth*
```
06              insomma noiosissimo.
```
insomma *extremely boring.*

Susanna is a professor and she is speaking, during her class, about watching the recording of a fashion show, as opposed to watching it live at the venue. At line 1, she produces a pre-detailing component, corresponding to her evaluation (*it is the most boring thing there can ever be*), and then continues her turn with a list of examples. The speaker produces a clear example of ad hoc category in the sense of Barsalou (1983), as she is actually referring to a set of "boring things that characterize a filmed fashion show", which is thus anchored to a specific activity (i.e. watching the recording) and to a specific context that the hearers are supposed to know; consider the use of definite articles for elements that were not previously introduced. The list is organized as a series of partially independent and syntactically heterogeneous increments: at line 2 there is a verbless predication, which is further expanded at line 3 by a prepositional phrase; at line 4 we have a finite main clause and finally, at line 5, another verbless predication containing a relative clause. We analyze these elements as syntactically parallel as they all occur in the slot corresponding to the projection of the complementizer *perché* 'because' at line 1. The post-detailing component produced at line 6 links this otherwise unorganized array of elements to the speaker's previous formulation providing a key for its interpretation in context. Therefore, in this case, *insomma* retains both its repair-like anaphoric function, as it links its host unit to the previous context in order to specify the intended reading, and its forward-oriented function, as it signals the conclusion of the list and leaves the floor open for the beginning of a new activity.

---

10. Italian *cioè* is the univerbated form of *ciò è*, literally 'that is'. This was translated here with the English *I mean* due to its most common use as a marker of reformulation.

A similar case is represented by those contexts in which the list is syntactically homogeneous, but the semantic contribution given by the exemplars is deemed unsatisfactory for the felicitous construction of a category. Consider Example (29):

(29) *il gatto e' sempre stato_ eh un cioe' domestico // ma anche in un certo senso tuttora non lo e' cioe' ha un rapporto con // l'uomo che non e' tipicamente quello dell'animale domestico non e' // quello del bestiame non e' quello del cane non e' quello del cavallo // eh* **insomma** *il gatto non e' mai stato sfruttato dall'uomo*

(LIP, FD15)

'The cat has always been, eh, domestic // and yet in a way it is not properly domestic even now, I mean, it has a relationship with // the man, which is not typically that of a domestic animal, it is not // the same as cattle, it is not the same as the dog, it is not the same as the horse // eh INSOMMA the cat has never been exploited by man'

Here the sole list of exemplars is treated as potentially ambiguous: simply putting together *dogs, horses* and *cattle* is regarded as insufficient in the present context to convey the type of denotation intended by the speaker, as the list has different potential readings. By introducing a post-detailing component, *insomma* helps thus to disambiguate possible interpretations, providing the key for this passage.

The second context has to do with the sequential properties of *insomma* in lists: as said, one of its core functions is to mark its host unit as the conclusion of a previously initiated discourse unit. Within a list, this translates in the production of a unit that signals the end of the construction and thus leaves the floor open for contributions by other speakers. Consider Example (30):

(30) KIP, BOC1009

```
01  Serena    mentre invece la faccia negativa e' l'interesse che il
                proprio territorio non venga minacciato=
                while instead, negative face is the interest that one's own
                territory is not threatened
02              =e quindi che venga preservata,
                and so that they preserve
03              una liberta' di azione e di e::h,
                some freedom of action and of ehm
04              di::, cioe',
                of, I mean
05              una liberta' di:: movimento di azione:: insomma (.) globale
                °ecco°.
                some freedom of movement, of action, insomma, global
                freedom, yes.[11]
06  Carolina  va bene.
                well.
```

---

11. We translate the Italian discourse marker *ecco*, that in Italian is used to give particular emphasis to some constituent (see Bazzanella 1995) with the English *yes*.

Serena is taking an oral exam and answering a question posed by Carolina, the professor. In structuring the last part of her answer, she gives a first definition at line 3 of *freedom of action*. This is treated immediately after as unsatisfactory, and in fact the speaker performs self-repair by using a list pattern where her previous formulation is only one of the listed items. *Insomma* is used here to introduce the post-detailing component which, again, provides a general label subsuming all the listed items. At the same time, *insomma* is interpreted as relevant for turn-taking, as can be seen by the fact that Carolina takes the floor and puts an end to Serena's answer (and to the exam) only after this component is produced.

3.3   Theoretical discussion

This case study on the employ of *insomma* within lists allows for some generalizations concerning lists and categorization. The data analyzed so far show that *insomma* may have, within a list, the function of introducing a post-detailing component, the production of which may be seen in relation to different motivations. We therefore identified:

1. Reference oriented motivations: that is, cases where *insomma* solves a potential clash between elements with different semantic characterization and introduces a general label;   (Examples (27) and (29) above)
2. Processing oriented motivations: that is, cases where there is a formal clash between syntactically heterogeneous elements produced as subsequent increments and retrospectively "forced" into the list pattern; (Example (28) above)
3. Interaction oriented motivations: that is, cases where *insomma* marks the end of the list and leaves the floor open to the beginning of a new activity.
(Example (30))

These findings have implications for our representation of the list structure as it was described in Section 1.1. As already mentioned, much attention has been given in current research on the elements that close a list, and particularly general extenders (see Section 3.1), while the internal structure of the post-detailing slot has never been analyzed in detail. What our data on *insomma* show is that the post-detailing component may be preceded by another structural position, which may be named 'post-detailing component introducer' (POST-C-I) and which is filled by items like *insomma*. Figure 5 illustrates the list skeleton in Figure 1 revised accordingly. Example (26) is mapped onto to it (in translation) for the sake of exemplification.

PRO-C	projection component	*saltier things*
LI	list introducer	*so for example*

X₁	**conjunct 1**	***toasts***
CO	coordinator / connective	-
X₂	conjunct 2	-
CO	coordinator / connective	-
X₃	conjunct 3	-
…	…	-
CO	coordinator / connective	-
X_LAST	**conjunct last**	***eggs***
LC	list completer	-

(INSERTIONS — bracket to the left of the boxed section)

POST-C-I	post-detailing component introducer	**INSOMMA**
POST-C	post-detailing component	*the classic more American-style breakfast*

Figure 5. List skeleton revised

The individuation of this new structural position within the list skeleton has consequences also for the construction grammatical framework outlined by Masini et al. (2018) (Section 2.1). Obviously, we need to upgrade the formal representation of the list construction in (10) so that it adheres to the new skeleton in Figure 5:

(31)   ABSTRACT LIST CONSTRUCTION REVISED
       Form: ([PRO-C]) ([LI]) {X_1 | (^CO) (X_2) | (^CO) (X_3) | … | (^CO) X_{LAST} | (LC)} ([POST-C-I]) ([POST-C])
       Function: 'function f over the set of Xs + presupposition p = common categorization underlying Xs'

Moreover, the fact that lists with *insomma* in POST-C-I position (and not any other position, see Section 3.2) are associated with a categorizing function, with the category explicitly clarified through the POST-C itself, calls for the opportunity to add, in our constructional network, a daughter construction of the categorizing list construction which displays a number of specific properties with respect to other categorizing lists: in this daughter construction, *insomma* is lexically specified, the POST-C is obligatorily expressed, and the semantics and especially pragmatics of the construction is not strictly predictable from the combination of properties inherited from the mother construction plus these specifications.[12]

Indeed, we claim that there is a fundamental distinction on the pragmatic level between lists with and without a post-detailing component, that could be phrased in the following way. As we have seen, all denotation lists more or less explicitly involve some process of categorization, whereby syntactically parallel items are constructed

---

12. See also Example (13) in Section 2.1 and Mauri et al. (2019b) for similar cases.

and processed in discourse as part of the same set. However, the act of categorizing reality in a particular way is a potentially face-threatening act: this is why we have a number of cases where speakers do not produce a post-detailing component, in an attempt to avoid direct categorization; in these cases, speakers seem to prefer the use of vague expressions such as general extenders. Fiorentini (2018a) in fact proposes to link this "indirect" categorization to the pragmatic function of mitigation. On the other hand, we have cases, such as the ones discussed throughout this section, where the speakers do produce the post-detailing component in order to: (i) facilitate syntactic processing of a complex unit that might not be easily accessed as a list in online production, (ii) disambiguate between different possible interpretations of a list, (iii) mark the conclusion of their turn and ease turn-taking. Therefore, the *insomma* categorizing list construction carries an overall pragmatic function of clarification or precision (quite opposite to the vague categorization and mitigation function that characterizes other kinds of lists) as well as "closure".

As mentioned above, this semi-specified construction would be an instantiation of the categorizing list construction, hence a lower node in the hierarchy with a life of its own, which inherits the core formal and functional features from the overarching list construction and at the same time overrides it by specifying its own formal and functional properties. This state of affairs is depicted in Figure 6, which reproduces and enriches Masini et al.'s (2018) representation of the constructional network for lists.
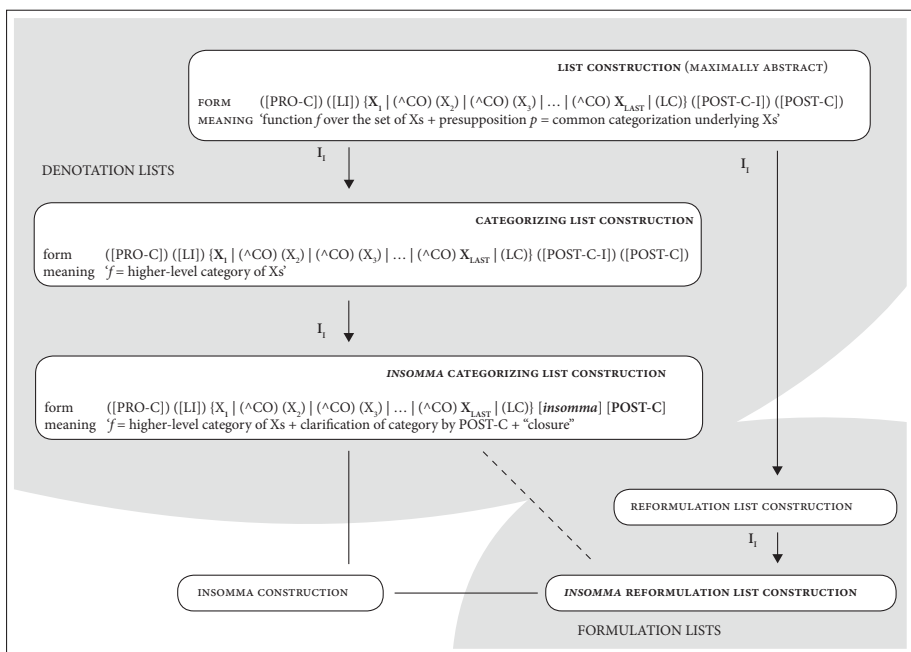

Figure 6. The constructional network of the *insomma* categorizing list construction

As we can see from Figure 6, the *insomma* categorizing list construction would also be linked to some more general *insomma* construction, which cannot be analyzed here. In fact, our proposal paves the way to a separate, thorough investigation of *insomma*, since one of the by-products of our analysis consists in isolating and identifying one precise use of *insomma*, which, as we know, is a highly polyfunctional item. By encapsulating *insomma* within larger constructions (much in line with the methodology proposed by Masini and Pietrandrea 2010 for the equally polyfunctional *magari*), it becomes possible to disambiguate different meanings of *insomma*. The general *insomma* construction would also be linked to the (plausible) reformulation list construction featuring *insomma* as a reformulating marker (schematically: x_1 *insomma* x_{LAST}): the latter is in turn an instantiation of a more general reformulation list construction, which we regard here as a daughter construction of the abstract list construction, on a par with denotation lists, in compliance with a unified approach to grammar and interaction, as advocated for in Section 2.3. Under this view, what have been called denotation lists and formulation lists (the two grey areas in Figure 6) would result from the same maximally abstract mechanism. It is therefore not surprising that specific constructions belonging to this network would share many properties, like the *insomma* categorizing list construction and the *insomma* reformulation list construction (here connected by a dotted line).

Finally, this constructional network, which aims at taking a snapshot of the synchronic situation regarding categorizing lists and the role of *insomma* within lists, might be exploited for a (desirable) diachronic analysis. In particular, given the structural resemblance of the two patterns under discussion – the *insomma* categorizing list construction and the *insomma* reformulation list construction – the role of *insomma* as an introducer of post-detailing component in categorizing lists most likely emerges from its role as a reformulating marker (the most common in use). If it turned out to be true, this situation would strengthen even more the hypothesis that sees denotation and formulation as two domains that are deeply intertwined. We leave this line of diachronic research for future investigation.

4.   Conclusions

The aim of this contribution was to discuss the list as a prominent linguistic device that is used to create categories. Two key aspects have proven crucial for our account: the formal description of the list, and the study of its interactional properties and of how they relate to linguistic structure. In the pursuit of this twofold research roadmap, we drew the foundations of our analysis from two different theoretical paradigms, namely Construction Grammar and Interactional Linguistics.

Far from being competing approaches, these two models prove to be compatible and in fact complementary. A Construction Grammar approach is necessary in order to capture generalizations across instantiations of the list pattern displaying (sometimes great) differences in their surface forms and in their functions. It is therefore a powerful tool in the description of particular types of lists as abstract constructions stored in a speaker's representation of grammar and connected to each other by inheritance links. However, this view cannot be disentangled from a reflection on *how* (the actual instantiations of) such constructions are deployed in real contexts, and *for what purposes*: these are the questions specific to Interactional Linguistics. Furthermore, taking into account evidence from oral corpora provides us, in the end, with a more accurate description of the scrutinized construction, that is, one that is able to include also pragmatic and interpersonal values (see Finkbeiner 2019).

What we have now achieved is a formal representation that takes in greater account formal (and in some cases also functional) similarities between denotation lists (specifically category-building lists) and formulation lists (such as reformulation). The close relationship between categorizing lists and reformulation lists was investigated through a specific case-study on the discourse marker *insomma* in spoken Italian, which allowed us to bridge the gap between the two. *Insomma* is primarily used as a reformulation marker. However, when occurring in a specific position at the end of categorizing lists, *insomma* acquires a new disambiguating function closely related to (and most likely derived from) the reformulation one: indeed, by reformulating an incrementally built list of (often unorganized or partially-organized) linguistic elements, it introduces a general category label, inducing thus a "list-reading" on the previous material.

Finally, based on our exploration of list constructions in spoken Italian, we may draw some conclusions concerning the relationship between this device and categorization. Our analysis of lists suggests that we move away from a rigid cognitive view, whereby categories are seen uniquely as the reflection of some pre-linguistic organization of knowledge in the human mind. The very existence of constructions that have the role of "building" categories points to the need for a more "fluid" and dynamic approach to categorization. Even more crucially, the fact that, in spoken language, lists are temporally organized structures that are adapted to local contexts, often with the participation of more than one speaker, suggests that categorization should be regarded as a process and as a contextually bound activity that is cooperatively carried out by using a number of linguistic resources, like the list construction. This view is also shared by other discourse-based approaches including, for example, discursive psychology: Derek Edwards (1991, 1997) was in fact one of the first to emphasize the idea of categorization as an action. Based on this discussion, categories should be seen as the product of interaction rather than one of its pre-conditions.

Acknowledgements

Funding

References

Ariel, Mira & Mauri, Caterina. 2018. Why use "or"? *Linguistics* 56(5): 939–993. https://doi.org/10.1515/ling-2018-0020

Auer, Peter. 2000. On-line syntax—Oder: Was es bedeuten könnte die Zeitlichkeit der mündlischen Sprache ernst zu nehmen. *Sprache und Literatur* 85(31): 43–56. https://doi.org/10.30965/25890859-031-01-90000005

Auer, Peter. 2009. Projection and minimalistic syntax in interaction. *Discourse Processes* 46(2–3): 180–205. https://doi.org/10.1080/01638530902728934

Auer, Peter. 2015. The temporality of language in interaction: Projection and latency. In Deppermann & Günthner (eds), 27–56. https://doi.org/10.1075/slsi.27.01aue

Auer, Peter & Pfänder, Stefan. 2007. Multiple retractions in spoken French and spoken German. A contrastive study in oral performance styles. *Cahiers de Praxématique* 48: 57–84. https://doi.org/10.4000/praxematique.758

Barotto, Allessandra. 2018. The role of exemplification in the construction of categories: The case of Japanese. *Folia Linguistica* 52: 37–68. https://doi.org/10.1515/flih-2018-0002

Barotto, Allessandra & Mauri, Caterina. 2018. Constructing lists to construct categories. *Rivista di Linguistica* 30: 95–134.

Barsalou, Lawrence W. 1983. Ad hoc categories. *Memory & Cognition* 11(3): 211–227. https://doi.org/10.3758/BF03196968

Barth-Weingarten, Dagmar & Couper-Kuhlen, Elisabeth. 2011. Action, prosody and emergent constructions: The case of "and." In Auer & Pfänder (eds), 263–292. https://doi.org/10.1515/9783110229080.263

Bazzanella, Carla. 1995. I segnali discorsivi. In *Grande grammatica italiana di consultazione, Vol. III: Tipi di frase, deissi, formazione delle parole*, Lorenzo Renzi, Giampaolo Salvi, & Anna Cardinaletti (eds), 225–257. Bologna: il Mulino.

Blanche-Benveniste, Claire. 1990. Un modèle d'analyse syntaxique "en grilles" pour les productions orales. *Anuario de Psicología* 47: 11–28.

Bybee, Joan L. 2006. From usage to grammar: The mind's response to repetition. *Language* 82(4): 711–733. https://doi.org/10.1353/lan.2006.0186

Bybee, Joan L. 2010. *Language, Usage and Cognition*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511750526

Bybee, Joan L. 2013. Usage-based theory and exemplar representation. In *The Oxford Handbook of Construction Grammar*, Thomas Hoffmann & Graeme Trousdale (eds), 46–49. Oxford: OUP.

Bonvino, Elisabetta, Cortés Velásquez, Diego & Fiorenza, Elisa. 2010. "Sopratavola soprammobile come dite voi": Lists in L1 and L2. *Italian Journal of Linguistics* 30(1): 201–230.

Channell, Joanna. 1994. *Vague Language*. Oxford: OUP.

Conte, Maria-Elisabeth. 1996. "Anaphoric encapsulation". In *Coherence and anaphora* (= Belgian Journal of Linguistics, 10) Walter De Mulder & Liliane Tasmowski (eds.), 1–10.

Croft, William. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective* (Reprinted). Oxford: OUP. https://doi.org/10.1093/acprof:oso/9780198299554.001.0001

De Mauro, Tullio, Mancini, Federico, Vedovelli, Massimo & Voghera, Miriam. 1993. *Lessico di frequenza dell'italiano parlato*. Milano: Etaslibri.

De Mauro, Tullio. n.d. 'Insomma'. In *Il nuovo De Mauro*. <https://dizionario.internazionale.it/parola/insomma> (24 March 2021).

Deppermann, Arnulf & Günthner, Susanne (eds). 2015. *Temporality in Interaction* [Studies in Language and Social Interaction 27]. Amsterdam: John Benjamins. https://doi.org/10.1075/slsi.27

Diessel, Holger. 2015. Usage-based construction grammar. In *Handbook of Cognitive Linguistics*, Eva Dąbrowska & Dagmar Divjak (eds), 296–322. Berlin: De Gruyter.

Diessel, Holger. 2017. Usage-based linguistics. In *Oxford Research Encyclopedia of Linguistics*, Mark Aronoff (ed.). Oxford: OUP. https://doi.org/10.1093/acrefore/9780199384655.013.363

Du Bois, Jack. 2014. Towards a dialogic syntax. *Cognitive Linguistics* 25(3): 359–410. https://doi.org/10.1515/cog-2014-0024

Du Bois, John W., Chafe, Wallace L., Meyer, Charles, Thompson, Sandra A., Englebretson, Robert & Martey, Nii. 2000–2005. *Santa Barbara Corpus of Spoken American English*, Parts 1–4. Philadelphia PA: Linguistic Data Consortium.

Edwards, Derek. 1991. Categories are for talking. On the cognitive and discursive bases of categorization. *Theory and Psychology* 1(4): 515–542. https://doi.org/10.1177/0959354391014007

Edwards, Derek. 1997. *Discourse and Cognition*. Thousand Oaks CA: Sage.

Fillmore, Charles J. 1982. Frame semantics. In *Linguistics in the Morning Calm*, The Linguistic Society of Korea (ed.), 111–137. Seoul: Hanshin.

Fillmore, Charles, Kay, Paul & O' Connor, Mary Catherine. 1988. Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language* 64(3): 501–538. https://doi.org/10.2307/414531

Finkbeiner, Rita. 2019. Reflections on the role of pragmatics in Construction Grammar. *Constructions and Frames* 11(2): 171–192. https://doi.org/10.1075/cf.00027.fin

Fischer, Kerstin. 2010. Beyond the sentence: Constructions, frames and spoken interaction. *Constructions and Frames* 2(2): 185–207. https://doi.org/10.1075/cf.2.2.03fis

Fiorentini, Ilaria. 2018a. Eccetera eccetera e così via di seguito. I general extenders dell'italiano contemporaneo. In *Club Working Papers in Linguistics*, Vol. 2, Francesca Masini & Fabio Tamburini (eds), 20–39. Bologna: Circolo Linguistico dell'Università di Bologna (CLUB).

Fiorentini, Ilaria. 2018b. 'E le rimanenti cose'. "Eccetera" tra reticenza e inferenza. In *Spazi bianchi. Le espressioni letterarie, linguistiche e visive dell'assenza*, Alfonsina Buoniconto, Gerardo Salvati & Raffaele Cesaro (eds), 249–260. Soveria Mannelli: Rubbettino.

Fox Tree, Jean E. & Schrock, Josef C. 2002. Basic meanings of *you know* and *I mean*. *Journal of Pragmatics* 34(6): 727–747. https://doi.org/10.1016/S0378-2166(02)00027-9

Fried, Mirjam. 2010. Grammar and interaction. New directions in constructional research. *Constructions and Frames* 2(2): 125–133. https://doi.org/10.1075/cf.2.2.00int

Ghezzi, Chiara & Andorno, Cecilia. 2014. Vagueness markers as politeness strategies in an Italian radio phone-in show. *Beiträge Zur Fremdsprachenvermittlung* 20: 15–40.

Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago IL: University of Chicago Press.

Goldberg, Adele E. 2006. *Constructions at Work*. Oxford: OUP.

Goria, Eugenio. 2020. The discursive construction of categories. Categorisation as a dynamic and co-operative process. *Language Sciences* 81: 101233.
https://doi.org/10.1016/j.langsci.2019.06.003

Goria, Eugenio, & Mauri, Caterina. 2018. Il corpus KIParla: Una nuova risorsa per lo studio dell'italiano parlato. In *Club Working Papers in Linguistics*, Vol. 2, Francesca Masini & Fabio Tamburini (eds), 96–116. Bologna: Circolo Linguistico dell' Università di Bologna (CLUB).

Gumperz, John J. 1982. *Discourse Strategies*. Cambridge: CUP.
https://doi.org/10.1017/CBO9780511611834

Haspelmath, Martin. 2007. Coordination. In *Language Typology and Syntactic Description, Vol. II: Complex Constructions*, Timothy Shopen (ed.), 1–51. Cambridge: CUP.
https://doi.org/10.1017/CBO9780511619434.001

Hilpert, Martin. 2019. *Construction Grammar and its Application to English*, 2nd edn. Edinburgh: EUP.

Hoffmann, Thomas & Trousdale, Graeme (eds). 2013. *The Oxford Handbook of Construction Grammar*. Oxford: OUP. https://doi.org/10.1093/oxfordhb/9780195396683.001.0001

Hopper, Paul J. 1987. Emergent grammar. In *Berkeley Linguistics Society, 13: General Session and Parasession on Grammar and Cognition*, Jon Aske, Natasha Berry, Laura A. Michaelis & Hana Filip (eds), 139–157. Berkeley CA: BLS. https://doi.org/10.3765/bls.v13i0.1834

Imo, Wolfgang. 2015. Interactional Construction Grammar. *Linguistics Vanguard* 1: 69–77.
https://doi.org/10.1515/lingvan-2015-0008

Jefferson, Gail. 1990. List-construction as a task and resource. In *Interactional Competence*, George Psathas (ed.), 63–92. New York NY: Irvington Publishers.

Jefferson, Gail. 2004. Glossary of transcript symbols with an introduction. In *Conversation Analysis. Studies from the First Generation* [Pragmatics & Beyond New Series 125], Gene H. Lerner (ed.), 13–34. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.125.02jef

Kahane, Sylvain & Pietrandrea, Paola. 2012. Types d'entassement en français. *Actes du 3e Congrès Mondial de Linguistique Française. SHS Web of Conferences* 1: 1809–1828.

Kilgarriff, Adam, Baisa, Vit, Bušta, Jan, Jakubíček, Miloš, Kovář, Vojtěch, Michelfeit, Jan, Rychlý, Pavel & Suchomel, Vit. 2014. The Sketch Engine. Ten years on. *Lexicography* 1(1): 7–36.
https://doi.org/10.1007/s40607-014-0009-9

Lerner, Gene H. 1991. On the syntax of sentences-in-progress. *Language in Society* 20: 441–458.
https://doi.org/10.1017/S0047404500016572

Linell, Per. 2005. *The Written Language Bias in Linguistics: Its Nature, Origins and Transformations*. London: Routledge.

Lo Baido, Maria Cristina. 2018. Mitigation via exemplification in present-day Italian: A corpus-based study. *ELUA. Estudios de Lingüística Universidad de Alicante*, (ELUA-2018, Anexo 4).
https://doi.org/10.14198/ELUA2018.Anexo4.02

Masaryk University, NLP Centre. 2011. *enTenTen, LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL)*, Faculty of Mathematics and Physics, Charles University, <http://hdl.handle.net/11858/00-097C-0000-0001-CCDF-8> (24 March 2021).

Masini, Francesca. 2006. Binomial constructions: Inheritance, specification and subregularities. *Lingue e Linguaggio* 5(2): 207–232. https://doi.org/10.1418/23144

Masini, Francesca & Arcodia, Giorgio Francesco. 2018. Listing between lexicon and syntax: Focus on frame-naming lists. *Rivista di Linguistica* 30: 135–172.

Masini, Francesca, Mauri, Caterina & Pietrandrea, Paola. 2018. List constructions: Towards a unified account. *Italian Journal of Linguistics* 30(1): 49–94.
https://doi.org/10.26346/1120-2726-116

Masini, Francesca & Pietrandrea, Paola. 2010. Magari. *Cognitive Linguistics* 21(1): 75–121.
https://doi.org/10.1515/cogl.2010.003

Mauri, Caterina. 2008. *Coordination Relations in the Languages of Europe and Beyond*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110211498

Mauri, Caterina. 2017. Building and interpreting ad hoc categories: A linguistic analysis. In *Formal Models in the Study of Language*, Joanna Blochowiak, Cristina Grisot, Stehanie Durrleman-Tame & Christophe Laenzlinger (eds), 297–326. Berlin: Springer.
https://doi.org/10.1007/978-3-319-48832-5_16

Mauri, Caterina, Ballarè, Silvia, Goria, Eugenio, Cerruti, Massimo & Suriano, Francesco. 2019a. KIParla Corpus: A New Resource for Spoken Italian. In *Proceedings of the Sixth Italian Conference on Computational Linguistics. Bari, Italy, November 13–15, 2019*, Raffaela Bernardi, Roberto Navigli & Giovanni Semeraro (eds). <http://ceur-ws.org/Vol-2481/paper45.pdf> (24 March 2021).

Mauri, Caterina, Goria, Eugenio & Fiorentini, Ilaria. 2019b. Non-exhaustive lists in spoken language. *Constructions and Frames* 11(2): 290–316. https://doi.org/10.1075/cf.00032.mau

Mosegaard Hansen, Maj-Britt. 1998. The semantic status of discourse markers. *Lingua* 104(3–4): 235–260. https://doi.org/10.1016/S0024-3841(98)00003-5

Östman, Jan-Ola. 1999. Coherence through understanding through discourse patterns: Focus on news reports. In *Coherence in Spoken and Written Discourse* [Pragmatics & Beyond New Series 63], Wolf Bublitz, Uta Lenk & Eija Ventola (eds), 77–100. Amsterdam: John Benjamins.
https://doi.org/10.1075/pbns.63.08ost

Östman, Jan-Ola. 2005. Construction discourse: A prolegomenon. In Östman & Fried (eds), 121–144. https://doi.org/10.1075/cal.3.06ost

Östman, Jan-Ola & Fried, Mirjam (eds). 2005. *Construction Grammars. Cognitive Grounding and Theoretical Extensions* [Constructional Approaches to Language 3]. Amsterdam: John Benjamins. https://doi.org/10.1075/cal.3

Overstreet, Maryann. 1999. *Whales, Candlelight, and Stuff Like That: General Extenders in English Discourse*. Oxford: OUP.

Pekarek-Doehler, Simona. 2011. Emergent grammar for all practical purposes: The on-line formatting of left- and right dislocations in French conversation. In *Constructions: Emerging and Emergent*, Peter Auer & Stefan Pfänder (eds), 45–87. Berlin: Mouton de Gruyter.
https://doi.org/10.1515/9783110229080.45

Pfeiffer, Martin. 2017. The syntax of self-repair in German: An explanatory model. *Journal of Pragmatics* 119: 63–80. https://doi.org/10.1016/j.pragma.2017.08.001

Selting, Margret. 2007. Lists as embedded structures and the prosody of list construction as an interactional resource. *Journal of Pragmatics* 39(3): 483–526.
https://doi.org/10.1016/j.pragma.2006.07.008

Selting, Margret & Couper-Kuhlen, Elizabeth (eds). 2001. *Studies in Interactional Linguistics* [Studies in Discourse and Grammar 10]. Amsterdam: John Benjamins. https://doi.org/10.1075/sidag.10

Schegloff, Emanuel A. 1979. The relevance of repair to syntax-for-conversation. In *Syntax and Semantics, 12: Discourse and Syntax*, Talmy Givón (ed.), 261–286. New York NY: Academic Press.

Voghera, Miriam. 2013. Tipi di "tipo" nel parlato e nello scritto. In *Di linguistica e di sociolinguistica. Studi offerti a Norbert Dittmar*, Immacolato Tempesta & Massimo Vedovelli (eds), 185–195. Roma: Bulzoni.

Voghera, Miriam, Iacobini, Claudio, Savy, Renata, Cutugno, Francesco, De Rosa, Aurelio & Alfano, Iolanda. 2014. VoLIP: A searchable Italian spoken corpus. In *Complex Visibles Out There*, Ludmila Veselovská & Markéta Janebová (eds), 628–640. Olomouc: Palacký University.

Wälchli, Bernhard. 2005. *Co-compounds and Natural Coordination*. Oxford: OUP. https://doi.org/10.1093/acprof:oso/9780199276219.001.0001

Waltereit, Richard. 2006. The rise of discourse markers in Italian: A specific type of language change. In *Approaches to Discourse Particles*, Kerstin Fischer (ed.), 61–76. Oxford: Elsevier.

Wilson, Deirdre & Carston, Robyn. 2007. A unitary approach to lexical pragmatics: Relevance, inference and ad hoc concepts. In *Pragmatics*, Noel Burton-Roberts (ed.), 230–259. Houndmills: Palgrave. https://doi.org/10.1057/978-1-349-73908-0_12

Are new words predictable?

A pilot study on the origin of neologies by means of natural selection

Dietmar Zaefferer

Ludwig-Maximilians University Munich

Research on neologisms has so far been dominated by retrospective investigations (cf. e.g. Schmid 2008; Kerremans & Prokić 2018). By contrast, the present study is probably the first to directly tap into speakers' linguistic competence by eliciting names for unfamiliar entities. Building on early assumptions by Darwin (1859, 1871) and others an evolutionary, fitness-based approach is developed to predict und thus explain the selections made by groups of German speakers of successful neologisms from self-generated pools. Four main factors are identified as crucial: quantity (brevity), quality (truthfulness), relation (distinctiveness) and manner (attractiveness). In order to test the viability of the theory an experiment is conducted and replicated. The results provide preliminary support for the assumptions made.

Keywords: language change, prediction, experimental linguistics, lexicon, neologisms, Darwinian evolution, natural selection, fitness

1. Introduction

1.1 Predictive language theories and their moving target

Arguably one of the most fundamental questions in linguistic theorizing is this: why are human languages the way they are? The correct answer is not difficult to find: because they became to be that way. But this only leads to the next question: how did this happen? What are the forces that determined their prehistory and history up to their present stages? And this is a deep question that probably will never find a definite and complete answer. But scientific progress consists in giving provisional and partial answers, answers that slowly are moving towards higher degrees of reliability and completeness.

Whereas the evolution of full-fledged human languages from more primitive communication systems is still bound to be highly speculative (in spite of the considerable progress that has been made in the last decades),[1] the constantly ongoing dynamics of languages is open to less conjectural scientific investigation: based on the observation of changes in the past, theories can be devised in order not only to explain those historic changes and similar changes in progress, but also to predict changes to come.

Research on neologisms has so far been dominated by retrospective investigations (cf. e.g. Schmid 2008). With the expansion of the World Wide Web the detection of new words has become an interesting strand of neologism research (cf. Kerremans & Prokić 2018). In both cases only pre-existing products of language competence are being investigated. Only very recently the concept of 'Predictive Lexicology' (cf. Veale & Butnariu 2006) has entered the stage in the domain of computational creativity. By contrast, the study to be presented below is to do with natural creativity, and as such it is probably the first to directly tap into the linguistic competence by eliciting new names for unfamiliar entities.

Similarities with biological evolution have already been pointed out by Darwin, but the helpfulness of this analogy is still a controversial issue (cf. Pinker & Bloom 1990; Mesoudi et al. 2006; Progovac 2019). The present pilot study proves that at least in the area of neologisms the comparison with biological natural selection fosters theory building in that it leads to falsifiable predictions about the fate of names for new phenomena. It does so by interpreting fitness as preference in the context of selection from a pool: Given a set of competing labels for a new category, the one that survives as a result of its being accepted by the community can be assumed to be the most preferred one. Conversely, starting from a theoretical preference ranking it should be possible to predict acceptance by a community at least to a certain degree.

These considerations naturally lead to the following research question: What are the factors that determine the preference ranking of competing labels in a community and hence the selection of the survivor? Once they are found, new words become predictable in the sense that in a given pool of candidates for naming a new phenomenon, those that are more likely to win the competition can be distinguished from the others.

If the theory is successful it still leaves open other aspects of the predictability of new words, first and foremost the emergence of new phenomena to be labeled (which of course for the most part is a question linguistic theory is neither called for nor capable to answer), and second the factors that determine the creation of

---

1. Cf. the proceedings of the biennial International Conferences on the Evolution of Language (EVOLANG) that took place twelve times so far since 1996.

the pool (which is a question only a joint effort of linguistics and psychology can answer), but at least a first step will have been made towards a more comprehensive theory of the ways neologisms emerge, and a proof of concept will have been given that making new words predictable is not a completely hopeless enterprise.

1.2 Language change and neology

Languages and words are artifacts of a special kind insofar as in everyday life they are perceived as something natural rather than artificial. On the other hand, words are those elements of language whose nature as artifacts becomes most easily visible, especially when new categories require the creation of new linguistic forms or of new readings of existing ones. This is regularly the case when new phenomena such as new artifacts are to be dealt with in a community and become important enough for the need to agree on its categorization and linguistic coding to arise.

Given this hybrid nature of language as a quasi-natural social artifact, the creation and spread of a label for an unfamiliar category can be assumed to combine conscious reflection and unconscious processes of adaptation both within and across individuals.

The central idea of the investigation to be presented here was to probe experimentally into the identity and relative weight of those factors that determine the form and early fate of words created for communicating about novel categories. It is a pilot study in the sense of a pretest of an unprecedented research procedure.

The dynamics of natural living languages can be compared to that of glaciers: at first glance they look motionless, but upon closer inspection they turn out to be in constant motion, snowfalls add new substance which gets transformed into ice, the ice moves slowly downward, ice blocks fall from their terminus, and substance gets lost as melt water runs off.

Similarly, in the body of a living language four regions corresponding to four kinds of language change can be distinguished:

1. The core region comprises those elements that stay the same in the slow movement of the whole: *Zero language change* is the limiting case of linguistic dynamics, and it is of crucial importance for the functionality of language.[2]
2. The lateral regions contain those elements that undergo modification: *Transforming language change* applies to either form or function or both form and function.

---

2. "Every language changes at a rate which leaves contemporary persons free to communicate without disturbance." (Bloomfield 1926: 162)

3. The lower region holds the elements that go out of existence ('melt away'): *Subtractive language change* is to do with the forms or functions or form-function pairs that become obsolete and finally disappear.

4. The upper region is where elements come into existence ('new snow accumulates'): *Additive language change* happens when fresh elements enter the language.

The focus of this paper, of course, is on additive change, but the other regions are relevant as well: Most morphemes used in neologisms come from the stable core and move to the lateral regions of transformative language change in that their meaning potential is modified by the addition of a new, figurative usage. Finally, as is well known, successful new words that compete with pre-existing ones can push down the former to the lower region by marginalizing and ultimately ousting them (subtractive change).

An important question that forms the backdrop of all neologism research is: Why do new elements of language emerge at all? I submit that the most general answer is one that is best illustrated by an economical metaphor: because there is demand for them. Although demand can also come from various minor sources such as taboo avoidance,[3] the strongest motivation for demand is provided by new phenomena that come into existence, requiring thus that new categories be built for dealing with these phenomena. In order to communicate about such novel categories, in the beginning deictic expressions ('the thing over there') or lengthy phrases ('the gizmo with a blue ball on top and three feet') will suffice, but in line with Zipf's Law this becomes too cumbersome as soon as communication about a new category occurs sufficiently often and becomes sufficiently relevant for a community. This is when demand for a compact way of dealing with the concept exceeds the relevant threshold value and a new linguistic element is born.

These assumption leads to the first working hypothesis the present study is based on:

H1 Market of Words Hypothesis:
Neologisms are supplied if and when there is sufficient demand.

The *Market of Words Hypothesis* leads directly to our central research question: how is this demand satisfied? The answer to this question is given by our second working hypothesis:

---

3. Cf. e.g.: "It is a likely speculation that the Norman French title Count was abandoned in England in favour of the Gemanic Earl […] precisely because of the uncomfortable phonetic proximity to *cunt*" (Hughes 1991: 20).

H2 Selection of Fittest Hypothesis:
Neologism supply takes place in two steps: (a) a candidate pool of possible names is generated, and (b) a selection process eliminates all candidates with an insufficient degree of fitness such that only one (rarely more) candidate survives and wins.[4]

At this point a new question arises: Is this artificial selection or natural selection? Before addressing this issue, it seems appropriate to embark on a short excursion into the history of science regarding the relation between Darwinian evolution and language change.

1.3 Language evolution: From Darwin via Haeckel, Schleicher and Müller back to Darwin

In November 1861 the famous and highly influential German biologist Ernst Haeckel[5] 'buries himself' (his own words in a letter to his fiancée) in the German translation of Darwin's (1859) masterpiece *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life* (Richards 2008: 68). Excited about the new ideas he urges his friend August Schleicher, the famous linguist (who earlier had introduced tree diagrams for language families) to read it too. Schleicher, immediately excited as well, reacts with an open letter to Haeckel that comments on the intriguing analogies between the evolution of species and the evolution of languages (*Die Darwinsche Theorie und die Sprachwissenschaft* [Darwin's Theory and Linguistics], Schleicher 1863).

The loop circles back to England when Schleicher's text is translated under the header "Darwinism tested by the science of language" (a title echoed by Aronoff 2017) and published as Schleicher 1869. Max Müller, the famous German philologist in Oxford, reviewed it approvingly in the very first number of *Nature* (Müller 1870). Most interestingly in view of the present topic, Müller points out a parallelism overlooked by Schleicher: "A much more striking analogy, therefore, than the struggle for life among separate languages, is the struggle for life among words and grammatical forms which is constantly going on in each language. Here the better,

---

4.   This is the well-known avoidance of narrow synonyms ("je ne crois pas qu'il y'ait de mot synonime dans aucune Langue [I do not believe that there is a synonymous word in any language]." Girard 1718: xxx [roman number 30]). In line with our approach there have been attempts at explaining this avoidance in evolutionary terms, e.g., by Aronoff (2016), who draws on Gause's principle of competitive exclusion: "… as a result of competition two similar species scarcely ever occupy similar niches …" (Gause 1934: 19); cf. also Goldberg 2019: 26.

5.   Famous among other things for coining the neologism 'ecology' (Haeckel 1866: 286).

the shorter, the easier forms are constantly gaining the upper hand." Moreover, Müller reveals himself as a functionalist: "… though we cannot in every instance explain the causes of victory and defeat, we still perceive, as a general rule, that those words and those forms carry the day which for the time being seem best to answer their purpose." (Müller 1870: 257)

In *The descent of man, and selection in relation to sex*, his second landmark book on evolutionary theory, Darwin includes a section on language (1871: 85ff.), where he quotes (somewhat inaccurately) from Müller's review of Schleicher's open letter:

> We see variability in every tongue, and new words are continually cropping up; but as there is a limit to the powers of the memory, single words, like whole languages, gradually become extinct. As Max Müller[69] has well remarked: – 'A struggle for life is constantly going on amongst the words and grammatical forms in each language. The better, the shorter, the easier forms are constantly gaining the upper hand, and they owe their success to their own inherent virtue.' To these more important causes of the survival of certain words, mere novelty and fashion may be added; for there is in the mind of man a strong love for slight changes in all things. The survival or preservation of certain favoured words in the struggle for existence is natural selection.   [[69] 'Nature,' January 6th, 1870, p. 257]

This should suffice for motivating the short excursion into the history of science. Almost all the relevant topics of the present paper are already addressed here: the perpetual emergence of lexical innovations as well as the causes of the survival of certain words, such as the preference for better, shorter, easier words (Müller), and the love for novelty and fashion (Darwin) at work in the lexicon. In honor of these outstanding scholars the former idea will be called *Müller's Hypothesis* and the latter one *Darwin's Hypothesis*.

H3 Müller's Hypothesis:
In their constant struggle for life the better, the shorter, the easier words and forms are the ones that survive.

H4 Darwin's Hypothesis:
In addition to quality, briefness and ease mere novelty and fashion contribute to the survival of words and forms.

One more topic that is crucial for the present discussion is addressed by Müller, who in his Schleicher review adds the following caveat (Müller 1870: 258):

> But these analogies should not be carried too far. At all events we should never forget that, if we speak of languages as natural productions, and of the science of language as one of the natural sciences, what we chiefly wish to say is, that languages are not produced by the free-will of individuals, and that if they are works of art, they are works of what may be called a natural or unconscious art – an art in which the individual, though he is the agent, is not a free agent, but checked

> and governed from the first breath of speech by the implied co-operation of those to whom his language is addressed, and without whose acceptance language, not being understood, would cease to be language.

This takes us back to the question asked at the end of the preceding section: Is the selection of neologies artificial selection or natural selection? Darwin's answer is unmistakable: "The survival […] of […] words […] is natural selection." But Müller warns us that biology and works of art are different in that only the latter are the product of agents, and that the common denominator in the selection processes is only the absence of consciousness. But is human agency not also the hallmark of what inspired Darwin's theory, namely artificial as the opposed natural selection? Does Müller not imply that what is at work in language is in a way not really natural, but artificial selection?

And indeed, a closer look at Darwin's terminology seems to support this view. Discussing the category of artificial selection in *The origin of species* he makes an often ignored important distinction (emphasized by Sterrett 2002) between what he calls 'methodical selection,' which is based on conscious, intentional action, and what he calls 'unconscious selection,' which although also being based on conscious, intentional action, is an unintended consequence of it on the social level:

> I can see no more reason to doubt this [that the swiftest and slimmest wolves would have the best chance of surviving], than that man can improve the fleetness of his greyhounds by careful and *methodical selection*, or by that *unconscious selection* which results from each man trying to keep the best dogs without any thought of modifying the breed.   (Darwin 1859: 90f.; emphasis added)

Since the creation, selection and survival of new words is clearly not a matter of biological, but of cultural evolution (cf. Richerson & Boyd 2005; Mesoudi et al. 2006), the term pair methodical selection and unconscious selection could be adopted, the former for cases of intentional choice (e.g. "The government doesn't like 'inequalities' in health. Even the word is banned: 'variations' is the acceptable word." Smith 1997: 51), and the latter for processes in which several competing names are used in a community until a winner emerges.

But since there is a common denominator of biological and cultural natural selection in that both are non-intentional uncontrolled processes, I will not speak of methodical selection and unconscious selection, but rather extend the use the term pair artificial selection and natural selection to the domain of cultural evolution, seemingly vindicating Darwin's verdict that the survival of words is natural selection (although he never wrote about cultural evolution). The resulting picture comprises thus four categories: Artificial and natural biological selection on the one hand, and artificial and natural cultural selection on the other. The use of 'natural selection' in the title of this paper means of course the last-mentioned category.

1.4   The incredible career of a neologism and Engelbart's Question

During presentations of preliminary results of this study the author used to challenge his audience as a little warm-up exercise with the question and picture shown in Figure 1.

What is this?

Figure 1. Picture of an unfamiliar object

Since nobody knew the correct answer, he gave it himself, first in the following form:

(1)   *X-y* position indicator for a display system.[6]

Then he presented other possible formulations of the correct answer:

(2)   Position indicator for a display system

(3)   Visual display input device

(4)   Cursor position controller for a display device

(5)   Apparatus for controlling movement of a curser [sic!] in a computer display system

(6)   Optical cursor control device

(7)   Cursor control device for use with display systems[7]

None of them seemed to be especially enlightening. So, one more answer was added that solved the riddle:

---

6.   This was the name used by Engelbart in patent US 3541541 A, filed on Jun 21, 1967. <https://patents.google.com/patent/US3541541A/en>

7.   All examples are titles of other patents to be found in the text of the abovementioned patent as citing it.

(8)   Computer mouse

Together with the device it denotes this neologism made an incredible career: in 1981 the compound *mouse click* (the second member was chosen for disambiguation) appeared only once in the 46,107 books Google's Ngram Viewer is based on for this year, twenty years later 'mouse click' occurs as many as 2,535 times in the 104,147 books sifted through in 2001.

How did this incredible career start off? Here is what is known about it:

In the early sixties Doug Engelbart and his collaborators at Stanford Research Institute were testing a two-wheeled wooden box (Figure 1) that was supposed to complement the functions of the computer keyboard and that later was to be patented under the formal name '*x-y* position indicator for a display system.' Several names were proposed.

> We set up our experiments and the mouse won in every category, even though it had never been used before,' said Mr Engelbart. 'It was faster, and with it people made fewer mistakes. Five or six of us were involved in these tests, but no one can remember who started calling it a mouse. I'm surprised the name stuck.' […] 'We thought that when it had escaped out to the world it would have a more dignified name,' said Mr Engelbart. 'But it didn't.' (Beaumont 2008: 1)

Engelbart had no viable theory of the form and fate of names for new categories. With such a theory he wouldn't have been surprised that the name stuck. In honor of Douglas Engelbart, the underappreciated pioneer of augmenting the human intellect and boosting collective IQ (Engelbart 1995) who died in 2013, the following neology based on his name will be used:

Q1 Engelbart's Question:
Why do some neologies stick, whereas others don't?

Using this neology, the purpose of the present study can also be characterized as trying to give an empirically supported answer to Engelbart's Question.

1.5   Naming and categorization: Clarifications and sample case observations

The story of the computer mouse and its collective christening is told here among other things in order to give a real-life illustration of the issues to do with the relation between categorization and naming, both individually by an agent and interactively by a group. Before coming back to the example a few clarifications may be helpful in preventing possible misunderstanding.

First, it is worth emphasizing that naming and categorization are not the same. As the focus of this paper is on naming it is certainly not central to the topic of the

present volume. But since categorization is difficult to access without looking at language, and since naming and categorization are closely related the interactive naming processes studied here will offer also some insights into the categorization processes involved.

Second, whereas categorization is most of the time understood as assignment to an already established category it is clear that this volume is rather to do with categorization in the sense of category formation, i.e., the creation of entities that allow for categorization in the former sense to occur in the first place. The same holds for the senses of naming: in most contexts, for instance in psychological naming tests, the task consists in finding a pre-existing name that fits the stimulus. In the present context naming is the creation of new names, i.e., name formation.

Third, concepts and categories are taken here to be abstract entities that are different both from their instantiations (or sets thereof) and their (neural) representations in individual minds. Instead, the following definition will be assumed: for an agent or a group of agents to dispose of a concept or category is equivalent to being able in the context of a given purpose (a) to distinguish between instantiations and non-instantiations of the concept and (b) to neglect the distinctions between different instantiations of it. A very simple example is the thermostat, which disposes of three temperature categories: inside the desired range, above it, and below, because it regards the difference of temperatures in distinct categories and disregards temperature differences within the same category.

Returning to the case of the computer mouse it obviously belongs to the category of artifacts and more specifically to the tools. As with all tools its closer categorization has to be in functional terms. Since the task of Engelbart's team was defined in functional terms (to develop a device that allows for controlling the position of a movable mark on a computer screen) there can be no doubt that every team member disposed of a clear and shared concept of it. The words that were used to refer to it before being ousted by the animal metaphor are not known, what is known, however, is that there was a competition not only among expressions, but also among devices: "A total of five different hardware devices were tested [...] The devices included the light pen, and four types of bug-positioning devices: a joystick, a 'mouse', a Grafacon, and a knee control." (English et al. 1965: 41)

Interestingly, the quote mentions another neology, 'bug' in the sense of the mark on the screen the positioning device points at. Since it had the shape of a plus sign, the motivation by similarity with a little animal is quite transparent.[8] In this hardware competition situation for a team member to speak of a bug-positioning device was not enough to refer specifically to what was later to be called a mouse:

---

8.   Nevertheless, as is well known, this use of the insect metaphor was later superseded by its reference to the concept of a defect in either hard- or software.

bug-positioning device is the strongest common superordinate concept (*genus proximum*), but there was demand for a more specific concept with a distinctive feature *(differentia specifica)* to tell the new device developed by the group from its competitors. Given the cubical shape of the first mouse (see Figure 1) one can speculate that something like 'the box device' may have been used. Interestingly, the animal metaphor was so successful that already in the abovementioned technical report the nickname "mouse" was used together with the other rather technical designations.

But what could have inspired its creation? In view of the original box shape that was quite different from the round bodies of later mice, the form was probably not the main motivation. The association of the flexible cable coming out from the device's body with the tail of a mouse seems plausible, but since the competitors were connected in the same way the presence of a cable was not distinctive. In the absence of more information every plausible explanation is bound to be speculative, including the following guess: perhaps there was a double motivation by two conspiring factors, the comparatively small size and the motion that it causes in the 'bug.' Given that a mouse is a prototypically small animal and that mice do eat bugs that could have been a motivation. An additional factor may have been the iconic relation between sign and signified constituted by the monosyllabicity and hence minimal size of the word and the tiny size of the animal. This may be the answer to Engelbart's Question in this special case.

And of course after the group consensus about the name had been achieved the spread of its use beyond the Engelbart lab may well have been supported by factors such as the resemblance of the cable and a tail, the similarity of the shape of later mice with that of the animal, and the sympathy many people have for mice (as opposed to rats).[9]

For the present purpose of developing a predictive theory these observations on a sample case may be more or less helpful depending on its degree of prototypicality. In order to get beyond the single case a large amount of additional cases had to be analyzed.

---

9. Other motivations are of course possible. At the Bologna meeting the present volume is based on Larry Barsalou (p.c.) commented that he associates the device with the animal because both typically scurry to and fro.

2.   Investigating form and fate of neologisms: An experimental approach

2.1   Creating demand for names of unknown objects: Meet the Fribbles

In order to come up with an empirical and generalized answer to Engelbart's Question, and in view of the fact that it is impossible to wait in the lab until a new phenomenon appears in the public domain, a new experimental kind of neologism research had to be devised. In the first place, under the assumption that *H1*, the *Market of Words Hypothesis* is correct, the following question had to be answered: how can an experimenter artificially create demand for neologisms in his subjects' minds?

To begin with, stimuli had to be found that were completely unknown and new to the participants. Fortunately, there exists a whole family of this kind of stimuli, the artificial animal-like objects called *Fribbles* (created by Michael J. Tarr, Brown University, www.tarrlab.org; cf. also Barry et al. 2014).[10]

Five of the 36 Blue Fribbles (the three most-different exemplars of all 12 Fribble species) were selected. They are shown in Figure 2.
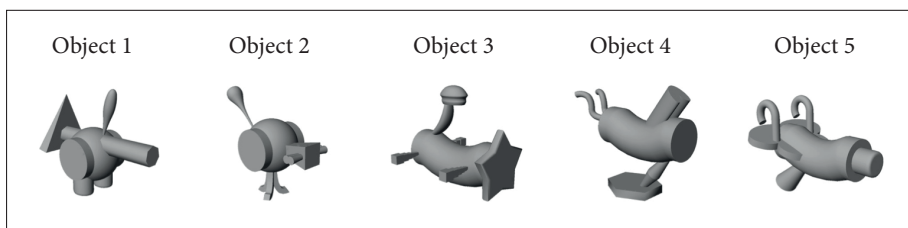
Figure 2. The five selected unfamiliar object stimuli

Fribbles make it easy for the experimenter to create demand: all he has to do is to instruct the subjects to think up names for these objects.

2.2   Preparatory data collection: Five objects, five name pools and an initial hypothesis

The project took off with a rather informal first step: nine students (fluent speakers of German, six females and three males) of a seminar on essential statistics for linguists were given twenty minutes to fill a one-page questionnaire with the header 'Naming of unknown objects,' showing the five object numbers as column headers and three empty lines below each one. Slides of the objects were projected

---

10. The author had been made aware of the Fribbles' existence a decade ago by Daniel Casasanto, who had used them successfully himself (Casasanto 2009a).

for approximately four minutes each. Subjects were asked to come up with three possible names they would give the relevant object, without talking to each other. There was no remuneration. Out of the 135 slots, 55 were filled with mostly novel German words. Although there was considerable overlap concerning the roots (the eleven proposals for Object 3 all contained the German root *Stern* 'star'), there was no overlap at all between the full names.

The preparatory pool for Object 1 contained 17 names, among them *Hammerschweinchen* 'hammer piglet' and *Sparschweinkanone* 'piggy bank cannon', with six occurrences of the root 'pig' and three occurrences of the root 'hammer.'

Object 2 yielded only 9 responses, including *Kastenkugel* 'box sphere,' *Kopfhörerhund* 'earphone dog' and *Ohrenvogel* 'ear bird.' The root 'bird' occurred three times, the roots 'ear' and 'sphere' twice.'

With eleven different responses the harvest inspired by Object 3 was slightly bigger. As already mentioned, all used the root 'star' as component, ten in modifier position (e.g. *Sterngurke* 'star cucumber,' *Sternstempel* 'star stamp'), and one as head: *Aussichtsstern* 'vista star.'

Object 4 elicited nine words, and there was no overlap at all, neither of words nor of roots. Examples are *Schneckenvogel* ('snail bird') and *flügelloser Schmetterling* ('wingless butterfly').

Object 5 was again given nine names, among them *Hakenzahnpastatube* 'hook toothpaste tube' and *Aufhängesitz* 'hang up seat'. The roots 'hang' and 'tube' occurred three times each, the root 'hook' twice.

2.3 Venturing a first hypothesis

The recurring features of the 55 items in the five preparatory pools were encouraging: they hinted at the possibility that new words are predictable at least in their form (noun-noun compounds are preferred) and the general semantic categories involved. Most roots denote either animals (pig, elephant, bird, snail) or artifacts (hammer, hook, tube, seesaw), and shape-denoting roots (star, sphere, hook) seem to have an important role as well.

Therefore, the following tentative hypothesis was formulated:

H5 Initial Word Prediction Hypothesis:
German names for novel objects are preferred to the degree that they meet the following criteria:
– They have the form of a binary NN compound,
– The head of the compound denotes an animal or artifact and
– The modifier of the compound has a shape-related meaning.

Fitting examples from the five preparatory pools are *Dreiecksstempel* ('triangle stamp'), *Rundtröte* ('round toy trumpet'), *Sternsaurier* ('star dinosaur'), *Schneckenvogel* ('snail bird'), *Hakenhebe* ('hook lifter').

But the preparatory pools were of course by far too small in order to derive predictors for an experiment. In addition to poor size, another shortcoming of the pools was that they originated from the activity of isolated subjects, whereas the creation of new words is not restricted to individuals, rather it is probably most of the time or at least quite often a social process where ideas are exchanged and proposals are compared. And finally, speakers may have a good sense for the difference between their own preferred words and those neologisms that are likely to be accepted in a community.

Therefore, a pretest was designed in order to get rid of these shortcomings by means of involving group processes and of tapping into the intuition of speakers about social preferences that are possibly distinct from their own particular likings.[11]

3. Pretest: Deriving preference hypotheses from estimated natural selection

3.1   Pretest

Five subjects were recruited with the help of the mail service 'Infodienst' of Ludwig-Maximilians University in Munich, three females and two males. They received 9 Euros each for their participation. The session lasted for about 60 minutes and took place in the university's MELESSA lab; it was recorded with the lab's audio and video devices. The experimenter (the author of this paper) was present during the whole session. After signing two consent forms, one for the experiment and one for the recording, the subjects were handed a five-page questionnaire in German, one page per object.

Under the heading *Naming unknown objects* each page was divided into three sections corresponding to three phases:

Phase 1 (non-interactive; 3 minutes):
individual pools are created.

Instruction:
Think up three names for this object and write them down.

---

11.   Note that Lewis' (1969) concept of convention can be seen as being based on a second-order preference: The preference for shared sign use is preferred over possible personal preferences for the use of a sign.

Phase 2 (interactive; 6 minutes):
the individual pools are joined. The resulting items, up to fifteen, constitute the joined pool from which the subjects have to select in phase 3.

Instruction:
Share your suggestions with the group and tell them why they are good. Write down the suggestions of the others and compare.

Phase 3 (non-interactive; 1 minute):
subjects have three lines each for (a) making an artificial selection and (b) estimating a natural selection from the candidates in the joined pool.

Instruction:
Write down (a) the names you would personally prefer to use and (b) those terms that in your opinion would best prevail in a larger community.

Each object picture was projected on the lab's whiteboard for ten minutes, divided into three, six and one minute for the first, second and third phase, respectively; then the picture changed.

In order to merge the projected rankings under 3 (b) it was decided to give the entries in the third column single weight, those in the second one double weight and the winners triple weight. Thus, a unified ranking resulted with e.g. *Hamming* (not an existing common noun, but the root is that of hammer and the suffix *-ing* is very frequently used) getting the weight ten, because two subjects had it in the first position and two in the second one.

The entries resulting from phase 3 (b) served as basis for deriving the features of natural selection, i.e. those properties that separate the winners from the rest: the winners are supposed to show the highest degree of fitness, the runners-up the second highest, and the third-place finishers the third highest.

3.2   Pretest results: Five pools resulting from artificial and estimated natural selection

Already the first cursory look at the data showed such a big difference from what has been the result of the preparatory non-interactive data collection that only one conclusion could be drawn: *H5*, the *Initial Word Prediction Hypothesis* underestimates the inventiveness at work in name creation, especially when interaction comes into play, to such a dramatic extent that this hypothesis had to be completely discarded.

The top three names for each object according to the estimated natural selection are presented in Table 1.

Table 1. Winners and runners-up of the estimated natural selection

O 1:	1. *Hamming*	2. *Trimolin*	3. *Togolino*
O 2:	1. *blue Willi*	2. *Fantaan*	3. *Geometros*
O 3:	1. *Starcopter*	2. *Starnaut*	3. *Burcumber*
		2. *Wormster*	
O 4:	1. *Bumbleblue*	2. *Piggyback*	3. *Pipester*
O 5:	1. *Tubedang*	2. *Mobile Puff*	3. *Poolfun*
	1. *Volereur*	2. *Skyfall Shuttle*	3. *Blue Gun Pro X8*
	1. *Kinderjet*		
	1. *Poolboard*		

There were three binary ties, one with Object 3 and two with Object 5. The latter was also special because there was a five-fold tie in the winner's slot. Strikingly, no knowledge of German is required to understand the proposals except for the modifier of the compound *Kinderjet: Kind* is 'child' and *-er* is the linking element.

Fortunately for the replacement of *H5* by a better hypothesis, the estimated natural selections gave several hints for getting at the real determinants, especially when compared with the artificial selection.

First, items that were too long were estimated to lose out on their competitors. So, quantity, less in terms of roots than in terms of syllables, seems to play a key role.

Second, the ontological categories assumed by *H5* were much too specific. So, a more abstract and general way of accounting for the descriptive aspect of the candidates had to be found.

Third, the by far most important change on the way from *H5* to a new hypothesis based on the results of the pretest consisted in turning a problem into an asset: instead of being considered only an impediment for predictability, the unexpected degree of inventiveness was transformed into an additional predictor.

Thus *H4, Darwin's Hypothesis* mentioned above, has been partially vindicated, except that 'mere novelty and fashion' turned out to belong among the most important causes for the survival of new words, and not among the less important ones, as Darwin surmised.

3.3 Developing a selection predictor: Factors of fitness ranking

In order to come up with a more viable hypothesis than *H5* the following question had to be asked: what are the features that make the difference between the 21 winners listed above and the remaining 49 names? Some examples from the latter pool are shown in Table 2.

Comparing Tables 1 and 2 provided support both for *H3, Müller's Hypothesis* (the better, the shorter, the easier survive), and for *H4, Darwin's Hypothesis* (novelty

Table 2. Sample losers from the estimated natural selection

O 1:	*abstrakte Kanone*	*Dreikopftier*	*blaues Rudikolon*
	'abstract cannon'	'three head animal'	'blue rudicolon'
O 2:	*blauer Vogelstrauß*	*Bolzenstandgerät*	*Edler Klopapierhalter*
	'blue ostrich'	'bolt foot device'	'upscale toilet paper holder'
O 3:	*Wurstkolben*	*Stangenrohr*	*Sternenraupe*
	'sausage piston'	'rod pipe'	'star caterpillar'
O 4:	*Garderraupe*	*Akropferd*	*Skybike*
	'garder caterpillar'	'akro horse'	'sky bike'
O 5:	*Flug-wiege*	*Kinderprise*	*Schwimmlandschaft Delux*
	'flight cradle'	'child pinch'	'swimming pool landscape deluxe'

and fashion foster survival), and thus inspired the following proposal for a word prediction model that consists of four categories.

Echoing Grice (Grice 1975: 45), who was echoing Kant (Kant 1781 [1881]: 71), who in turn had borrowed from Aristotle, these categories will be called Quantity, Quality, Relation, and Manner. In the current context they serve to denote four aspects of fitness for the survival of neologisms:

– Quantity: closeness to optimal length
– Quality: number of fitting object features captured
– Relation: number of distinctive object features captured
– Manner: degree of originality and fashionability

H6 Final Word Prediction Hypothesis
A German name for a novel object is more likely preferred to its competitors
1. the closer its length is to the optimum,
2. the higher the number of object features it codes,
3. the higher the number of distinctive object features it codes,
4. the higher its degree of originality and fashionability.

The next step was one of the most challenging tasks of the project: developing valid operationalizations of these categories.

1. *Quantity*
In order to deal with the second part of *Müller's Hypothesis* ('the shorter ones win') the number of words, roots, morphemes and syllables in the winners and the losers of the pretest pools for each object were counted. The resulting means are shown in Table 3 and visualized in Figure 3.

This confirmed part two of *Müller's Hypothesis* in all possible specifications, but since the difference was most pronounced among syllables the Quantity aspect was operationalized as number of syllables.

Table 3. Length differences between winners and losers of the pretest pools

	Words	Roots	Morphemes	Syllables
winners	1.3	1.8	2.15	2.9
losers	1.34	2.28	2.8	3.86
difference	0.040	0.480	0.650	0.969
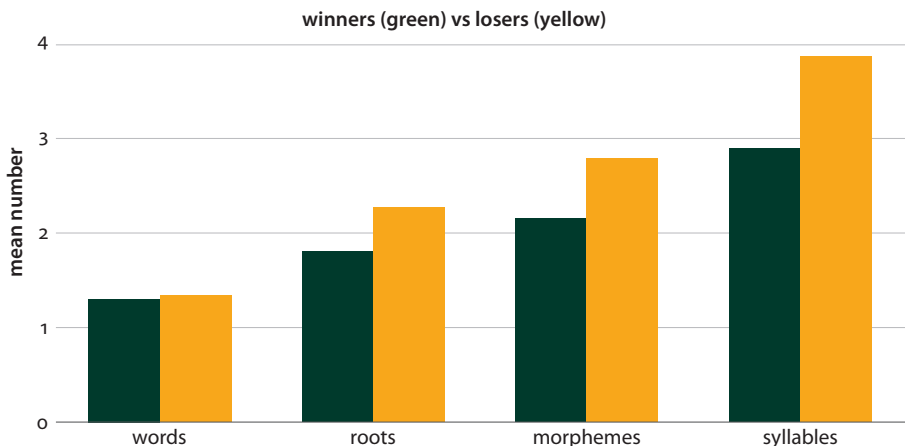
Figure 3. Comparing winners and losers in terms of mean number of units

However, the preference for shorter names is of course not without limitations, else monosyllabic names would be always preferred. This may be true to some extent for English words (see below), but in the German pretest pools it certainly does not hold. A chart of the distribution of length in terms of syllable count among the 70 items in the five pools together is shown in Figure 4.
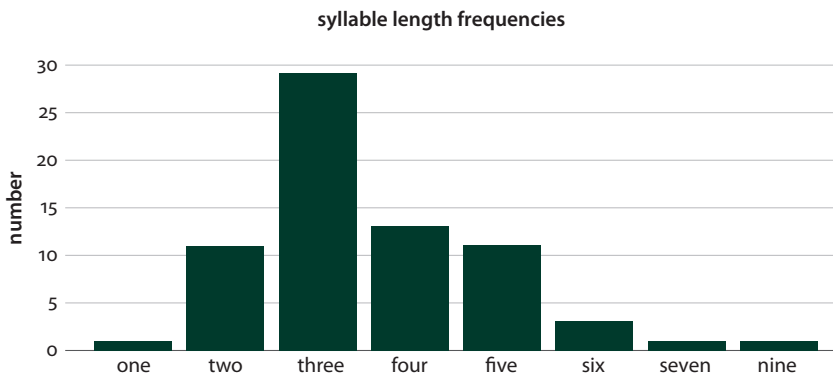
Figure 4. Distribution of syllable counts in the pretest pools

Hence the decision was made against taking the quantity aspect of *Müller's Hypothesis* literally and in favor of interpreting it as a preference for the optimal number of syllables, which according to the sample is three.

2. *Quality*
The third part of *Müller's Hypothesis* ('the easier ones win') was much more difficult to operationalize. It was decided to interpret 'easy' as 'easy to remember' or in other words degree of motivation: for instance, when a star is part of the object, names with 'star' are well-motivated and easy to remember, in short, they fit this feature of the object. The more features a name fits, literally or figuratively, the higher its degree of Quality.

3. *Relation*
The most difficult to operationalize was the first part of *Müller's Hypothesis* ('the better ones win'). Since 'good' can mean anything that increases fitness, it seemed plausible to subsume under Relation the degree to which a name is helpful in discriminating between the five objects. As all objects are blue, the adjective in *blauer Vogelstrauß* 'blue ostrich' fits the color of the object, but it is not distinctive.[12] By contrast, *Three Feet Bird* is distinctive because only Object 2 stands on something similar to three feet.

4. *Manner*
According to *Darwin's Hypothesis* there is a fourth fitness increasing factor, in addition to the three parts of *Müller's Hypothesis*, namely 'mere novelty and fashion.' The chosen operationalization incorporates both originality ('novelty') and number of foreignisms ('fashion') such as anglicisms or italianisms.

The diversity of word formation devices was astonishing. The pool with its 70 elements contained about 37% German nominal compounds such as *Umschalt-knopfgerät* 'toggle button device' or *Formentier* 'forms animal', 30% foreignisms, most of them anglicisms such as *Chick-a-brick* or *Piggyback*, but also italianisms like *Togolino* and one Frenchism: *Volereur*. Quite a few suggested names could not be assigned to any current word formation type, among them *Heta, Trimolin, Rudikolon, Fantaan*, and *Flovotov*. All in all, the degree of lexical heterogeneity was too high to allow for an account of word formation types beyond the very superficial observations just stated.

Coming back to the topic of categorization and its relation to naming, especially to the question what insights into category formation can be gained from an

---

12. Interestingly, but not unexpectedly this noun phrase occurs as a name for both Object 1 and Object 2; it was suggested, of course, by different subjects.

analysis of name building, it is obvious that the first and the last categories, the formal factors Quantity and Manner, are unhelpful. What is more, the comparison of the items in Table 1 (winners) and Table 2 (losers) shows a strong tendency towards preferring short and original names over more transparent, ie, better motivated and distinctive ones. And indeed, there is a built-in tradeoff between the formal and the functional categories: the more features are literally or figuratively captured, the higher the probability that the Quantity restriction is exceeded. In sum, the selection process that makes the difference between those expressions that stick, and the others that don't tends to decrease the degree to which naming processes are helpful in learning about the categorization procedures behind them.

This is one lesson to be learned from the pretest data: if one is interested in the category formation behind the naming, it is more informative to look at the fleeting stars among the neologies, those that disappear immediately after their creation, whenever they are available, than to analyze those expressions that make it into the lexicon.

The other lesson is to do with the general ontological categories of artifact versus living creature. Interestingly, even though the Fribbles are often described as animal-like creatures the suggested animal names like *Kugelfant* 'sphere (ele)phant' are outnumbered by expressions that imply a conceptualization as artifact. This is interesting because, as mentioned above, artifacts are best characterized in terms of their function and not their visual appearance, whereas the Fribbles stimuli only provide visual information such that possible functions have to be guessed from the appearance.

The question whether this can be interpreted as confirming a human tendency towards conceptualizing the environment in terms of affordances or potential functions must be left to a different occasion.

3.4   Assigning degrees of fitness: The coding algorithm

In order to come up with an algorithm that enables different coders to consistently calculate fitness scores, a further working hypothesis was required: in line with the principle of indifference the four categories were given equal weight, i.e. they were assumed to contribute equally to the overall degree of fitness, with an approximate maximum of 10 in each category. Regarding Quantity and Manner this is easy to achieve by setting the maximum to 10 and $8 + x$, respectively. For Quality and Relation things are different, since in principle there is no upper limit. But as mentioned above there is a trade-off between these categories and Quantity: it is difficult to increase the number of fitting elements without increasing the number of syllables. Therefore, the algorithm was designed in a way that practically excluded

Quality and Relation scores higher than 10, such that the overall maximal score was close to 40.

The definition of the complete algorithm for calculating fitness scores is presented in Tables 4a through 4e.

1. *Quantity*

Motivated by the syllable count distribution shown in Figure 4 the coding scheme assigns the maximum of 10 points to words with 3 syllables and decreasing scores with shorter and longer items according to the function defined in Table 4a.

Table 4a. Quantity score as function of syllable count

Syllable number	1	2	3	4	5	6	7	8	9
Quantity score	6	8	10	8	6	4	2	1	0

2. *Quality*

Here the number of literally fitting aspects *lif* is given twice the weight of the number of figuratively fitting aspects *fif* and the result is doubled. Table 4b presents the corresponding formula.

Table 4b. Quality score as function of the number of literally and figuratively fitting aspects

Quality score := $(lif * 2 + fif) * 2$

3. *Relation*

The Relation score can at most equal the Quality score. The formula is the same except that the number of distinctive literally fitting aspects *dlif* and the number of distinctive figuratively fitting aspects *dlif* take the place of *lif* and *fif*, respectively. It is shown in Table 4c.

Table 4c. Relation score as function of the number of distinctive aspects

Relation score := $(dlif * 2 + dfif) * 2$

4. *Manner*

The first factor of the Manner score is the most subjective part of the whole algorithm in asking the coder for an *edo*, an estimate of the degree of originality (Darwin's 'novelty') on a scale ranging from 0 to 8. The second factor *nfe* simply counts the number of foreign elements. Table 4d presents the corresponding formula.

Table 4d. Relation score as function of originality degree and foreignisms

Manner score := *edo + nfe*

Table 4e shows the overall fitness score formula, which is simply the sum of the four category scores.

Table 4e. Fitness score as function of quantity, quality, relation, manner

Fitness: = Quantity + Quality + Relation + Manner

4.   Experiment: Testing the fitness rank predictor with forced natural selection

4.1   Experiment

The procedure was the same as in the pretest (cf. Section 3.1. above), except that this time six new subjects were recruited, four females and two males, and that the five-page questionnaire they were given had been modified in several partly crucial respects.

After the title *Naming unknown objects* each page was again divided into three sections corresponding to three phases (*n* is a number ranging from 1 to 5):

Phase 1:
Thinking up suggestions (non-interactive; 2:30 minutes)

Instruction:
How would you want to talk about Object *n*? Come up with three names for this object and write them down.

Phase 2:
Comparing proposals (interactive; 5 minutes)

Instruction:
Share your suggestions with the group and tell them why they are appropriate for communicating about Object *n*. Write down the suggestions of the others and compare.

Phase 3:
Finding a consensus proposal (interactive; 2:30 minutes)

Instruction:
Write down in column (a) the three top names you would recommend to a larger community for communicating about Object *n*. Compare and discuss the proposals until you find a consensus. Write down the consensus recommendation in column (b).

Each object picture was projected on the lab's whiteboard for ten minutes altogether, two and a half minutes for the first, five minutes for the second, and again two and a half minutes for the third phase; then the picture changed.

Resulting thus from a process of forced natural selection, the entries obtained in phase 3 section (b) were the ones to be used for testing the quality of the fitness rankings developed on the basis of the pretest data.

4.2   Comparing observed and predicted rankings

4.2.1   *Object 1 experiment ranking*

For the first object only two of the six subjects came up with the full range of three proposals, three provided two and one participant a single one, such that the result were 13 names in 18 cells (72%). The group reached a consensus only for the first rank and left open the other ranks. The winner was *Walking Cam*. The fitness scores provided by the algorithm presented in Section 3.4. above for the thirteen names are shown in Table 5 and range from 8 to 28, yielding ten ranks (there are three ties). And the item with rank one is *Walking Cam*: perfect match!

Table 5. Experiment fitness scores and ranks for Object 1

fitness score	28	24	23	22	18	18	15	15	14	13	10	10	8
predicted rank	1	2	3	4	5	5	6	6	7	8	9	9	10
observed rank	1	–	–	–	–	–	–	–	–	–	–	–	–
difference	0	–	–	–	–	–	–	–	–	–	–	–	–

4.2.2   *Object 2 experiment ranking*

The second object was seemingly harder to deal with than the first one as witnessed by the records: only 11 of the 18 cells were filled (61%) and the final listings by the six subjects of three top candidates, where they were noted, showed some disagreement regarding the order. Three subjects assumed the consensus list to be (1) *Geometrievogel* 'geometry bird,' (2) *Space-Kugler* 'space blowfish' and (3) *Rundständer* 'round pedestal,' but one put *Space-Kugler* in first position and no other names in second and third place, whereas the other two participants left all three positions open. Therefore the algorithm used in drawing conclusions from the pretest (end of Section 3.1. above) was reactivated and triple, double and single weight was assigned to the first, second and third position, respectively. The result is a tie in the first rank (*Geometrievogel, Space-Kugler*), one second ranking word (*Rundständer*) and no third rank.

As documented in Table 6, there is a perfect match of the predicted and observed ranks for *Space-Kugler* and *Rundständer*. The predicted fourth rank for *Geometrievogel* underrates the observed first rank by 3.

Table 6. Experiment fitness scores and ranks for Object 2

fitness score	25	25	24	22	22	20	20	17	12	12	6
predicted rank	1	1	2	3	3	4	4	5	6	6	7
observed rank	–	1	2	–	–	–	1	–	–	–	–
difference	–	0	0	–	–	–	–3	–	–	–	–

4.2.3   *Object 3 experiment ranking*

Data harvest was much more bountiful for the third object: its picture elicited 16 out of the 18 possible proposals (89%) and the consensus ranking was unanimous.

The fitness scores listed in Table 7 show an unusually gentle slope: They predict two tied winners and seven tied runners-up. The observed ranks, however, are quite different, and there is no perfect match at all. The winner *Sternentaucher* 'star diver' is only minimally underrated by the prediction, the third-ranking *Schnuppenstern* 'snuffing star' is overrated by two ranks, but *Galaxieplattform* 'galaxy platform,' that finished second, is rather strongly underrated, namely by 5.

Table 7. Experiment fitness scores and ranks for Object 3

fitness score	28	28	26	26	26	26	26	26	26	24	23	22	20	16	10	6
predicted rank	1	1	2	2	2	2	2	2	2	3	4	5	6	7	8	9
observed rank	3	–	1	–	–	–	–	–	–	–	–	–	–	2	–	–
difference	+2	–	–1	–	–	–	–	–	–	–	–	–	–	–5	–	–

4.2.4   *Object 4 experiment ranking*

The fourth object was apparently less inspiring for the group: they produced only 13 out of 18 (72%) possible names. Except for one dissenting vote the consensus ranking was as follows: 1. *Breakdancer*, 2. *Insektenständer* 'insect stand' and 3. *Krauler* 'crawler.'

Table 8 shows that observed and predicted ranks again form a perfect match, as with Object 1, but this time with three pairs.

Table 8. Experiment fitness scores and ranks for Object 4

fitness score	26	25	24	22	20	17	16	14	14	12	12	10	8
predicted rank	1	2	3	4	5	6	7	8	8	9	9	10	11
observed rank	1	2	3	–	–	–	–	–	–	–	–	–	–
difference	0	0	0	–	–	–	–	–	–	–	–	–	–

4.2.5 Object 5 experiment ranking

Even fewer names were proposed for Object 5, namely only 12 out of 18 (66%). The group agreed (again with one exception) on the following top candidates: (1) *Schlauchdüse* 'hose nozzle,' (2) *Stöpsler* 'plugger' and (3) *Tubenelefant* 'tube elephant.' Due to the dissenting vote the latter rose to second rank, yielding a tie with *Stöpsler*.

The fitness scores shown in Table 9 predict three tied winners: *Schwimmbadtube* 'swimming pool tube' and *Henkelröhre* 'handle pipe' in addition to *Schlauchdüse*. Only the latter matches the observed winner in a perfect match. Regarding the tied second positions the prediction hit the mark with *Stöpsler* and underrated slightly (by 1) *Tubenelefant*.

Table 9. Experiment fitness scores and ranks for object 5

fitness score	28	28	28	22	21	20	20	18	17	14	13	12
predicted rank	1	1	1	2	3	4	4	5	6	7	8	9
observed rank	–	–	1	2	2	–	–	–	–	–	–	–
difference	–	–	0	0	−1	–	–	–	–	–	–	–

4.2.6 Overview of experiment rankings

Table 10 shows the main results of the experiment: given the total number of 13 pairs for comparison, and counting over- and under-rating score differences of 1 as near matches and bigger differences as strong mismatches, there are 8 full

Table 10. Experiment results for objects 1 to 5

Names by object	Syll	Quan	Qual	Rel	Mann	Fitn	Pre R	Obs R	M
Walking Cam	3	10	4	4	10	28	1	1	Full
Geometrie-Vogel	6	4	6	4	6	20	4	1	u_3
Space-Kugler	3	10	6	4	5	25	1	1	Full
Rundständer	3	10	8	4	2	24	2	2	Full
Sternentaucher	4	8	6	6	6	26	2	1	u_1
Galaxie-Plattform	5	6	2	0	8	16	7	2	u_5
Schnuppenstern	3	10	6	4	8	28	1	3	o_2
Breakdancer	3	8	4	4	10	26	1	1	Full
Insektenständer	5	6	6	6	7	25	2	2	Full
Krauler	2	8	4	4	8	24	3	3	Full
Schlauchdüse	3	10	6	6	6	28	1	1	Full
Stöpsler	2	8	4	4	6	22	2	2	Full
Tubenelefant	5	6	6	4	5	21	3	2	u_1
MEAN	3.615	8	5.230	4.153	6.692	24.07			u_0,2
SD	1.211	1.921	1.475	1.459	2.089	3.407			

matches (62%) and 10 full or near matches (77%), corresponding to 5 mismatches (39%) and only 3 strong mismatches (23%). The average mismatch value was a 0.62 under-rating score.

In view of these results it seemed justified to note that the four fitness categories derived from the pretest performed surprisingly well.

4.3 Discussion of the experiment results

4.3.1 *Limits of ecological validity*

The experiment and its outcome left of course many questions open. It was clear right from the beginning that the question of how close the lab setting may come to situations of everyday life was an open issue.[13] As a fallback position (position 1) one could hold the view that at least regarding small group decision processes where a pool of candidates is generated the ecological validity is sufficiently good. But this may also be an overly restrictive option.

The following real-life examples are selected to shed some light on this issue.

Paradise Papers. A more recent case than the Engelbart story and one that by contrast did not meet the small group restriction was the christening process for the set of 13.4 million confidential electronic documents about offshore investments leaked in 2017 to Frederik Obermaier and Bastian Obermayer, two German investigative journalists at the newspaper *Süddeutsche Zeitung*, who shared them with an international network of more than 380 journalists. Here is an excerpt of what they wrote (Obermaier & Obermayer 2017; translation DZ):

> *How the Paradise Papers originated*
> Answers to the most important questions regarding the new data leaked to the SZ
> *Whence the name Paradise Papers?*
> This time, unlike the case of the Panama Papers, it was not one country that played the central role, but many countries. Since these are all so-called *Steuerparadiese* ('tax havens'), the name Paradise Papers was born.

This excited the curiosity of the author and so he sent Frederik Obermaier an email asking for the way the consensus for the final name Paradise Papers was reached. He kindly and immediately answered as follows (translation DZ):

> The naming process involved the entire international Paradise Papers team, more than 300 journalists from around the world. There were a number of aspects that played a role in the discussion, including:

---

13. At the Bologna meeting the present volume is based on, Mira Ariel gave voice to a rather strong skepticism about the transferability of the results beyond the case of Engelbart and his team.

- Does it reflect the core of the research?
- Is the name also understandable in other languages, such as French or Spanish, or is it easy to translate?
- Is alliteration possible? This has proven itself in the past (Panama Papers)
- Does the name also work in social media, for example in conjunction with a hashtag? #ParadisePapers (according to experience, it is often incomprehensible for more than two words)

In my view, "Paradise Papers" has prevailed, as this project affected several countries and therefore a limitation to individual countries would not have been fair to the project. Many partners liked the analogy with the Panama Papers and the Pentagon Papers as well as the alliteration.

Regrettably, the follow-up request for other members of the pool from which the winner had been selected was rejected; in view of future occasions the journalist would rather not disclose them.

Still the newspaper case can be seen as encouraging a stronger position (position 2) to the effect that the ecological validity of the approach is sufficiently good at least regarding group processes of selection from a pool of candidates, dropping thus the small group requirement.

Brexit. It also has been argued that the prior emergence of a candidate pool from which the neology is selected may be more the exception that the rule. Here is another real-live example that relates to this issue.

On 25 December 2016 Tom Moseley, a political reporter at BBC wrote an article *The rise of the word Brexit*. Some passages are revealing in the present context (Moseley 2016):

Who coined the phrase?
The Oxford English Dictionary awarded this honour to Peter Wilding when it added Brexit to its volumes recently.
Mr Wilding is the founder and director of the British Influence think tank – and campaigned for the UK to Remain in the EU in June's referendum.
He wrote about "Brexit" in May 2012, eight months before the then Prime Minister David Cameron had announced he would be holding a referendum.
"Unless a clear view is pushed that Britain must lead in Europe at the very least to achieve the completion of the single market then the portmanteau for Greek euro exit might be followed by another sad word, Brexit," he predicted. […]
Brexit not Brixit
It could have all been (slightly) different.
Brexit was far from set in stone and faced early competition with an alternative version featuring the following month in an Economist article predicting that "a Brixit looms for several reasons".

In August 2012, investment bank Nomura made waves when it warned the City in a report that a "Brixit" was "increasingly likely", while the same term was used in a Daily Mail column urging: "Bring on the 'Brixit'."

But Brexit prevailed, although it was another three years before its use really took off.

Even though this is only anecdotal evidence, it hints at the possible tenability of a third position that is still one grade stronger than its predecessors. The seeming absence of an overt pool of candidates may often result from the fact that competing expressions are forgotten, as in the case just mentioned. And even if there is neither a group nor an overt pool, the assumption that in word production there is always a selection among competing alternatives at least in the mind is rather well supported by current neuroscience (cf. Hagoort et al. 2009).

These considerations indicate that the more cautious views of position 1 and position 2 may be safely given up, reaching the admittedly somewhat bold conclusion that neither the restriction to small group size nor to situations with an overt candidate pool is appropriate and that the ecological validity of the approach is possibly as general as can be (position 3). There are, however, other constraints that will be discussed in the following section.

4.3.2   *Other possible constraints*

As already mentioned in Section 3.3. above the fitness score for item length in terms of syllable count is not universal. It is rather highly plausible that the finding of a preference for trisyllabicity in new words of German does not necessarily carry over to other languages. Here is a real-life example that points in this direction.

Cell phone vs. mobile phone. Preliminary evidence for a possible language specificity of optimal item length comes from comparing the proportions of *cell phone* (two syllables) and *mobile phone* (three syllables) in American and in British English, respectively, as documented in Google's Ngram viewer: whereas British English shows a strong preference for the longer over the shorter term (Figure 5), the situation in American English is exactly the opposite (Figure 6).
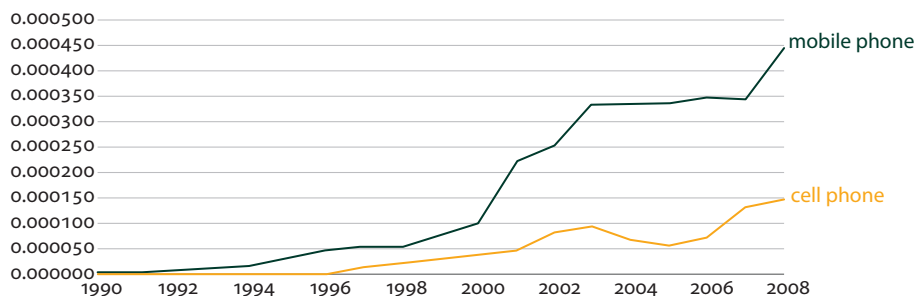

Figure 5. Frequencies of *mobile phone* and *cell phone* in British English
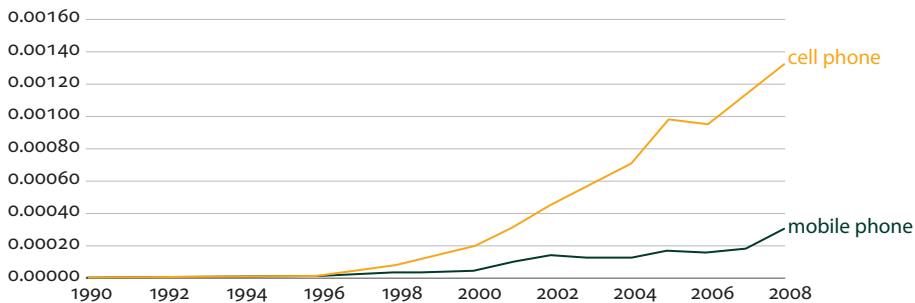
Figure 6. Frequencies of *mobile phone* and *cell phone* in American English

Moreover, it seems reasonable to assume that not only Quantity, but also the other three fitness categories may vary from language to language and from culture to culture requiring thus an adequate parameterization.

The high proportion of words with foreignisms (23%) among the 13 winners of the experiment, for instance, is probably among other things due to the age of the subjects: most of them were in their twenties. Hence age may be another relevant parameter.

Not only cultural differences, including those based on membership in different generations, play a role in parameter setting for the fitness categories, but also individual preferences and distinct kinds of group dynamics, as shown by the different degrees of participant productivity in the pretest and in the experiment. And as pointed out by Frederik Obermaier, the internet-based exchange of about 380 journalists constituted an intriguing kind of consensus finding where an even richer spectrum of fitness criteria seemed to play a role.

4.3.3 *Tackling the replicability issue*

The so-called replication crisis does not only affect psychological science (Pashler & Wagenmakers 2012), but the entire range of empirical humanities including linguistics (Peels 2019). This situation and the fact that the four fitness categories derived from the pretest performed possibly a little too well in the experiment were the main reasons for postponing the publication of its results until a replication had been performed. This is the topic of the following section.

5. Replication: Retesting the fitness rank predictor

5.1 Replication procedure

In order to control for possible hidden confounds it was the idea of the replication to keep the procedure as close as possible to the experiment procedure described above in Section 4.1. It took place in the same MELESSA lab as the pretest and the initial experiment, five months after the latter. Six new subjects were recruited through the same university service, this time three females and three males, and they were given exactly the same questionnaire and instructions as the participants in the initial experiment.

5.2 Comparing observed and predicted replication rankings

5.2.1 *Object 1 replication ranking*

Displaying a remarkable degree of creativity, the replication team invented 18 different names for the 18 available slots. However, their consensus ranking left the third place empty. This time the prediction based on the fitness scores missed the observed ranking dramatically: the winner was *Elefantenwasserhahn* 'elephant water tap' and the runner-up *Geometrino*, the former was underrated by 5 and the latter even by 6 (Table 11).

Table 11. Replication results for Object 1

fit	28	28	26	26	23	20	20	19	18	18	18	17	15	15	14	13	8	5
pr	1	1	2	2	3	4	4	5	6	6	6	7	8	8	9	10	11	12
o	–	–	–	–	–	–	–	–	1	–	–	–	2	–	–	–	–	–
dif	–	–	–	–	–	–	–	–	−5	–	–	–	−6	–	–	–	–	–

5.2.2 *Object 2 replication ranking*

By contrast with the experiment group, the replication team members didn't have any problems with finding 18 different names for the second object as well. There was one deviation in the consensus on the top three items (first and second rank were switched by one participant), but without consequences for the outcome. The winner and runners-up were *Chippihocker* 'chippy stool', *Kugelkamera* 'spherical camera' and *Einrichtungsstrauß* 'furniture ostrich/bouquet,' respectively, and the prediction missed the latter two just by one. However, the winner was rather strongly underestimated, namely by 5 (Table 12).

Table 12. Replication results for Object 2

fitness	26	25	24	24	24	23	23	22	22	22	22	21	21	20	20	16	13	12
predict	1	2	3	3	3	4	4	5	5	5	5	6	6	7	7	8	9	10
observ	–	–	–	2	–	3	–	–	–	–	–	1	–	–	–	–	–	–
diff	–	–	–	−1	–	−1	–	–	–	–	–	−5	–	–	–	–	–	–

5.2.3 Object 3 replication ranking

As with the first two objects the replication team came up with the maximum of 18 different names, and this time the consensus was faultless: the top three proposals were *Fastfoodstern* 'fast food star', *Gourmetwurst* 'gourmet sausage' and *Sternen-krabble*r 'star crawler', in this order. Fitness scores predicted the winner exactly and missed the third-place finisher just by 1, whereas the second-ranking item was grossly (by 4) underrated (Table 13).

Table 13. Replication results for Object 3

fit	32	30	30	28	28	27	26	26	26	25	24	22	22	22	21	21	20	20
pr	1	2	2	3	3	4	5	5	5	6	7	8	8	8	9	9	10	10
ob	1	3	–	–	–	–	–	–	–	2	–	–	–	–	–	–	–	–
diff	0	+1	–	–	–	–	–	–	–	−4	–	–	–	–	–	–	–	–

5.2.4 Object 4 replication ranking

Once more 100% of the provided cells were filled with 18 different names and a flawless consensus was achieved with the following order of finishers: *Wurstventilator* 'sausage fan', *Kabelhalter* 'cable holder' and *Hakenprothese* 'hook prosthesis'. The fitness scores predicted a ranking that matched this outcome perfectly for the second and third position, the winner, however, was substantially (by 5) underrated (Table 14).

Table 14. Replication results for Object 4

fit	30	26	26	24	24	23	22	22	21	21	20	20	19	18	18	17	16	14
pre	1	2	2	3	3	4	5	5	6	6	7	7	8	9	9	10	11	12
obs	–	2	–	3	–	–	–	–	1	–	–	–	–	–	–	–	–	–
dif	–	0	–	0	–	–	–	–	−5	–	–	–	–	–	–	–	–	–

5.2.5 Object 5 replication ranking

This was the only object for which one member of the highly creative replication group found only two names, reducing thus the completeness score to 99%. The consensus again was unanimous, and these were the three top ranking proposals: *Zahnpastapistole* 'toothpaste pistol', *Liegerollator* 'recumbent walker' and

Knopfaufhänger 'button hanger.' Here, the performance of the predictor was rather poor: the best-fitting score was the one for *Knopfaufhänger*, which was overrated by 2, the winner *Zahnpastapistole* was underrated by 3, and the under-rating score of 7 for the second-ranking *Liegerollator* was the strongest mismatch occurring in experiment and replication together (Table 15).

Table 15. Replication results for Object 5

fit	30	30	28	27	26	26	24	24	23	22	21	18	18	18	18	17	16
pre	1	1	2	3	4	4	5	5	6	7	8	9	9	9	9	10	11
obs	3	–	–	–	1	–	–	–	–	–	–	2	–	–	–	–	–
dif	+2	–	–	–	−3	–	–	–	–	–	–	−7	–	–	–	–	–

5.2.6   *Overview of replication rankings*

The extract from the replication results shown in Table 16 looks much less promising than its experiment counterpart in Table 10 above: given the total number of 14 pairs of predicted and observed ranks and counting again over- and under-rating score differences of 1 as near matches and larger differences as strong mismatches, the number of full matches dropped from 8 (62%) to 3 (21%) and the number of full or near matches fell from 10 (77%) to 7 (50%); in other words the number of mismatches increased dramatically from 5 (39%) to 11 (79%), among them 7 (50%) strong mismatches as opposed to only 3 (23%). Finally, the average mismatch value has almost quadrupled from a 0.62 to a 2.43 under-rating score.

Table 16. Replication results for Objects 1 to 5

Names by object	Syll	Quan	Qual	Rel	Mann	Fitn	Pre R	Obs R	Match
Elefantenwasserhahn	7	2	4	4	8	18	6	1	u_5
Geometrino	6	6	2	0	7	15	8	2	u_6
Chippihocker	4	8	2	0	10	20	6	1	u_5
Kugelkamera	5	6	6	6	6	24	3	2	u_1
Einrichtungsstrauß	4	8	2	2	7	19	4	3	u_1
Fastfoodstern	3	10	6	6	8	30	1	1	**Full**
Gourmetwurst	3	10	4	4	9	27	6	2	u_4
Sternenkrabbler	4	8	6	6	7	27	2	3	o_ 1
Wurstventilator	5	6	4	4	7	21	1	6	u_5
Kabelhalter	4	8	6	6	6	26	2	2	**Full**
Hakenprothese	5	6	4	4	8	24	3	3	**Full**
Zahnpastapistole	6	4	8	8	8	28	4	1	u_3
Liegerollator	5	6	4	4	8	22	9	2	u_7
Knopfaufhänger	4	8	6	6	5	25	1	3	o_2
MEAN	4.571	6.857	4.571	4.285	7.428	23.28			u_2.4
SD	1.049	2.099	1.761	2.249	1.237	4.130			

Understandably these figures suffice to curb any enthusiasm about the performance of the fitness score as valid predictor for neologisms. However, since the replication results could have been worse, it was decided to run a less superficial statistical analysis, which will be presented below in Section 6.1., after some remarks on the replication.

5.3 Discussion of the replication

5.3.1 *Group dynamics*

Why was there such a strong difference in the productivity of the experiment group compared to the replication group? Both had 90 cells to fill, but the first group filled less than three out of four cells (72%), whereas the second one missed the 100% score only by one. (For comparison, the productivity of the pretest group was 93%.)

The observations of the experimenter, who was present during all three sessions, support the plausibility of an explanation in terms of a pronounced difference in the group dynamics: the replication group was by far the liveliest team. Right from the beginning a male subject took the lead and moderated efficiently the consensus finding process that progressed somewhat laboriously in the experiment group. In addition, there was also a remarkably good spirit, everybody enjoyed the game and was having fun. By contrast, members of the experiment group were relatively shy and proceeded much more cautiously.

Assuming that in general groups are more similar to the experiment team than to that of the replication could vindicate the performance of the prediction tool to some extent.

5.3.2 *Optimizing fitness categories?*

Another consideration that suggested itself as a consequence of comparing the outcomes of experiment and replication was the option of fine-tuning the fitness score algorithm.

Taking the category of quantity as an example, another look at Figure 4 above, charting the distribution of the 70 items in the pretest pool according to their length in terms of syllable counts shows that the difference between the values for four and five syllables is rather small. This could motivate replacing the coding scheme of Table 4a, where the decline of the score with an increasing number of syllables after the optimum is 10–8–6–4, by the sequence 10–8–8–4 as in Table 17:

Table 17. Alternative quantity score as function of syllable count

Syllable number	1	2	3	4	5	6	7	8	9
Quantity score	6	8	10	8	8	4	2	1	0

As a result, the number of full matches in the replication would decrease from 3 (21%) to 2 (14%), and that number of full or near matches from 7 (50%) to 6 (42%), with the corresponding increases in the numbers of (strong) mismatches. On the other hand, the average mismatch value would decrease considerably, by about 30% (from an under-rating of 2.43 to 1.71), so if the average mismatch value is considered as most important this change in the fitness score algorithm could indeed slightly improve the performance.

However, since the advantage of this modification seems rather slim, it was not taken into account by the statistics presented below, which are based on the original unmodified schema.

6.   Overall analysis

6.1   Ways of estimating correlation strength of incomplete rankings with ties

The primary problem for a statistical analysis posed by our data can be easily read off a sample data set. Table 13 (page 141) shows the fitness scores of the 18 names proposed for Object 3 in the replication together with the resulting predicted ranking, the observed ranking and the differences, showing one perfect match, one over-ranking by 1 and one under-ranking by 4.

The considerable number of ties (four double and two triple ties resulting in a rather smooth decline of only 12 points over 18 items) is not a big problem, but the main challenge comes from the high percentage of missing data: Whereas the predicted ranking is complete the observed ranking regards only three words, such that out of 36 data cells pairing predicted with observed ranks only 21 are filled. A simple way of dealing with this situation would of course be a radical cut-down on the data by throwing away all incomplete pairs of predicted and observed ranks, keeping only 6 of the 36 data points.

An application of Kendall's tau-b rank correlation coefficient (Kendall 1945),[14] ranging from −1 to +1, to all ten data sets yields the following numerical (Tables 18 and 19) and graphical (Figure 7) results.

---

14. The main reason for choosing this algorithm is its being designed for dealing with ties, which occur frequently in the data of this study.

Table 18. Minimal separate results of experiment and replication

Minimal separate results						
Experiment	**Object 1 e**	**Object 2 e**	**Object 3 e**	**Object 4 e**	**Object 5 e**	**Mean**
tau-b	1	0	−0.2	1	0.82	0.524
p95greater	0	0.5	0.65	0.17	0.11	0.286
Replication	**Object 1 r**	**Object 2 r**	**Object 3 r**	**Object 4 r**	**Object 5 r**	**Mean**
tau-b	1	−0.333	0.5	−0.333	−0.333	0.100
p95greater	0.5	0.833	0.333	0.833	0.833	0.666

Table 19. Minimal results of experiment and replication together

Minimal total results	Mean
tau-b total	0.319
p95greater total	0.512
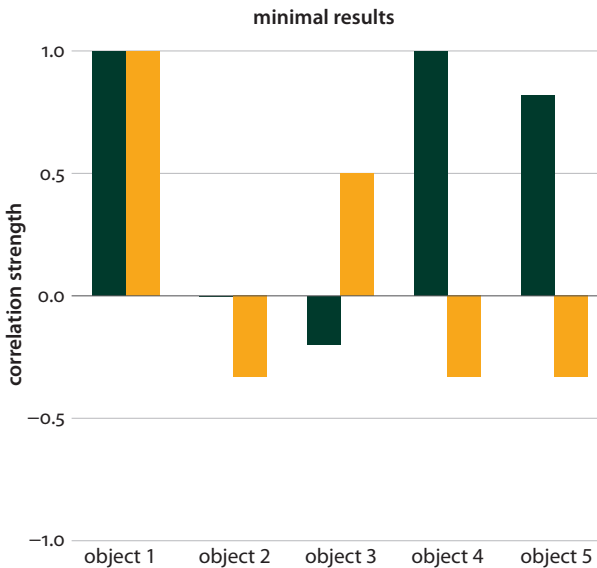
Figure 7. Minimal results of experiment (black) and replication (yellow)

Obviously, disregarding the majority of the collected data results in a notable contrast between the experiment and the replication: except for Object 1 they disagree in all cases on the support of the null hypothesis which predicts there to be no or even a negative correlation between predicted and observed rankings. Given that in view of the small number of data points the p-values are expectably poor, the quality

of the predictor is mainly reflected in the strength of a positive correlation, which on average is much better in the experiment's data set than in that of the replication.

But the predicted rank values for those items that lack an observed rank number contain information that does not have to be entirely discarded: the missing values are not completely random, but rather strongly constrained by the requirement of constituting a ranking as well. Unfortunately, although there is an impressively rich literature on missing data,[15] nothing seemed to fit exactly the rank correlation patterns at stake in the present data.

Therefore, a specific custom-tailored algorithm had to be developed. In order to get all the missings in one sequence the Tables were reordered according to the observed ranking, transforming thus for instance Table 13 into Table 20 by rearranging the three complete data pairs:

Table 20. Replication results for Object 3 rearranged by observed ranks

fit	32	25	30	30	28	28	27	26	26	26	24	22	22	22	21	21	20	20
pr	1	6	2	2	3	3	4	5	5	5	7	8	8	8	9	9	10	10
ob	1	2	3	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
diff	0	−4	+1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–

Then a somewhat more sophisticated way of dealing with the missing data than the radical cut leading to the minimal results consists in taking the average of a sufficiently high number of random samples (vectors of length 15 in the sample case) used to fill the gaps.

That could be adequate if there were no further constraints on those randomly chosen vectors apart from length. But as mentioned above there are further restrictions: since ties are possible, the first vector element has to be either 3 again or 4, and the last vector element some integer between 3 and 18. In other words, the condition on the sequence of numbers requires a continuous increasing weak order, which reduces the number of options from 15^{15} to 2^{15}, i.e., from a number with eighteen digits to 32,768. As a result, the random sampling has to choose only from this much smaller set.

But what happens to the predicted partners of the thus imputed 'observed' ranks? Looking at Table 20 may be misleading, since although it has been sorted, except for the first three columns, in terms of the 'predicted rank' row, for the missing data the order of the corresponding predicted ranks and fitness scores has not been changed. But the predicted rank orders that would result from the different completions of the 'observed rank' line are entirely unknown, meaning that with

---

15. As of November 6, 2019, Google Scholar records 10,428 publications quoting the highly influential article 'Missing data: our view of the state of the art' (Schafer & Graham 2002).

the exception mentioned above Table 20 is only one of a enormous number of equivalent arrangements. What is known is the identity and frequency of the rank numbers that possibly have to be rearranged.

Unfortunately, by contrast with the completed observed rank row where the number of possibilities for n missing values is 2^n, the number of permutations of a given predicted rank order rises factorially in the worst case. In the given example the number of possible orderings is 59,875,200, and generating all these permutations exceeds the limits of the computing power of a desktop.

The algorithm that was developed in the light of these considerations[16] in order to satisfy the relevant constraints proceeds in five steps.

First, the set of possible completions of the predicted rank order is generated, which consists of the permutations of the rank positions without an observed partner. In the case of Table 20 the permutation input is the single-row Table 21 attained from Table 20 by clipping off the first three cells of the second line:

Table 21. Partial replication results for Object 3: Predicted ranks without corresponding observations

pr	2	3	3	4	5	5	5	7	8	8	8	9	9	10	10

In view of (a) the potential factorial growth of the number of permutations, mitigated only by the amount of ties, and (b) the computational limits mentioned above vectors exceeding the length of 12 like the one in Table 21 are made to fit the ceiling by cutting off the last three (shaded) elements.

Second, the set of possible completions of the observed rank row is calculated. In the present example this is the set of possible continuations (monotonic weak orders) of length 12 of the three top ranks.

Third, for control purposes, sample matrices are built that combine instances of the two completed rows. Tables 22 and 23 show two examples.

Table 22. Replication results for Object 3 with randomly completed rankings

fitness score	32	25	30										
predicted rank	1	6	2	8	3	2	5	5	7	3	9	4	8	5	8
observed rank	1	2	3	3	3	4	4	4	4	5	5	6	7	7	7
difference	0	−4	+1	−5	0	+2	−1	−1	−3	+2	−4	+2	−1	+2	−1

---

16. The initial idea was inspired by a suggestion from Felix Schönbrodt (p.c.).

Table 23. Replication results for Object 3 with different randomly completed rankings

fitness score	32	25	30											
predicted rank	1	6	2	5	8	8	2	3	4	5	9	7	5	3	8
observed rank	1	2	3	3	3	3	3	3	4	4	4	4	5	5	5
difference	0	−4	+1	−2	−5	−5	+1	−1	0	−1	−4	−3	0	+2	−3

Fourth, a Kendall tau-b correlation test with confidence level 0.95 and the alternative hypothesis of a positive correlation coefficient is run 10,000 times, drawing thus up to 10,000 different random samples from the set of possible orderings of predicted ranks on the one hand and of possible continuations of the observed ones on the other.

In the present case this repetition procedure made the resulting tau-b and *p*-values stable at least in the first two decimal positions.

Fifth and last, the procedure is run three times and the means of the results are calculated.

6.2 Statistical overview of experiment and replication results

A comparison of minimal (Tables 18, 19, and Figure 7) and full results (Tables 24, 25, and Figure 8) elucidates the benefit gained by incorporating the (almost) full amount of data on the prediction level and the restricted set of possible continuations on the observation level: the extremes, especially the strong contradictions between experiment and replication, are gone and the number of negative results has dropped from three to one.

Table 24. Full separate results of experiment and replication

Full separate results						
Experiment	**Object 1 e**	**Object 2 e**	**Object 3 e**	**Object 4 e**	**Object 5 e**	**Mean**
tau-b	0.16	0.26	0.11	**0.44**	0.30	0.254
p95greater	0.30	0.20	0.33	0.05*	0.15	0.206
Replication	**Object 1 r**	**Object 2 r**	**Object 3 r**	**Object 4 r**	**Object 5 r**	**Mean**
tau-b	*−0.04*	*0.01*	0.21	0.28	*0.05*	0.102
p95greater	0.56	0.49	0.20	0.14	0.42	0.362

Table 25. Full results of experiment and replication together

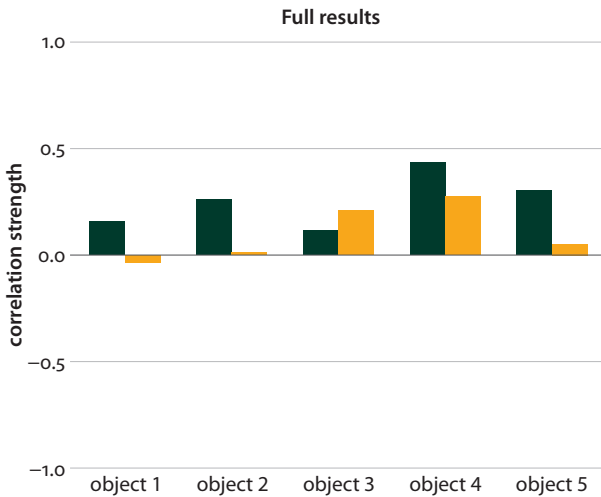| Full total results	Mean
tau-b total	0.178
p95greater total	0.284

Figure 8. Full results of experiment (black) and replication (yellow)

According to Cohen's (1988) frequently used suggestion for evaluating the correlation coefficient as effect size (here correlation strength) estimators, this size is called as follows:

– *small*   if the coefficients are between .10 and .29,
– *medium*   for coefficients between .30 and .49, and
– *large*   in the case of coefficients of .50 and higher.

Using this terminology we can say that among the ten full individual results there is one medium size effect (0.44 for Object 4 in the experiment; printed in boldface), six small effects (for the other four objects in the experiment and for two in the replication), and three negligible ones (all in the replication; printed in italics), one of them even negative. The mean of the experiment effect size is in the upper quarter of the range of small (0.254), that of the replication less than half as big, but still in the same category, albeit by a rather tiny margin, and the overall mean is therefore small as well.

The probabilities of the obtained (or worse) results under the assumption of a one-tailed null hypothesis (no or even a negative correlation between predicted and observed rankings) at a .05 significance level (.05 p-values) are almost all by far too high, which is what one would expect given the small amount of data. There is one exception, the result for Object 4 in the experiment, which is just significant (marked by an asterisk) in that sense. In any case, the p-value average of 0.206 for the experiment means that in four out of five cases the null hypothesis fails to predict the obtained results.

Moreover, there is a well-motivated agreement among most researchers that the usefulness of a pilot study cannot be read off standard significance levels like the usual .05-level, the reason being that virtually by definition pilot studies are underpowered in view of this level. Here is a pertinent quote from an article dealing with medical research methodology:

> We recommend that in pilot trials the focus should be on descriptive statistics and estimation, using confidence intervals, rather than formal hypothesis testing and that confidence intervals other than 95% confidence intervals, such as 85% or 75%, be used for the estimation.   (Lee at al. 2014: 1)

In view of the fact that a 95% confidence interval and a 75% confidence interval correspond to a 05% and a 25% significance level, respectively, we can state that five out of the altogether ten results for individual objects are in line with this recommendation, and so does the average of the five results of the initial experiment.

All in all, the data collected in this pilot study provide some preliminary evidence for the hypothesis of a positive correlation, albeit a small one, between the predicted and the observed rankings and hence for the usefulness of *H3* and *H4, Müller's* and *Darwin's Hypothesis* in our operationalization, in predicting some core features of neologisms.

7.   Summary and outlook

In the light of the considerations and results discussed and presented in this paper, what is the answer to the question asked in its title? Is it really the case that new words are predictable? A short, straightforward and rough answer is simply: no.

But there is evidence that the short answer is true only to some extent. Admittedly, predicting the emergence of demand for a new word is possibly beyond of the scope of any theory, let alone linguistics. But regarding their linguistic aspects a longer and less simplifying answer is this: new words are at least not completely unpredictable. There is reason to assume that in line with *H6*, the *Final Word Prediction Hypothesis*, the four factors quantity (word length), quality (descriptive fit), relation (discriminativeness) and manner (attractiveness) play crucial roles in their selection.

As outlined above (Section 1.3) this hypothesis is far from being new, it has been formulated in a rather speculative way already one and a half centuries ago by two outstanding scientists of their epoch who mentioned those factors as determining the survival of certain new words at the expense of others: *H3, Müller's Hypothesis*, comprises the first three of these factors, and *H4, Darwin's Hypothesis*, states the last one.

What is new is that the present pilot study has provided some experimental support for these hypotheses. The proposed four fitness categories did reasonably, in part even surprisingly well. Even though due to its lack of power this exploration could not come up with really strong evidence, its results are compatible with the notion that a more extensive investigation, possibly with an improved weighting of factors, could lead to a truly predictive theory of core determinants of the creation, selection and survival of new words.

In view of these findings there is reason to assume that *Q1*, Engelbart's Question, to wit 'Why do some neologies stick, whereas other don't?' is not hopelessly hard to answer.

Last, but not least I want to submit that the present study encourages using the concept of natural selection via community preference with both its conscious and its uncontrollable aspects for the investigation of other phenomena, linguistic and non-linguistic, in the domain of cultural evolution, and that it definitely seems to be worth of further exploration.

Acknowledgements

References

Aronoff, Mark. 2016. Competition and the lexicon. In *Livelli di analisi e fenomeni di interfaccia*, Annibale Elia, Claudio Iacobini & Miriam Voghera (eds), 39–52. Roma: Bulzoni.

Aronoff, Mark. 2017. Darwinism tested by the science of language. In *On Looking into Words (and Beyond)*, Claire Bowern, Laurence Horn & Raffaella Zanuttini (eds), 443–456. Berlin: Language Science Press.

Beaumont, Claudine. 2008. Computer mouse celebrates 40th birthday. <https://www.telegraph.co.uk/technology/news/3538800/Computer-mouse-celebrates-40th-birthday.html> (2 March 2020)

Barry, Tom J., Griffith, James W., De Rossi, Stephanie & Hermans, Dirk. 2014. Meet the Fribbles: Novel stimuli for use within behavioural research. *Frontiers in Psychology*, 5(article 103): 1–8. https://doi.org/10.3389/fpsyg.2014.00103

Bloomfield, Leonard. 1926. A set of postulates for the science of language. *Language* 2(3): 153–164. https://doi.org/10.2307/408741

Casasanto, Daniel. 2009a. Embodiment of abstract concepts: Good and bad in right-and left-handers. *Journal of Experimental Psychology: General* 138(3): 351.
https://doi.org/10.1037/a0015854

Casasanto, Daniel. 2009b. When is a linguistic metaphor a conceptual metaphor? In *New Directions in Cognitive Linguistics* [Human Cognitive Processing 24], Vyvyan Evans & Stéphanie Pourcel (eds), 127–145. Amsterdam: John Benjamins. https://doi.org/10.1075/hcp.24.11cas

Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale NJ: Lawrence Erlbaum Associates.

Darwin, Charles. 1859. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray.

Darwin, Charles. 1871. *The Descent of Man, and Selection in Relation to Sex*. London: John Murray.

Engelbart, Douglas Carl. 1995. Toward augmenting the human intellect and boosting our collective IQ. *Communications of the ACM* 38(8): 30–33. https://doi.org/10.1145/208344.208352

English, William K., Engelbart, Douglas C. & Huddart, Bonnie. 1965. Computer-Aided Display Control. *Final Report*, Contract NASl-3988, SRI Project, 5061(1).

Gause, Georgij F. 1934. *The Struggle for Existence*. Baltimore MD: Williams and Wilkins.
https://doi.org/10.5962/bhl.title.4489

Girard, Gabriel. 1718. *La justesse de la langue françoise ou les différentes significations des mots qui passent pour synonimes*, Paris: Laurent d'Houry.

Goldberg, Adele E. 2019. *Explain me This: Creativity, Competition, and the Partial Productivity of Constructions*. Princeton NJ: Princeton University Press.

Grice, H. Paul. 1975. Logic and conversation. In *Syntax & Semantics: Speech Acts*, Vol. 3, Peter Cole & Jerry L. Morgan (eds), 41–58. New York NY: Academic Press.

Haeckel, Ernst. 1866. *Generelle Morphologie der Organismen. Allgemeine Grundzüge der organischen Formen-Wissenschaft, mechanisch begründet durch die von Charles Darwin reformirte Descendenz-Theorie*, Band 2. Berlin: Reimer.

Hagoort, Peter, Baggio, Giosuè & Willems, Roel M. 2009. Semantic unification. In *The Cognitive Neurosciences* (4th edn), Michael S. Gazzaniga (ed.), 819–836. Cambridge MA: The MIT Press.

Hughes, Geoffrey. 1991. *Swearing: A Social History of Foul Language, Oath and Profanity in English*. Cambridge MA: Blackwell.

Kant, Immanuel. 1781 [1881]. *Critique of Pure Reason*. Translated by Friedrich Max Müller as *Immanuel Kant's Critique of Pure Reason: In Commemoration of the Centenary of Its First Publication*. Londen: Macmillan and Company, 1881.

Kendall, Maurice G. 1945. The treatment of ties in ranking problems. *Biometrika* 33(3): 239–251. https://doi.org/10.1093/biomet/33.3.239

Kerremans, Daphné & Prokić, Jelena. 2018. Mining the web for new words: Semi-automatic neologism identification with the NeoCrawler. *Anglia* 136(2): 239–268. https://doi.org/10.1515/ang-2018-0032

Lewis, David. 1969. *Convention: A Philosophical Study*. Cambridge MA: Harvard University Press.

Lai, Randy. 2018. Arrangements: Fast generators and iterators for permutations, combinations and partitions. *R Package Version 1.1.5*. <https://CRAN.R-project.org/package=arrangements> (2 March 2020).

Lee, Ellen C., Whitehead, Amy L., Jacques, Richard M. & Julious, Steven A. 2014. The statistical interpretation of pilot trials: Should significance thresholds be reconsidered? *BMC Medical Research Methodology* 14: 41. https://doi.org/10.1186/1471-2288-14-41

Mesoudi, Alex, Whiten, Andrew & Laland, Kevin N. 2006. Towards a unified science of cultural evolution. *Behavioral and Brain Sciences* 29(4): 329–347. https://doi.org/10.1017/S0140525X06009083

Moseley, Tom. 2016. The rise of the word Brexit. *BBC News*, 25 December 2016. <https://www.bbc.com/news/uk-politics-37896977> (2 March 2020).

Müller, Max. 1870. Darwinism tested by the Science of Language (translation 1869 of Schleicher 1863). Review. *Nature* 1: 256–259

Obermaier, Frederik & Obermayer, Bastian. 2017. Wie die Paradise Papers entstanden. Antworten auf die wichtigsten Fragen rund um das neue Datenleck, das der SZ zugespielt wurde. *Süddeutsche Zeitung*, Dienstag, 7. November 2017 (255): 14.

Pashler, Harold & Wagenmakers, Eric Jan. 2012. Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science* 7(6): 528–530. https://doi.org/10.1177/1745691612465253

Peels, Rik. 2019. Replicability and replication in the humanities. *Research Integrity and Peer Review* 4(1): 2. https://doi.org/10.1186/s41073-018-0060-4

Pinker, Steven & Bloom, Paul. 1990. Natural language and natural selection. *Behavioral and Brain Sciences* 13(4): 707–727. https://doi.org/10.1017/S0140525X00081061

Progovac, Ljiljana. 2019. *A Critical Introduction to Language Evolution: Current Controversies and Future Prospects* [Springer Expert Briefs in Linguistics]. Cham: Springer. https://doi.org/10.1007/978-3-030-03235-7

R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/> (25 March 2021).

Richards, Robert J. 2008. *The Tragic Sense of Life: Ernst Haeckel and the Struggle over Evolutionary Thought*. Chicago IL: The University of Chicago Press. https://doi.org/10.7208/chicago/9780226712192.001.0001

Richerson, Peter J. & Boyd, Richard. 2005. *Not by Genes Alone: How Culture Transformed Human Evolution*. Chicago IL: The University of Chicago Press.

Schafer, Joseph L. & Graham, John W. 2002. Missing data: Our view of the state of the art. *Psychological Methods* 7: 147–177. https://doi.org/10.1037/1082-989X.7.2.147

Schleicher, August. 1863. *Die Darwinsche Theorie und die Sprachwissenschaft – offenes Send-schreiben an Herrn Dr. Ernst Häckel, a. o. Professor der Zoologie und Director des zoologischen Museums an der Universität Jena.* Weimar: Hermann Böhlau.

Schleicher, August. 1869. *Darwinism Tested by the Science of Language*, Translated by Dr. Alex V. W. Bikkers. London: John Camden Hotten. https://doi.org/10.5962/bhl.title.49464

Schmid, Hans-Jörg. 2008. New words in the mind: Concept-formation and entrenchment of neologisms. *Anglia-Zeitschrift für englische Philologie* 126(1): 1–36. https://doi.org/10.1515/angl.2008.002

Smith, Richard. 1997. Keeping the bad news from journalists. *British Medical Journal* 314: 81. https://doi.org/10.1136/bmj.314.7073.81

Sterrett, Susan G. 2002. Darwin's analogy between artificial and natural selection: How does it go? *Studies in History and Philosophy of the Biological and Biomedical Sciences* 33(1): 151–168. https://doi.org/10.1016/S1369-8486(01)00039-5

Veale, Tony & Butnariu, Cristina. 2006. Exploring linguistic creativity via predictive lexicology. In *The Third Joint Workshop on Computational Creativity*, ECAI 2006.

The Camel Humps prosodic pattern
Listing for disaffiliating in spoken Hebrew

Nadav Matalon

Hebrew University of Jerusalem

This chapter presents an analysis of a common prosodic pattern in spoken Hebrew. The pattern is characterized by a repetition of high rise-fall pitch movements, which visualize as successive "humps" in the pitch curve.[1] Based on analysis of naturally occurring conversations, I show that speakers use this pattern to construct open lists of reinforcements for a counter-stance they take. In some cases, the use of the pattern is itself the act of taking a counter-stance, that is, in such cases the disaffiliation between the stance-takers is only prosodically cued. Such cases demonstrate that prosodic list patterns are used not only for the task of enumerating items under one category, but also as a resource for complex interactional activities.

Keywords: prosody, Hebrew, list construction, stance, affiliation

The structure of the chapter is as follows. In Section 1, I introduce the premises underlying this study and describe the data. Section 2 includes a formal analysis of the pattern, followed by a functional analysis in Section 3. Conclusions are drawn in Section 4.

1. Introduction

This study follows the principles of Interactional Linguistics (IL, Couper Kuhlen & Selting 2018), which builds on the understanding that ordinary conversation is a structurally organized phenomenon, and aims to "describe linguistic structures and meanings as they serve social goals in [...] talk-in-interaction" (Lindström

---

1. a sample of three instances is available at: 🎧
   Audio files can be found at: https://doi.org/10.1075/slcs.220.06mat.audio
   🎧 marks the availability of an audio file

2009: 96). IL assumes a bi-directional relationship between language and interaction. On the one hand, recurrent interactional needs of interactants lead them to use linguistic forms over and over again for the same purposes in similar contexts, thus leading forms to become conventionalized and identified with specific interactional functions. On the other hand, synchronically speaking, each specific language (and specific dialect, sociolect etc.) lends itself to the members of the speech community as a limited toolbox of conventionalized forms, thus, in turn, shaping (particular) interactions (Lindström 2009).

In interaction-oriented approaches to prosody (Couper-Kuhlen & Selting 1996), prosodic patterns are viewed as contextualization cues (Gumperz 1982; Auer 1992), i.e., as elements which create a favorable frame or context for the interpretation of a particular utterance. According to Auer (1992), contextualization cues do not have a referential meaning of the kind lexical items have, but they are often necessary for participants in order to reach a plausible interpretation of what is going on in a given piece of interaction. Couper-Kuhlen (2015) claims that "prosodic contextualization cues help interactants make inferences about turn-taking and floor management, on the one hand, and about what actions or activities are being carried out, how they are being carried out, and how this might impinge upon participants' social image or face, on the other" (2015: 86).

This study is based on The Haifa Corpus of Spoken Hebrew (Maschler et al. 2017), which contains recordings of spontaneous conversations between students, their friends and relatives. The recordings are transcribed according to the Discourse Transcription system (Du Bois et al. 1992; see Appendix for full transcription conventions). At the time I conducted the study, the corpus consisted of 243 interactions, collected during the years 1993–2014. The recordings range in length from 1 to 9 minutes each, comprising over 11 hours of talk. In total, 701 different speakers participated in the recordings, with 2–5 interlocutors per interaction. I detected 69 examples of the studied pattern in the corpus, which were produced by 53 different speakers. I used the software PRAAT (Boersma & Weenink 2016) for acoustic analyses.

2.   The form

The Camel Humps pattern is one of the open-list patterns in spoken Hebrew, i.e., it is a pattern that speakers use to enumerate items under one category, without limiting themselves to a specific number of items. Selting (2007) discusses the distinction between closed and open lists:

> (a) closed lists […] suggest a closed number of items, and (b) open lists […] suggest an open number of items. These classes of lists are produced with different kinds of practices. It seems to be the prosody that is used to suggest the intended kind of list, irrespective of its syntactic embedding.   (Selting 2007: 9)

Typically, the last item in a closed list is prosodically different than the rest of the list items in that it carries a fall-to-low pitch movement, which marks the completeness of the list.[2] In open lists, on the other hand, all items share a similar prosodic design (according to the specific list pattern) that marks a continuation.[3] Thus, each item in an open list projects more-to-come (Jefferson 1990), giving the impression of a (potentially) never-ending list. The Camel Humps pattern is highly identifiable,[4] and it is perceived as distinct from other open-list patterns in Israeli Hebrew (see Section 3.6).

Extract 1 provides a first illustration of the pattern. The conversation is held by Tslil and Oren (Tslil is a friend of Oren's wife), and the topic under discussion is exercising by oneself or in a group.[5]

Extract 1. "The kayak rower" (1:48–1:57 min.)[6]   🎧

```
  136  Oren:  ze      mešaʕamem,
              this    boring
              'it is boring'
```

---

2.   For an example of a closed list, see Section 3.2, Extract 5, lines 3–4, and the discussion thereof.

3.   See Couper Kuhlen (1986: 150) for a description of several closed and open list patterns in English.

4.   Pronouncing examples of the pattern to other native Hebrew speakers always draws an "aha" reaction (Tannen 1984), that is, a reaction that shows that something they sensed intuitively was made explicit.

5.   See a full analysis of this example in Section 3.1.

6.   In the transcription, the breakdown to numbered lines denotes the segmentation into intonation units (Chafe 1994). The transcription consists a phonemic representation of the Hebrew text, an English gloss, and, when needed, a literal translation of the utterance (PTCL = particle). Underlining is used to mark the scope of the "hump" (see Section 2.1) in each intonation unit. The following characters are used to represent the Hebrew consonants that cannot be represented by a single English letter: 'c' for /ts/; 'š' for /sh/; and 'x' for /ch/ (voiced uvular fricative). See Appendix for full transcription conventions.

```
  137        levad.
             alone
  138  Clil: beemet?
             really?
  139        ani   sonet  laasot              im    axerim.
             I     hate   doing{exercising}  with  others
• 140        keilu  ani,
             like   I
• 141        ani  ohevet laruc   harey,
             I    love   to.run  PTCL
             'I love to run as you know'
• 142        az ani  raca,
             so I    run
→ 143        ba-kecev        šeli--,
             in.the-rhythm   of.1SG
             'at my [own] pace'
→ 144        ba-zman         šeli--,
             in.the-time     of.1SG
             'in my [own] time'
⇒ 145        ani sonet laruc       im   od    mišehu.
             I   hate  to.run      with more  someone
             'I hate running with others'
  146 Oren:                        at  raca baxuc?
                                   you run  outside?
```

Tslil's whole response (lines 138–145) meets Selting's (2004, 2007) description of the three-part gestalt in which lists are typically embedded:

[…] lists are always themselves embedded into a three-component structure, with:

a. *The projection component, projecting more-to-come, i.e. a multi-unit turn to be constructed, either a pre-detailing and/or a general formulation;*

b. *The list itself, preferably three-parted, suggesting the items as part of either a closed or an open number of list items, as a practice of detailing;*

c. *The gestalt closure, i.e. a post-detailing component, completing the structure around the list.* (Selting 2007: 58)

Lines 140–142 form the projection component (marked with the symbol •), projecting more information to come, i.e., the detailing of why the fact that Tslil loves to run is relevant for the fact that she prefers training by herself. Lines 143–144 (marked with the symbol →), are the list itself that is performed with the Camel Humps pattern. In her list, Tslil enumerates advantages, from her point of view, of training by oneself. Line 145 is the gestalt closure (marked with the symbol ⇒).[7]

---

7. This three-part structure is attested in many cases throughout the data. However, there are instances that lack a projection component, a gestalt closure, or both. For instance, some instances lack a projection component as the list itself is provided as a response to a challenging question. Since the focus here is the characteristics of the list itself, further reference to the three-part

Jefferson (1990) claims that there is a preference for lists of three items. The findings of this study only partially support this claim – although the average number of items is 2.9 ($N = 69$, $SD = 1.1$) and three-item instances are indeed common, two-item lists are slightly more common and the actual number of items in one list can be as high as six or as low as one (see Table 1).

Table 1. Distribution of the number of list-items across the instances of the camel humps pattern

Number of items	1	2	3	4	5	6	Total
Number of instances	5	24	22	14	1	3	69
%	7.2	34.8	31.9	20.3	1.4	4.4	100

2.1 Prosodic structure

The Camel Humps pattern is constructed by a concatenation of intonation units (henceforth IUs, see Chafe 1994; Du Bois et al. 1992; Himmelmann et al. 2018). IUs are prosodic "chunks" of speech, delimited by several prosodic cues such as disfluencies in speech rate and pitch reset (but cf. Barth-Weingarten 2016). On average, these units of talk are approximately 1-second long, consist of 4–5 words, and contain one primary accented syllable.[8]

The IUs that make up an instance of the Camel Humps pattern share two essential features: they end with a relatively high rise-and-fall pitch movement, and their last syllables are lengthened compared to the usual final lengthening of IUs. The fact that each list-item constitutes a separate IU is consistent with Selting (2007), who claims that "most open lists are constructed […] with each list item constituting a prosodically packaged separate TCU"[9] (2007: 32). Figure 1 is an illustration of an idealized instance of the Camel Humps pattern; it presents a division into two IUs, each ending with a "hump".

One of the two essential features of the Camel Humps pattern is the relatively high rise-and-fall pitch movement that defines each hump. The exact location of the rise and fall movements depends on the location of the primary accent within

---

structure will not be made, apart from marking the projection component and the gestalt closure in the transcriptions, when they exist.

8. A primary accented syllable is a prominent syllable that is realized as a pitch deviation from a neutral baseline and is either loud or lengthened or both (Chafe 1994: 60).

9. Turn Constructional Unit (Sacks, Schegloff & Jefferson 1974). A TCU is "a coherent and self-contained utterance, recognizable in context as 'possibly complete'" (Clayman 2013). Therefore, by definition, a TCU consists of at least one IU.
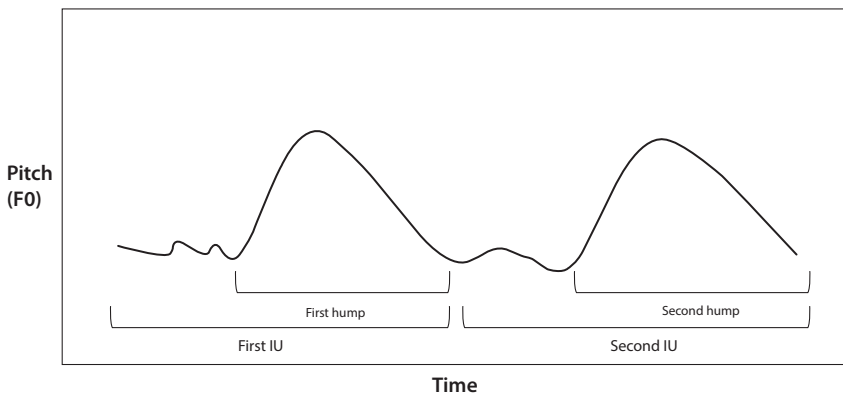
Figure 1. Illustration of the Camel Humps pattern: A pitch-track of a two-hump instance

the IU: when the primary accent is placed on the last syllable of the IU, the hump is mono-syllabic, i.e., the pitch rises and falls within this one syllable; when this is not the case, the hump is multi-syllabic, and the pitch rises and reaches its peak within the primary accented syllable, and falls over the following syllables, until the end of the IU.[10,11] The height of each hump is operationalized here as the difference in F_0 (measured in semitones) between the onset of the pitch rise and the pitch peak. [12] Table 2 presents measurements of the height of humps across the instances of the pattern.

Table 2. Height of humps across the instances of the Camel Humps pattern (in semitones)

	Minimal height	Maximal height	Average height	SD	Number of instances*
Female	4	14.2	8.75	2.46	51
Male	6	12.9	9.15	1.83	14
Total	**4**	**14.2**	**8.8**	**2.3**	**65**

* Four instances of the pattern were excluded from the analysis presented here due to poor sound quality or overlap speech, which made it impossible to extract acoustic data.

---

10. In the terms of the British school of intonation (e.g., Palmer 1922), each "hump" consists of the *nucleus* and the *tail* of the IU (if there is a tail).

11. Because the movement of pitch may cover multiple syllables, and for reasons of legibility, the transcriptions in this chapter do not contain symbols of rising and falling pitch. However, since the "humped" part of the IU is underlined in the transcription, the reader can rely on the description above to appreciate the location and scope of the rise and fall of pitch.

12. Since the fall in pitch (beginning right after the pitch peak and lasting until the end of the IU) is usually just as large as the rise in pitch, only the rise in pitch was systematically measured.

The second essential feature of the Camel Humps pattern is the lengthening of syllables within humps. The mean length of syllables within humps is presented in Table 3.

Table 3. Syllable length within humps across the instances of the Camel Humps pattern (in miliseconds)

	Minimal mean* length	Maximal mean length	Average length**	SD	Number of syllables
Female	117	600	352	126	188
Male	107	940	360	127	66
Total	**107**	**940**	**354**	**126**	**254**

* Per one hump.
** Across all instances.

To verify that, as perceived, syllables within humps are truly lengthened, I performed a comparison of syllable lengths along the following lines. I drew a random sample of 10 of the 53 speakers who provided the instances of the pattern. For each speaker, a stretch of talk was selected as a "counter example" on the following grounds: I primarily selected another open list (with a different prosodic pattern) if there was one. If there were no additional open lists, I selected a stretch of talk in which the speaker connected several IUs, each one ending with a continuous tone (Du Bois et al. 1992).[13] Then, in each such counter example, I calculated the mean length of syllables between the onset of the primary accent to the end of the IU.[14] Table 4 compares the mean length of syllables across instances of the Camel Humps pattern to that of syllables across the counter examples. The comparison confirms that the syllables within the humps are lengthened, as perceived. These findings are consistent with Silber-Varod (2011), who defined a 230 ms threshold for lengthened syllables in her work on Hebrew boundary tones.

Table 4. Syllable length of a sample of ten speakers (in miliseconds): Comparing instances of the Camel Humps pattern to counter examples

	Minimal mean length	Maximal mean length	Average length	SD	Number of syllables
Pattern instances	163	940	387	196	46
Counter examples	120	390	200	50	70

---

13. Only stretches of talk with completed, i.e., not truncated, IUs were included; considered were both "substantive" IUs, i.e., IUs that convey ideas of events, states, or referents, and "regulatory" IUs, i.e., IUs that regulate interaction or information flow (Chafe 1994: 63).

14. Again, in the terms of the British school (e.g., Palmer 1922), the nucleus and the tail (if there is one).

As mentioned above, some humps contain more than one syllable. In my data, humps are 1- to 7-syllable long (the average number of syllables per one hump is 1.69). Multisyllabic humps tend to be longer than monosyllabic ones, but the more syllables a hump contains, the shorter the syllables in it tend to be. Thus, syllables in multisyllabic humps might be shorter than 230 ms. In fact, in rare cases of extreme multisyllabic humps, even the mean syllable length might be shorter than 230 ms (as can be seen in Tables 3 and 4).[15]

2.2   Syntactic structure

The Camel Humps pattern presents some variation in terms of the syntactic structure of the listed items. In most cases, the list comprises simple declarative clauses, but smaller syntactic units such as noun phrases (e.g., as direct objects), prepositional phrases (e.g., as indirect objects or adjuncts), and verb phrases, can be found as well. When lists consist of syntactic units smaller than a clause, each consecutive item functions as an additional, or alternative, completion of the same syntactic slot in the same clause. "The kayak rower" example that was presented above is such a case:

Extract 1.   (Shortened) – "The kayak rower" (01:53–01:55 min.)

```
• 142       az ani raca,
            so  I   run
→ 143       ba-kecev         šeli--,
            in.the-rhythm    of.1SG
            'at my [own] pace'
→ 144       ba-zman          šeli--,
            in.the-time      of.1SG
            'in my [own] time'
```

The clause "(so) I run" (line 142) is completed with the first prepositional phrase "at my (own) pace", and then, without repeating the first part of the clause, an additional prepositional phrase is given – "in my (own) time". Thus, the list comprises two prepositional phrases.

The list items of each of the instances tend to be syntactically equivalent. Nevertheless, in some cases the items in a list are syntactically varied. Consider, for example, Extract 2. In this conversation, Beni, Uri and Michal are talking about a questionable social event that Beni and Uri participated in. At this point, the

---

15.   For example, see line 230 in Extract 2 below. The noun *gitara* (guitar) is left dislocated and emphasized, due to information-structural considerations. As a result, the speaker constructs a seven-syllable hump. The total length of this hump is 1250 ms, and the mean syllable length is 179 ms. In such cases, the perceived effect of lengthening is probably achieved by the total length of the hump, and the long, consistent, fall of pitch, which, in this case, covers the six syllables that follow the primary accented syllable *ta*.

conversation reaches high tones and Beni and Uri, which are arguing with Michal, address her simultaneously (overlap speech is omitted for convenience):

Extract 2. "Back to god" (2:39–2:45 min.)

```
  225  Beni:   banu,
                we.came
• 226           ve-od        xaver   ba,
                and-another  friend  came
  <one line is omitted>
→ 228           nige--n,
                he.played
  229           ve--,
                and
→ 230           gitara  afilu    hivi,
                guitar  even     he.brought
                'he even brought a guitar'
  <two lines are omitted>
→ 233           ve   darbuko--t,
                and  darbukas {drums}
```

The list items occur in lines 228, 230 and 233. Lines 228 and 230 contain verb phrases (there are no explicit subjects within these IUs – the subject is introduced in line 226), whereas line 233 contains a noun phrase – an additional complement of the verb *hivi* (brought) in line 230. In other words, these lines contain material for two separate lists (one embedded within the other), namely, a list of actions ('to play' and 'to bring') and a list of objects ('guitar' and 'drums'). However, the prosodic packaging of lines 228, 230 and 233 overrides these semantic/syntactic formal statuses and creates a three-item list of these lines. Selting (2007) has found that in German as well, the items of a list are not necessarily syntactically equivalent. In such cases, according to Selting, it is prosody that contributes most to the recognizability of lists as lists.

Finally, there is one further locus of variation in the syntactic structure of the lists. Quite commonly, when a speaker wishes to add another item to her list but perhaps has none in mind, she pulls out an item with no real content – it is empty or general. Jefferson (1990) refers to such items as "generalized list completers" and claims that in most cases speakers add such an item at the end of a list in order to maintain the preferred three-item list structure. Here, I will refer to these items as "dummy items", because they appear in the data not only as third "missing" items of three-part lists, but also, very commonly, as the second item of two-part lists (see Extract 6 in Section 3.2) and the fourth item of four-part lists. Moreover, in some cases the dummy item is not the last item of the list, as demonstrated in the following extract.

In the part of the conversation extracted here, Amir reminds Dalya of a conversation they once had, in which she commented about his bad mood.

Extract 3. "Intrusion of privacy" (00:40–00:45 min.)

```
•  57  Amir:   ve-az        amart,
                and-then     you.said
    58          ken,
                yes
    59          walla,
                PTCL
    60          ata  nira
                you  look
→   61          ata  nira  mevoa--s,
                you  look  bummed out {slang}
→   62          ve-ze--,
                and-so
→   63          roim  et-ze  al    ha-xiyux        šelxa,
                see   it     on    the-smile       yours
                'it is evident from your smile'
⇒   64          amarti  ken.
                I.said  yes
```

Lines 61–63 comprise the list itself, in which Amir reproduces Dalya's words in a constructed dialogue (Tannen 1986).[16] In line 62, the very common dummy item *ze*, which is literally a singular masculine demonstrative pronoun, appears as the second item in the three-part list. The reason for claiming that line 62 is a separate list-item, and not an extension of line 61, is prosody – this line is delivered as a separate IU, performed with a hump. The use of 2nd person in both line 61 *ata* 'you' and line 63 *šelxa* 'yours', together with the Camel Humps prosody, help determine the scope of the constructed dialogue list.

Table 5 summarizes the distribution of different syntactic structures across the instances of the Camel Humps pattern.

Table 5. Syntactic structure of list items across the instances of the Camel Humps pattern

	Syntactic structure	Number of instances	%
Syntactically equivalent list items	Clauses	30	45
	Polar questions	2	3
	WH questions	2	3
	Noun phrases	8	12
	Verb phrases	3	4
	Prepositional phrases	3	4
	Predicate-adjective phrases*	1	1
	Total	**49**	**72**

16. See Section 3.3 for a discussion of the use of the pattern in constructed dialogue.

Table 5. (*continued*)

	Syntactic structure	Number of instances	%
Syntactically non-equivalent list items	Clauses and dummy items	9	13
	Noun phrases and dummy items	1	1
	Mixed syntactic structure	10	14
	Total	**20**	**28**
	Grand total	**69**	**100**

* Present tense nominal clauses in Hebrew are usually constructed without a copula. In this single instance of the pattern (taken from the conversation "singles"), a nominal clause, *at yafa* 'you are pretty' (literally 'you pretty'), is the basis for a list of predicate-adjectives, with *yafa meod* 'very pretty' as the second item and *mexoeret* 'ugly' as the third item.

3. The function

In all inspected instances, speakers use the Camel Humps pattern to enumerate reinforcements for a counter-stance they take.[17] That is, speakers use the pattern to support a stance they take in response to some other stance (Englebretson 2007). Goffman (1976) argues that we should look at talk-in-interaction as "a sequence of response moves with each in the series carving out its own reference" (1976: 293). A reference, in his view, is whatever a move responds to, which by no means has to be the previous turn/s in the conversation. This view is consistent with the findings of this study – speakers use the Camel Humps pattern while responding to stances of various sources and types. Therefore, a stance to which a speaker responds with the Camel Humps pattern will be referred to here as "reference-stance".[18]

Instances of the Camel Humps pattern could be categorized by the type of reference-stance that they target: (a) reference-stance *taken* by another participant in the conversation; (b) reference-stance *attributed* to another participant in the conversation (by the producer of the Camel Humps pattern); and (c) reference-stance taken outside the current conversation and *reported* by the producer of the Camel Humps pattern. This categorization will be exemplified and

---

17. I use the term "reinforcements" throughout this chapter to describe the enumerated items in a Camel Humps list. This term captures the one common denominator of the enumerated entities in all inspected instances, namely, they are provided in order to support and strengthen the stance that the speaker takes.

18. The term "reference-stance" should not be mixed with the term "object of stance" (Du Bois 2007), which refers to the target of the evaluation. For example, in Extract 1 ("The kayak rower"), the "object of stance" is exercising alone and the "reference stance" is Oren's negative evaluation of it *ze mešaʕamem* 'it is boring'.

explained in the Sections 3.1–3.3. Table 6 presents the distribution of instances with respect to reference-stance type.

Table 6. Distribution of instances with respect to reference-stance type

Type of reference-stance	Number of instances	%
Taken by another participant	19	28
Attributed to another participant	14	20
Reported by the producer of the pattern	36	52
Total	**69**	**100**

3.1 Reference-stance taken by another participant

"The kayak rower" example was already discussed above (Extract 1 is repeated below). The conversation is held by Tslil and Oren (Tslil is a friend of Oren's wife), and the topic under discussion is exercising by oneself or in a group.

🎧 **Extract 1.** (repeated) – "The kayak rower" (1:48–1:57 min.)

```
  136  Oren:   ze      mešaʕamem,
                this    boring
                'it is boring'
  137            levad.
                alone
  138  Clil:   beemet?
                really?
  139            ani   sonet  laasot               im    axerim.
                I     hate   doing{exercising}  with  others
• 140            keilu  ani,
                like   I
• 141            ani   ohevet  laruc   harey,
                I     love    to.run  PTCL
                'I love to run as you know'
• 142            az ani raca,
                so I   run,
→ 143            ba-kecev       šeli--,
                in.the-rhythm  of.1SG
                'at my [own] pace'
→ 144            ba-zman        šeli--,
                in.the-time    of.1SG
                'in my [own] time'
⇒ 145            ani sonet laruc        im   od    mišehu.
                I   hate  to.run       with more  someone
                'I hate running with others'
  146 Oren:                            at   raca baxuc?
                                       you  run  outside?
```

The way Oren and Tslil are taking opposite stances in this example nicely expresses the Stance Triangle (Du Bois 2007): in lines 136–137, Oren negatively *evaluates*

exercising alone (*ze mešaʕamem, levad.* 'it is boring, alone'), and in response, in line 139, Tslil *positions* herself on the negative end of the affective scale with respect to exercising with others (*ani sonet laasot im axerim*, 'I hate exercising with others'). By doing so Tslil *disaffiliates*[19] with Oren. After Tslil expresses her counter-stance explicitly, she turns to support it by providing a list of reinforcements (lines 143–144), performed with the Camel Humps pattern (see an acoustic visualization of lines 143–144 in Figure 2). The reinforcements are that by exercising by herself, Tslil can run however she wants (at her own pace), and whenever she wants (in her own time).

Figure 2. PRAAT pitch-track of lines 143–144 in Extract 1 from "The kayak rower"

The next example is an additional instance in which the reference-stance is taken by another participant. However, unlike the previous example, here both the reference-stance and the counter-stance are not stated explicitly. The conversation is held by Yafit and Malka, Yafit's mother. The topic under discussion is the program TAGLIT (literally 'discovery'; this is the Birthright Israel program), which takes young Jews from around the world for a ten-day trip to Israel in order to strengthen their affiliation to Judaism and to the State of Israel. Before the part of

---

19. Du Bois (2007) uses the term "divergent alignment" for such cases, and Keisanen (2007) uses the term "disalignment". However, interactional linguists usually make a distinction between "alignment" and "affiliation" (Stivers 2008), according to which "aligning" means following the requirements of sequential positioning and activity, and affiliating means endorsing the coparticipant conveyed stance. Therefore, the term "disaffiliation" will be used here (see also Steensig & Drew 2008).

the conversation that is extracted below, Malka provides the reference-stance: she states that in a couple of generations, United States Jewry will disappear as a result of assimilation, thus implying that the program TAGLIT cannot achieve its aims. Yafit, who worked as a guide in one TAGLIT trip, responds:

🎧 **Extract 4.** "Discovery" (0:52–1:04 min.)

```
• 50  Yafit:  hiru          lanu  ba--,
               they.showed   us    in.the
• 51            haxana       le-taglit,
                preparations for-TAGLIT
   52  Malka:   aval ze      mamas  mm
                but  it-is   really mm
• 53  Yafit:                        kama xevre  še--,
                                    some people that
   54  Malka   ...še-ma?
                   that-what
   55  Yafit:  ..še--
                   that--
→ 56            davka        axre   taglit     hixlitu
                laalo--t,
                actually     after  TAGLIT     decided
                to.do.aliyah
   57           ..ve--
                   and
   58           ...ve--
                   and
→ 59            laasot cava--,
                to.do  army
                'to join the army'
→ 60            ..ve-lehišaer   ba-arec,
                   and-to.stay  in.the-country {Israel}
```

The disaffiliation revolves around the chances of the Birthright Israel program to achieve its aims. However, unlike Tslil in the previous example (see Extract 1 above), Yafit does not state explicitly that she is taking a counter-stance with respect to Malka, namely, that she believes the program will achieve its aims.[20] Instead, she turns straight away to provide reinforcements for her stance, and Malka needs to rely on that list in order to infer Yafit's stance. In lines 56 and 59–60, Yafit enumerates actions that are all symbols of strong affinity to the state of Israel: doing 'Aliyah', which means immigration of Jews to Israel; fulfilling the duty of military

---

20. Alongside the Camel Humps pattern, Yafit uses lexical means as a cue for disaffiliation (the word davka, 'actually'). However, since Malka's stance regarding the chances of TAGLIT program to achieve its aims is only implied, the use of the word davka at this point is not enough to understand what reference-stance Yafit is disaffiliating with. It is worth mentioning that the word davka has no straightforward translation to English. It might be translated 'actually' as it was here, or it might be translated 'in spite' or 'precisely', when it occurs with prosodic prominence which arguably is lacking here.

service, which indicates patriotism; and, eventually, staying in Israel permanently. (See an acoustic visualization of lines 56–60 in Figure 3). Stating that she knows of young people who participated in the Birthright Israel program and as a result did all of the above implies that, at least for some of its participants, the program does achieve its aims (contra Malka's beliefs).

In such cases, therefore, the enumeration of reinforcements with the Camel Humps pattern not only supports the stance but also functions as the act of *taking* a stance. That is, the speaker relies on the other participants' ability to infer from the list she provides the ad-hoc category (Barsalou 1983) "reinforcements of the stance X", i.e., to understand what is the stance that this list of reinforcements can support. This is consistent with Schiffrin (1994), who argues that when listing items, the accumulation of entities whose properties warrant their inclusion under one category tells not only about these entities, but also about the category itself. This is also similar to Ariel's (2015) "Higher Level Category *or* Construction", in which a higher-level category, which is the speaker's communicative goal, emerges from the alternatives she introduces.
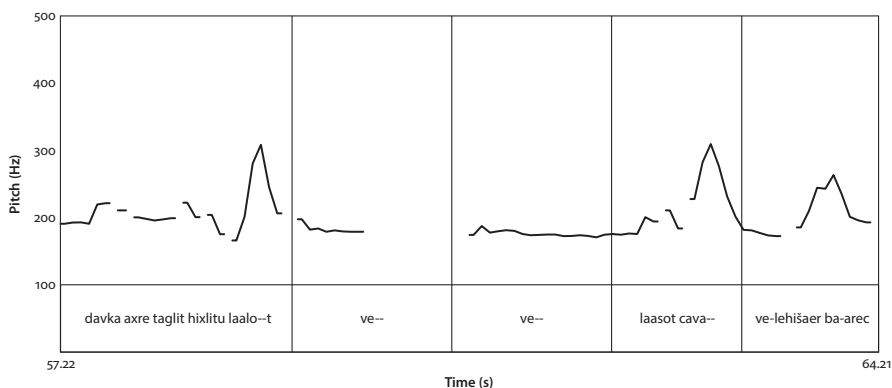
Figure 3. PRAAT pitch-track of lines 56–60 in Extract 4 from "Discovery"

3.2 Reference-stance attributed to another participant

The examples analyzed in Section 3.1 showed a situated conflict of stances between two participants. In the next two examples, the reference-stance is not taken or implied by another participant but rather is attributed to her. In such cases, the speaker who uses the pattern describes something that in her point of view is surprising,[21] and assumes that the co-participant/s will find what she describes

---

21. Usually because it stands in contrast to what is considered a "normal", or expected, state of affairs.

hard to believe. That is, she attributes to the co-participants/s an epistemic stance (e.g., Kärkkäinen 2003), which counters her own epistemic stance (naturally, she believes that what she describes is true). The Camel Humps pattern is used in such cases to cue the disaffiliation between these two opposite epistemic stances. The difference between the examples in Section 3.1 and the two following ones calls to mind Schiffrin's (1985) distinction between *oppositional* arguments and *rhetorical* arguments. According to Schiffrin, this is a distinction between "an interaction in which an opposition between speakers […] is negotiated through a conversation" (1985: 41), and "a genre in which the speaker establishes […] a position that is *potentially disputable*; in other words, speakers orient their presentation to a *potential opposition*." (1985: 40, emphasis mine).

The next example is taken from a conversation in which Nirit[22] is telling some relatives about her first experience in using a chatroom on the web:

Extract 5. "Nirit's chat" (00:03–00:17 min.)

```
  2  Nirit:   halaxti le-atarim    šel    e--h,
                I.went  to-websites of     eh
  3            ..arbaim     plus,
                 forty      plus
  4            ve-xamišim   plus.
               and-fifty    plus
  5            ...az xamišim    plu--s,
                 so  fifty      plus
  6            ..hayu,
                 there.were
  7            xamiša anašim.
               five   people
  8            ...šloša    lo        higivu,
                 three     did.not   respond
  9            ve-šnayim   niru    li    idyotim,
               and-two     seemed  to.me idiots
 10            al   ha-saf.
               on   the-doorstep {figure of speech}
               'right at the beginning'
• 11           ...arbaim    plus  hayu--,
                 forty      plus  there.were
→ 12           eize   matai--m,
               some   two.hundred
→ 13           šloš   meo--t,
               three  hundred
```

After describing the "fifty plus" (age) group as so small and lame (lines 5–10), it is possible to understand why Nirit assumes that her listeners would find the fact that

---

22. The corpus does not provide data regarding Nirit's age, but it can be inferred that she is not a young woman. At some point in the conversation, she refers to herself with humour as kšiša ('elderly').

the "forty plus" group was so busy hard to believe, or, in Schiffrin's (1985) words, "*potentially disputable*". After all, the titles of the groups do not suggest such a big difference. In other words, Nirit assumes something about her listeners' expectations regarding the "forty plus" group and treats these assumed expectations as an epistemic stance they hold. Her use of the Camel Humps pattern (lines 12–13) cues her disaffiliation with this (attributed) epistemic stance, i.e., her preemptive confrontation with this *"potential opposition"* (Schiffrin 1985). (See an acoustic visualization of lines 12–13 in Figure 4).

Nirit uses the pattern to enumerate her estimates regarding the number of users in the "forty plus" group. Her use of an open list pattern suggests that she could have added larger and larger estimates. Thus, it supports her stance that the number of users was surprisingly large.
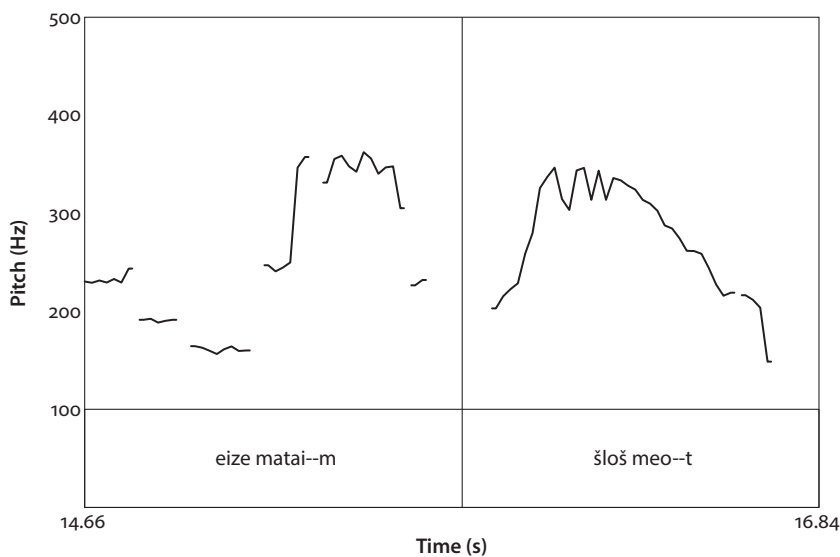
Figure 4. PRAAT pitch-track of lines 12–13 in Extract 5 from "Nirit's chat"

It is worth mentioning that this extract includes an example of a closed list (see Section 2 for the distinction between open and closed lists). In lines 3–4, Nirit enumerates the two chat rooms she visited. Line 3 ends with a rise-to-mid, which cues that the listing is about to continue, and line 4 ends with a fall-to-low, which cues that the listing is completed and that this list is exhaustive.

The next example is taken from a conversation in which Irit is telling about a time when she, her boyfriend (Yuval) and two other friends went to steal fruit from an orchard and from an avocado plantation.

Extract 6. "Picking off at avocado plantation" (1:01–1:07 min.)

```
• 79  Irit:    ve--  xataxnu          la--,
                and   we.cut-through   to.the
• 80             mm   Pardes,
                 mm   orchard
→ 81             ..alinu       šam          al   gdero--t,
                 ..we.climbed  over.there   on   fences
  82  Yuval:   rega,
               moment
               'wait  a moment'
→ 83  Irit:           ve-ze--,
                      and-this
  84  Yuval:              lifney   ze.
                          before   that
  85  Irit:    ma?
               what?
```

In lines 81 and 83 Irit constructs a two-item Camel Humps list, in which she describes the adventurous manner in which she and her companions got into the orchard (see an acoustic visualization of lines 81–84 in Figure 5).[23] Climbing an orchard fence in order to get in and steal fruit is not a normative behavior. In Irit's point of view, her listeners may find it hard to believe that she and her companions behaved this way. As in the previous example, Irit assumes something about her listeners' expectations and treats these assumed expectations as an epistemic stance they hold. She uses the Camel Humps pattern to cue her preemptive disaffiliation with this (attributed) epistemic stance.

It is worth mentioning that in this example something occurs which is very uncommon across the instances of the pattern. Usually, participants recognize the Camel Humps pattern as a cue for the construction of a list and allow the speaker to complete her long turn before claiming the floor. Here, Yuval interrupts Irit right after the first item of her list. Yuval's interruption might be explained by the fact that he himself is part of the narrative Irit is constructing – he was there with her in the orchard. Thus, he is "entitled" to be a narrator as well. Indeed, in an earlier part of the conversation, Irit and Yuval shortly discussed who should tell the story (lines 19–31 in the conversation) and decided that Irit would tell it.

---

23. Lines 82–84 are produced in overlap speech (indicated by the two text tiers in Figure 5). Because of the overlap speech, the pitch track of Irit's second list item is almost not visible. However, the peak of the second hump does appear (indicated by the two vertical dashed lines in Figure 5), since it overlaps with Yuval's voiceless labiodental fricative /f/, produced in the word lifney 'before'. (Note that in Figure 5 vertical lines do not indicate IU boundaries).
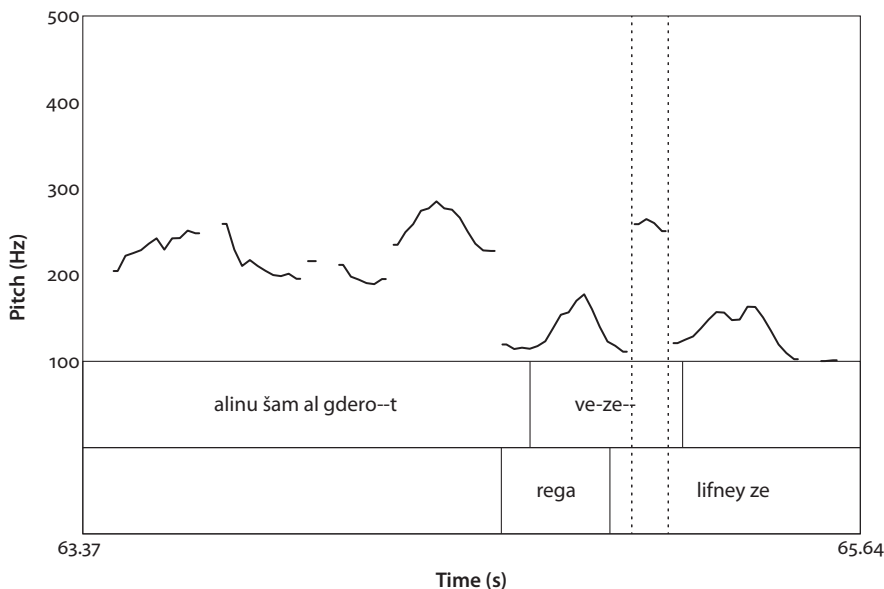
Figure 5. PRAAT pitch-track of lines 81–84 in Extract 6
from "Picking off at avocado plantation"

3.3 Reported reference-stance

The next two examples belong to a large subset of instances (36 out of 69) in which the reference-stance stems from *outside* the current conversation, that is, speakers disaffiliate with stance-takers that do not take part in the conversation. In all of these cases, speakers use the Camel Humps pattern to report the behavior of others – either their "doings" or "sayings" – and to evaluate it as illegitimate. The disaffiliation, therefore, revolves around the legitimacy of the reported behavior.

The next example is taken from a conversation in which Einat, who is a teacher, tells four fellow teachers about an act of vandalism performed by her students. The students broke into the school kitchen, turned on the ovens and left the place, thus causing a fire. In the following extract, Einat tells her colleagues how the school principle (Gabi) reacted to this incident. (See an acoustic visualization of lines 96–97 in Figure 6).

🎧 **Extract 7.** "Fire in school" (1:05–1:13 min.)

```
• 92    Einat:  ...gabi hitxil   lehaašim,
                   gabi started accusing
• 93            oti  ve-et  yafit.
                me   and    yafit
  94    B.C.:   wi,
                wi! {slang: a marker of astonishment}
  95            @
→ 96    Einat:  še-anaxnu @lo     @sagarnu   ta-tanuri--m,
                that-we   did.not shut.down the-ovens
→ 97            ...ve-še-lo     yitnu          yoter   maftexot
                le-mori--m,
                and-that-no  they.will.give  more    keys
                to-teachers
                'and that they will stop giving the [kitchen] keys to
                teachers'
```

Considering the fact that Einat knows that the students were to blame (and not her and the other teacher, Yafit), her evaluation of Gabi's reaction is rather clear – his accusations were out of place. The rest of the conversation turns to Gabi and his bad management, with a specific focus on similar cases in which he accused people for misbehaving, and even fired them, before finding out all the details of the event.
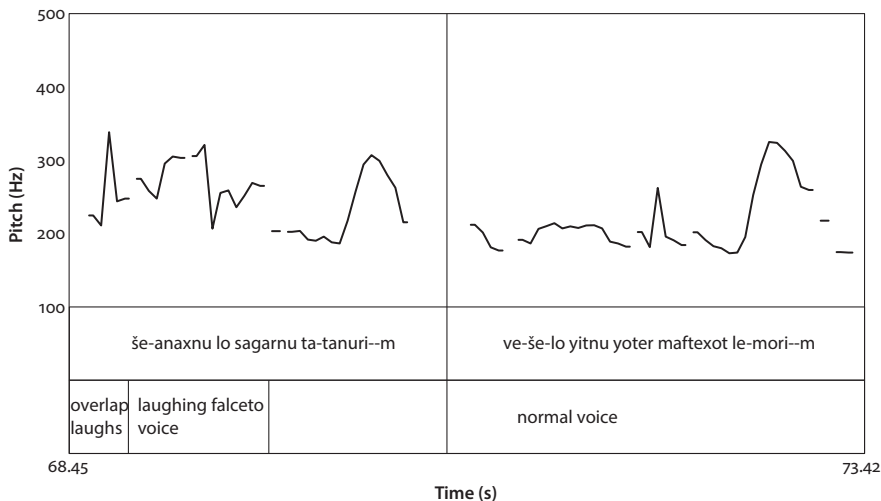
Figure 6. PRAAT pitch-track of lines 96–97 in Extract 7 from "Fire in school"

Another common case in which the reference-stance stems from outside the conversation (18 out of 36) is when speakers use the pattern in constructed dialogue (Tannen 1986), i.e., when speakers represent reported speech verbatim.[24] In all the

---

24. Although the reported action in the last Example (Extract 7 "Fire in school") is a "verbal" one, namely, an accusation, it is not a case of direct reported speech. It is rather a case of indirect

18 instances, speakers use the Camel Humps pattern in order to take a negative stance towards the reported speech – to evaluate it as "illegitimate". This is consistent with what Günthner (1999), following Bakhtin (1986), referred to as "polyphonic layering of voices": the speaker reconstructs past dialogue with one "voice", and, at the same time, evaluates the reconstructed utterances with another "voice". Günthner argues that it is prosody that enables speakers to add this evaluative layer.[25]

In the next example, Mary, who is a teacher, is telling a friend how she was forced to organize the Hanukkah celebration in school because no other teacher would do it. The celebration was a fiasco, and here Mary is reporting the reaction of the fellow teachers.

Extract 8. "Hanukkah celebration" (1:06–1:28 min.)   🎧

```
• 64  Mary:   bau    elay   morot,
              came   to.me  teachers
• 65          ..ve-amru    li,
                and-told   me
→ 66          ..ma   ze    ha-dava--r   ha-ze,
                what is    the-thing    the-this
              'what is this thing'
→ 67          ..ve-adif   še-lo   hayitem          osim
              tekes,
                and-better that-no you.would.have    do
              a.ceremony
              'it would have been better if you had not done a ceremony
              [in the first place]'
  68          ..ve-lo,
                and no...
  <ten lines omitted>
  79  Mary:   ..amarti     la,
                I.told     her
  80          ..az   eifo  hayit?
                so   where have.you.been
  81          ..lama  lo   irgant          et-ze?
                why   not  did.organize.you  this
              'why didn't you organize it?'
```

reported speech. Note, for example, the use of anaxnu 'we' and not of atem 'you'(2pl) in line 96. This differentiates the example in Extract 7 from the 18 examples in which the Camel Humps pattern is used in constructed dialogue. Therefore, I classified it as a case of reported "doings", together with examples in which the reported behavior is not verbal at all. This "forced" binary classification of a somewhat intermediate case does not affect the analysis of this example with respect to the pattern's function. In all cases in which the reference-stance stems from outside the conversation, the prosodic pattern is used as a means of adding an evaluative layer to the reported behavior, be it "doings" or "sayings".

25. It should be mentioned that when using the Camel Humps pattern for constructing a dialogue, a hearing of the prosody as if it were produced by the quoted person is not possible. This is a clear case of split voices.

Mary reports the unfair and illegitimate reaction of her colleagues in lines 64–68, namely, their harsh criticism about the celebration she organized. Lines 64–65 project the upcoming of the reported speech and lines 66–68 are the constructed dialogue itself, delivered with the Camel Humps pattern (see an acoustic visualization of lines 66–67 in Figure 7). Mary expresses her discontent with this reaction of her colleagues in her next turns. One example is found in lines 79–81 (this is another case of a constructed dialogue, only here Mary reproduces her own speech, and does not use the Camel Humps pattern). Mary did not ask the questions reported in lines 80–81 in order to receive information from her criticizer. Rather, these are what Günthner (1996) has identified as "reproach questions". In asking these questions, Mary is actually telling her criticizer that she has no right to criticize her, since she took no part in the organization.
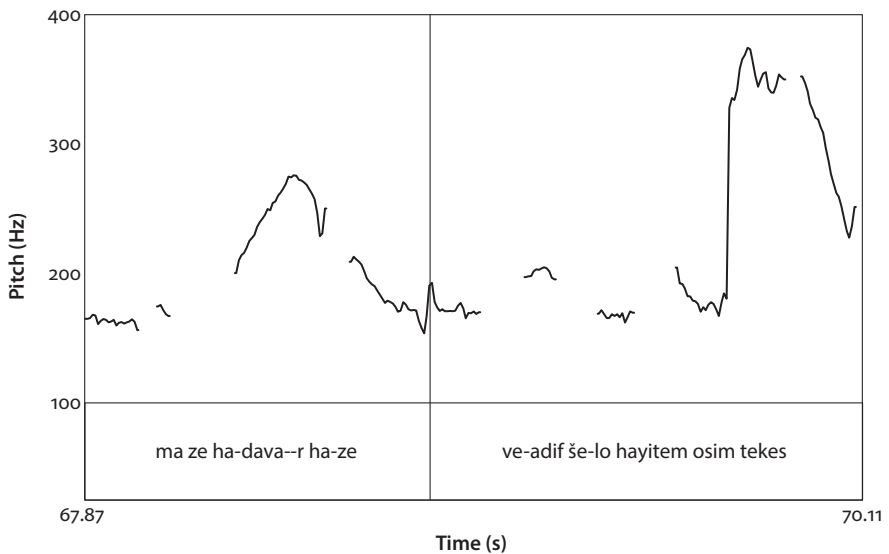
Figure 7. PRAAT pitch-track of lines 66–67 in Extract 8 from "Hanukkah celebration"

3.4   The Camel humps pattern as an exclusive cue of disaffiliation

The last example I analyze is another case in which the reference-stance is taken by another participant (see Section 3.1). However, in this example, the disaffiliation is cued solely by prosody. The example is taken from a conversation in which Liraz, Meggie, Yamit and Yosi argue about a hypothetical scenario, in which a woman that has a boyfriend meets her ex-boyfriend and gets emotionally confused. Throughout the conversation, the three women (Liraz, Yamit and Meggie) and the man (Yosi) take opposite stances. These opposite stances are exemplified in the following extract.

Extract 9a. "Love hurts" (0:40–0:48 min.)

```
 52  Liraz:  az    hi    crixa  lehipageš    im    ha
             so    she   needs  to.meet      with  the
 53          im    ha
             with  the
 54          ...ito       od     ha-paam,
             with.him    more   the-time
             'with him {the ex-boyfriend} once again'
 55          ..aval    lehagid    la-xaver          šela.
               but     to.say     to.the-boyfriend  her
             'but to tell her {current} boyfriend'
 56  Yosi:   ...ve-im    xaver        šela,
             and-if      boyfriend    her
             'and if her {current} boyfriend'
 57          eh,
             eh
 58          ..yiten       la     et-ze  ke-tnay?
               will.give   to.her  it    as-a.condition
             'will set it as a condition'
 59          ..im     at     nifgešet    ito,
               if     you    meet        with.him
 60          ..az     anxnu    mesaymim?
               so     we       are.done
```

In lines 52–55, Liraz phrases the position that she and the two other women hold throughout the conversation. Namely, that a woman in such a situation should share her doubts with her current boyfriend openly and meet her ex-boyfriend again in order to better understand what she feels. In lines 56–60, the man, Yosi, phrases his stance – a man in such a situation, i.e., the current boyfriend, would not agree for such a meeting to take place. Later in the conversation, Yosi says:

Extract 9b. "Love hurts" (2:19–2:23 min.)

```
218  Yosi:   ha-mexir    šel  ze     še-hi          im
             xaver       šela?
             the-price   for  that   that-she.is    with
             boyfriend   her
             'the price for being with her boyfriend'
219          ze   še-hi      lo     tipageš,
             is   that-she   not    will.meet
             'is that she won't meet'
220          im   af   exad   axer.
             with no   one    else
```

Here, unlike in Extract 9a, Yosi does not hypothesize about what an imagined boyfriend would say but uses his own words to state his stance explicitly – a woman cannot have a romantic meeting with another man while being in a relationship. With this background we can turn to analyze the example of the pattern:

🎧 **Extract 9c**. "Love hurts" (1:53–2:00 min.)

```
  177   Liraz: ...lifne,
                   before {meeting the ex-boyfriend}
  178          ...še-tagid            hakol        la-xaver
               šela,
                   that-she.should.say  everything    to.the-boyfriend
               her
                   'she should tell everything to her {current} boyfriend'
→ 179   Yosi:  še-tagid             še-hem    hayu paam
               that-she.should.say that-they were once
               xaveri--m,
               a.couple
  180   Liraz:     │ še-hem    hayu │ hakol,
                   │ that-they were │ everything
  181   Yamit:                      │ še-hem      hayu xaverim.
                                    │ that-they were friends
                                    │ 'that they were a couple'
  182   Liraz: ve-še-ze     bilbel      ota,
               and-that-it  confused    her
  183          ve-hi    roca   livdok,
               and-she  wants  to.check
  184          ve--
               and...
```

Liraz's contribution in lines 177–178 is the beginning of another attempt to phrase the stance that she took in lines 52–55 of Extract 9a. Line 178 ends with a slightly rising intonation that suggests that she was about to continue speaking. However, in line 179 Yosi interrupts her and produces a one-item Camel Humps list. The last syllable in line 179 *rim* is extremely lengthened (= 800 ms), and although it is impossible to extract F0 acoustically, the rise and fall pitch movement is clearly audible. Considering only the text of line 179 ('she should say that they were once a couple'), Yosi's contribution could be seen as going along with the position Liraz is taking, maybe even as a co-construction of a list (Lerner 1994). That is, according to Liraz, indeed everything should be told to the current boyfriend, including, of course, the fact that they were once a couple. However, it is very unlikely that Yosi is taking here a stance that stands in sharp contrast to the one he takes throughout the conversation. In Yosi's point of view, the things that should be told to the current boyfriend according to Liraz are all reinforcements for *his* stance. That is, these are things that should not, and cannot be told to a boyfriend. Yosi provides only one item to his list: "they were once a <u>couple</u>", but the Camel Humps prosody projects that he could have added more reinforcements to his stance. Yosi's stance is also supported by the implications of the one item he provides: they loved each other, they slept together and so on. Indeed, at an earlier point in the conversation, Yosi explicitly states that since the woman and her ex-boyfriend were a couple, and hence slept together, it is probable that they would do it again if they were to meet.

Liraz's response to Yosi's interruption (lines 180, 182–184) can serve as a "next turn proof" (Hutchby & Wooffitt 1998) to my analysis of Yosi's turn in two ways. First, when Liraz takes the floor back in line 180 she uses a louder voice and a higher global pitch than in lines 177–178. The difference in prosody makes it impossible to connect line 180 coherently with lines 177–178, and so it suggests a change of attitude that happened as a result of Yosi's interruption. Such "fight-back" responses, which follow several of the instances of the Camel Humps pattern, help prove that what was uttered with the Camel Humps pattern was indeed interpreted as an expression of a counter-stance.[26] Second, in line 180, Liraz produces an alternative clause to complete Yosi's subordinating construction "she should say X" (line 179). In this clause, Liraz changes Yosi's phrasing from *xaverim* 'couple' to *hakol* 'everything'. According to my interpretation, Liraz uses the word *hakol* 'everything' as an upgrade, which answers Yosi's challenge and the implications of his one-item list. This upgrade is yet another indication that Liraz interpreted Yosi's turn as an expression of disaffiliation.

Finally, there is an indication that Yosi's stance is cued solely by prosody, and not, for instance, by the above-mentioned implications. In line 181 Yamit, who holds the same stance as Liraz throughout the conversation, produces a completion of her own to Yosi's subordinating construction. Yamit in fact repeats Yosi's phrasing *še-hem hayu xaverim* 'that they were a couple' but uses a different prosodic pattern than he does. Thus, the two turns (line 179 and line 181) can be considered a "minimal pair", differing only in their prosodic design. I could not analyze this difference in prosody acoustically because of the overlap between turns, but I report here my auditory analysis of Yamit's turn in line 181. The first pitch movement in Yamit's turn is a jump up of ~6 semitones on the second syllable of the word *hayu* 'were'. Pitch is then held constant during the first and second syllables of the word *xaverim* 'couple', and eventually, on the last (accented) syllable *rim*, it rises another ~1 semitone and falls all the way back to the starting point. To my ears, this prosodic design in the current context conveys that saying 'that they were a couple' is not such a big deal as Yosi makes of it. Clearly, Yamit does not produce her turn to disaffiliate with Liraz. On the contrary, she recruits herself (Kendrick & Drew 2016) in order to aid Liraz in answering Yosi's challenge. The comparison of Yosi and Yamit's turns illustrates the role of prosody as a contextualization cue – nearly identical utterances performed in different prosodic designs form different actions altogether, namely, disaffiliation and recruitment.

---

26. Admittedly, the fact that Yosi interrupts Liraz might itself be a reason for her change of attitude. However, I do not believe this is the case here, and in other "fight-back" responses to the pattern this cannot be claimed either.

In summary, this example reveals the function of the Camel Humps pattern as a *cue of stance* and, more specifically, as a *cue of disaffiliation*. Of course, the prosodic pattern cues stance-taking and disaffiliation in all other instances as well, but it usually does so together with other means, such as an explicit stating of the stance (e.g., Extract 1 in Section 3.1), or the ad-hoc category inferred from the content of the list (e.g., Extract 4 in Section 3.1), in what Auer (1992) referred to as a "redundancy of cues". However, in this example there is no redundancy of cues: there is no explicit stating of stance, nor does the content of Yosi's one-item list reveal his stance. Therefore, it can be claimed that Yosi's disaffiliation with Liraz is cued solely by the prosodic pattern.

3.5   Interim summary

Speakers use the Camel Humps pattern to enumerate reinforcements for a counter-stance they take. In many cases, the stance is not stated explicitly, and the construction of the list is itself the overt act of taking a counter-stance. That is, enumerating items by using the Camel Humps pattern is sufficient to construct the ad-hoc category – reinforcements for a counter-stance. Moreover, in some cases, it seems that the stance-taking and the disaffiliation are only prosodically (and not lexically) cued. This leads to the claim that the Camel Humps pattern is a marker of disaffiliation in Hebrew.[27]

3.6   A different open list pattern in spoken Hebrew

The claim regarding the function of the Camel Humps pattern as a marker of disaffiliation could be supported by the presentation of open lists that are produced in a different prosodic structure and are not produced while taking a counter-stance. The following is a case in point.

The example is taken from a conversation in which Nurit describes a horrible episode she had with some medical doctors after her husband was injured.[28]

---

27. To clarify, the Camel Humps pattern is obviously not the only way to cue disaffiliation in spoken Hebrew, that is, it is not the case that whenever speakers disaffiliate, they use this pattern. However, whenever they use this pattern – they disaffiliate, and the very use of the pattern is one of the means to do so.

28. While describing this episode, Nurit provides an instance of the Camel Humps pattern. This instance is not analyzed in this chapter due to space limitations. However, it is the first instance in the audio clip of a sample of three instances provided in the beginning of the chapter.

However, there is one doctor named Dr. Abu whom she does not criticize. On the contrary, throughout the conversation she praises Dr. Abu and describes him as the only one who actually helped her husband. At one point in the conversation, Nurit reports one of Dr. Abu's attempts to aid her injured husband. Lines 162–163 are a constructed dialogue in which Nurit reconstructs an utterance of Dr. Abu's. These lines function as the projection component for the open list of Dr. Abu's actions, which are reported in lines 164–165.

Extract 10. "A vacation in Eilat" (2:10–2:14 min.)[29]   🎧

```
• 162  Nurit:   tov,
                 well
• 163            ..ten  li    ta-telefon    šelxa  ani    avarer,
                   give me    the-phone     yours  I
                 will.check
                 'give me your phone and I will check'
→ 164            ..ve-hu      hitkašer,
                   and-he     called,
→ 165            ve-hu  berer,
                 and-he checked
```

Although Nurit uses the open list to report the behavior of Dr. Abu (as in the examples analyzed in Section 3.3), it is clear that she does not do so in order to evaluate Dr. Abu's reported behavior as illegitimate. On the contrary, this list serves her to describe him as practical and willing to help. The acoustic visualization (Figure 8) makes clear the prosodic differences between this open list pattern and the Camel Humps pattern. Here, in each of the concatenated IUs (lines 164–165), the pitch curve starts relatively high and declines towards the intonation unit boundary. Importantly, unlike in the Camel Humps pattern, there is no rise-fall pitch movement on the primary accented syllables of the IUs (in both lines it is the last syllable). This example demonstrates two facts: (a) the Camel Humps pattern is not the only open list pattern in spoken Hebrew; (b) not every open list pattern in spoken Hebrew cues disaffiliation.

---

29. It is important to mention that although the symbol → in lines 164–165 marks the list items in this extract as in the other extracts in this chapter, this is not an instance of the Camel Humps pattern. Therefore, no part in these lines is underlined.
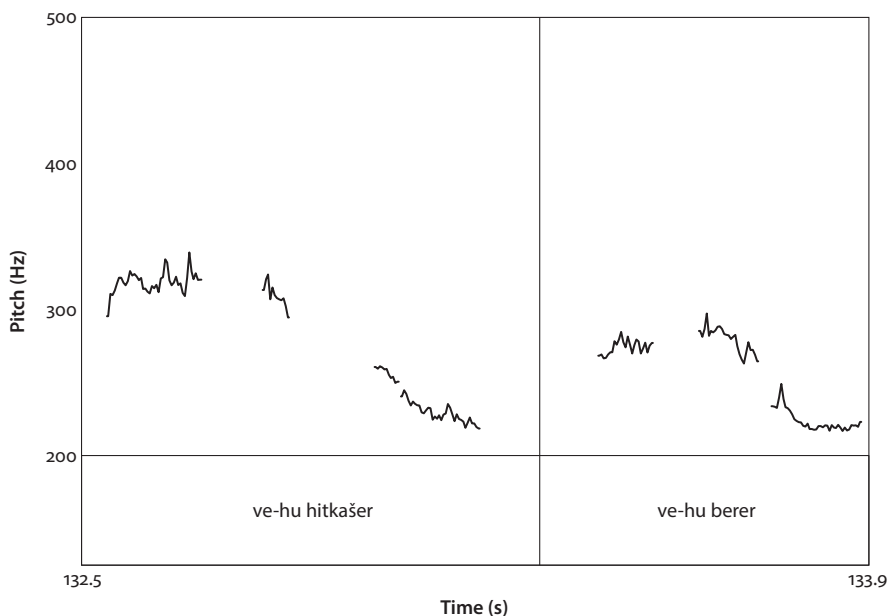
Figure 8. PRAAT pitch-track of lines 164–165 in Extract 10 from "A vacation in Eilat"

4. Conclusions

The findings of this study demonstrate that prosody plays a twofold role with re-
spect to lists in interaction. First, in its very basic ability to distinguish between
continuation and completion, and in its ability to equate utterances, prosody con-
tributes to the list-constructing mechanism, which enables speakers to include
enumerated entities under one category. Second, specific prosodic configurations,
such as the Camel Humps pattern, enable speakers to encode complex intersubjec-
tive meanings in the lists they construct (cf. Dori-Hacohen 2014).

The Camel Humps pattern is constructed by a concatenation of prosodically
similar intonation units that share two essential features: they end with a relatively
high rise-and-fall pitch movement, and their last syllables are lengthened compared
to the usual final lengthening of intonation units. The combination of these two
features is a marked prosodic behavior that makes the Camel Humps pattern a
distinct and identifiable prosodic form.

The Camel Humps pattern fulfills a cluster of functions in talk-in-interaction.
This cluster comprises functions from both the structural dimension and the inter-
actional dimension (Couper-Kuhlen 2009). In the structural dimension, since each
item in an open list projects more-to-come (Jefferson 1990), the pattern enables

speakers to take long turns, without being challenged for the floor. In the interactional dimension, at a primary level, the pattern serves speakers to accomplish the act of listing, i.e., to incorporate enumerated items under one category. This is so even if the speaker provides only one item (see Extract 9c in Section 3.4), or if she provides items of different semantic/syntactic statuses (see Extracts 2–3 in Section 2.2 and Extract 6 in Section 3.2). In addition, speakers use the pattern as a device for strengthening a stance they take. The act of listing reinforcements for their stance as an *open* list projects that more reinforcements could have been added. Furthermore, the marked prosody (the "humps") puts an extreme emphasis on each of the reinforcements the speaker provides for her stance, thus symbolizing the importance she attributes to each one of them.[30] Finally, the Camel Humps pattern functions as a contextualization cue which prompts an interpretation of what was said as an expression of a counter-stance, i.e., it is an overt act of disaffiliation. Unlike the rest of the functions that the pattern fulfills, there is no straightforward way to relate the pattern's function as a cue of disaffiliation to its prosodic features. This suggests that the Camel Humps pattern is a conventionalized marker of disaffiliation in spoken Hebrew.

Acknowledgements

References

Ariel, Mira. 2015. Higher-level category *or* constructions: When many is one. *Studies in Pragmatics* 17: 42–60.

Auer, Peter. 1992. Introduction: John Gumperz' approach to contextualization. In *The Contextualization of Language* [Pragmatics & Beyond New Series 22], Peter Auer & Aldo Di Luzio (eds), 1–38. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.22.03aue

Bakhtin, Mikhail. 1986. The problem of speech genres, trans. Vern W. McGee. In *Speech Genres & Other Late Essays*, Caryl Emerson & Michael Holquist (eds), 114–136. Austin TX: University of Texas Press.

Barsalou, Lawrence W. 1983. Ad hoc categories. *Memory and Cognition* 11(3): 211–227. https://doi.org/10.3758/BF03196968

---

30. See Günthner (1996: 287) on the use of vowel elongation and rise-fall pitch glide as amplificatio in German.

Barth-Weingarten, Dagmar. 2016. *Intonation Units Revisited: Cesuras in Talk-in-interaction* [Studies in Language and Social Interaction 29]. Amsterdam: John Benjamins. https://doi.org/10.1075/slsi.29

Boersma, Paul & Weenink, David. 2016. Praat: Doing Phonetics by Computer [Computer program]. Version 6.0.19. Amsterdam: University of Amsterdam.

Chafe, Wallace L. 1994. *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago IL: The University of Chicago Press.

Clayman, Steven E. 2013. Turn-constructional units and the transition-relevance place. In *The Handbook of Conversation Analysis*, Jack Sidnell & Tanya Stivers (eds), 150–166. Chichester: Wiley-Blackwell.

Couper-Kuhlen, Elizabeth. 1986. *An Introduction to English Prosody*. Tübingen: Niemeyer and London: Edward Arnold.

Couper-Kuhlen, Elizabeth. 2009. Prosody. In *The Pragmatics of Interaction* [Handbook of Pragmatics Highlights 4], Sigurd D'hondt, Jan-Ola Östman & Jef Verschueren (eds), 174–189. Amsterdam: John Benjamins. https://doi.org/10.1075/hoph.4.10cou

Couper-Kuhlen, Elizabeth. 2015. Intonation and discourse. In *The Handbook of Discourse Analysis*, 2nd edn, Deborah Tannen, Heidi E. Hamilton & Deborah Schiffrin (eds), 82–104. Malden MA: Blackwell.

Couper-Kuhlen, Elizabeth & Selting, Margret. 1996. Towards an interactional perspective on prosody and a prosodic perspective on interaction. In *Prosody in Conversation* [Studies in Interactional Sociolinguistics 12], Elizabeth Couper-Kuhlen & Margret Selting (eds), 11–56. Cambridge: CUP. https://doi.org/10.1017/CBO9780511597862.003

Couper-Kuhlen, Elizabeth & Selting, Margret. 2018. *Interactional Linguistics: Studying Language in Social Interaction*. Cambridge: CUP.

Dori-Hacohen, Gonen. 2014. More-than-three-part lists and their interactional achievements in radio phone-in shows. The 64th annual ICA Meeting. Seattle, WA.

Du Bois, John W. 2007. The stance triangle. In Englebretson (ed.), 139–182. https://doi.org/10.1075/pbns.164.07du

Du Bois, John W., Cumming, Susanna, Schuetze-Coburn, Stephan & Paolino, Danae. 1992. Discourse transcription. In *Santa Barbara Papers in Linguistics* 4. Santa Barbara CA: Department of Linguistics, UCSB.

Englebretson, Robert (ed.). 2007. *Stance Taking in Discourse: Subjectivity, Evaluation, Interaction* [Pragmatics & Beyond New Series 164]. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.164

Goffman, Erving. 1976. Replies and responses. *Language in Society* 5(3): 257–313. https://doi.org/10.1017/S0047404500007156

Gumperz, John. 1982. *Discourse Strategies*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511611834

Günthner, Susanne. 1996. The prosodic contextualization of moral work: An analysis of reproaches in 'why'-formats. In *Prosody in Conversation* [Studies in Interactional Sociolinguistics 12], Elizabeth Couper-Kuhlen & Margret Selting (eds), 271–302. Cambridge: CUP. https://doi.org/10.1017/CBO9780511597862.009

Günthner, Susanne. 1999. Polyphony and the 'layering of voices' in reported dialogues: An analysis of the use of prosodic devices in everyday reported speech. *Journal of pragmatics* 31(5): 685–708. https://doi.org/10.1016/S0378-2166(98)00093-9

Himmelmann, Nikolaus. P., Sandler, Meytal, Strunk, Jan & Unterladstetter, Volker. 2018. On the universality of intonational phrases: A cross-linguistic interrater study. *Phonology* 35(2): 207–245. https://doi.org/10.1017/S0952675718000039

Hutchby, Ian & Wooffitt, Robin. 1998. *Conversation Analysis: Principles, Practices and Applications*. Cambridge: Polity Press.

Jefferson, Gail. 1990. List construction as a task and resource. In *Interaction Competence*, George Psathas (ed.), 63–92. Lanham MD: University Press of America.

Kärkkäinen, Elise. 2003. *Epistemic Stance in English Conversation: A Description of its Interactional Functions, with a Focus on I Think* [Pragmatics and Beyond New Series 115]. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.115

Keisanen, Tiina. 2007. Stance taking as an interactional activity: Challenging the prior speaker. In Englebretson (ed.), 253–281.

Kendrick, Kobin H. & Drew, Paul. 2016. Recruitment: Offers, requests, and the organization of assistance in interaction. *Research on Language and Social Interaction* 49(1): 1–19. https://doi.org/10.1080/08351813.2016.1126436

Lerner, Gene H. 1994. Responsive list construction: A conversational resource for accomplishing multifaceted social action. *Journal of Language and Social Psychology* 13: 20–33. https://doi.org/10.1177/0261927X94131002

Lindström, Jan. 2009. Interactional linguistics. In *The Pragmatics of Interaction* [Handbook of Pragmatics Highlights 4], Sigurd D'hondt, Jan-Ola Östman & Jef Verschueren (eds), 96–103. Amsterdam: John Benjamins. https://doi.org/10.1075/hoph.4.06lin

Maschler, Yael, Polak-Yitzhaki, Hilla, Fishman, Stav, Miller Shapiro, Carmit, Goretsky, Netanel, Aghion, Gallith & Fofliger, Ophir. 2017. The Haifa Corpus of Spoken Hebrew. <http://textlink.ii.metu.edu.tr/haifa-corpus-spoken-hebrew> (29 March 2021).

Palmer, Harold E. 1922. *English Intonation, with Systematic Exercises*. Cambridge: W. Heffer and Sons.

Sacks, Harvey, Schegloff, Emanuel A. & Jefferson, Gail. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50: 696–735. https://doi.org/10.1353/lan.1974.0010

Schiffrin, Deborah. 1985. Everyday argument. In *Handbook of Discourse Analysis, Vol. 4: Discourse and Dialogue*, Teun van Dijk (ed.), 35–46. London: Academic Press.

Schiffrin, Deborah. 1994. Making a list. *Discourse Processes* 17(3): 377–406. https://doi.org/10.1080/01638539409544875

Selting, Margret. 2004. The 'upward staircase' intonation contour in the Berlin vernacular. In *Sound Patterns in Interaction. Cross-linguistic Studies from Conversation* [Typological Studies in Language 62], Elizabeth Couper-Kuhlen & Cecilia E. Ford (eds), 201–231. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.62.12sel

Selting, Margret. 2007. Lists as embedded structures and the prosody of list construction as an interactional resource. *Journal of Pragmatics* 39(3): 483–526. https://doi.org/10.1016/j.pragma.2006.07.008

Silber-Varod, Vered. 2011. The SpeeCHain Perspective: Prosody-Syntax Interface in Spontaneous Spoken Hebrew. PhD dissertation, Tel Aviv University.

Steensig, Jakob & Drew, Paul. 2008. Introduction: Questioning and affiliation/disaffiliation in interaction. *Discourse Studies* 10: 5–15. https://doi.org/10.1177/1461445607085581

Stivers, Tanya. 2008. Stance, alignment, and affiliation during storytelling: When nodding is a token of affiliation. *Research on Language and Social Interaction* 41(1): 31–57. https://doi.org/10.1080/08351810701691123

Tannen, Deborah. 1984. *Conversational Style: Analyzing Talk among Friends*. Norwood NJ: Ablex.

Tannen, Deborah. 1986. Introducing constructed dialogue in Greek and American conversational and literary narrative. In *Direct and Indirect Speech*, Florian Coulmas (ed.), 311–332. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110871968.311

Appendix. Transcription conventions

The lines in each transcription are numbered. Each line denotes an intonation unit. The lines consist of a phonemic representation of the Hebrew text, followed by an English gloss (PTCL = particle). Where this gloss is not close enough to an English utterance, it is followed by a third line supplying (usually) a literal translation. In the phonemic representation, the following characters are used to represent the Hebrew consonants that cannot be represented by a single English letter: 'c' for /ts/; 'š' for /sh/; and 'x' for /ch/ (voiced uvular fricative). The "humped" part of each IU is underlined. The transcription basically follows Chafe (1994), as adapted by Du Bois et al. (1992).

…	half second pause (each extra dot = another 0.5 second)
..	perceptible pause of less than half a second
(3.56)	measured pause of 3.56 seconds
,	(at end of line) continuing intonation ('more to come')
.	(at end of line) sentence-final falling intonation
?	(at end of line) sentence-final 'appeal intonation' (Du Bois et al. 1992)
	No marking of boundary tone at the end of a line indicates a truncated intonation unit
--	elongation of preceding vowel sound
@	a burst of laughter (each additional @ symbol denotes an additional burst)
⎣	Square bracket to the left of two consecutive lines indicates the beginning of overlapping speech, i.e. when two speakers are talking at once (e.g., lines 51–52 in Extract 4, "Discovery").
xxx	Lines aligned in such a way so that the right of the top line is placed over the left of the bottom line indicates latching, no inter-turn pause (e.g., lines 52–53 in Extract 4, "Discovery").
[xxx]	Material within square brackets in the gloss indicates exuberances of translation (what is not there in the original).
{xxx}	Transcriber's comments

Making the implicit explicit

Free enrichment in interaction

Christine Paul

Free University Berlin

Semantic underdeterminacy in the form of unformulated syntactic slots is a widely-discussed research theme, even though opinions differ about whether the interpretation is a result of pragmatic processes of free enrichment (e.g., Carston 2004; Recanati 2002, 2004; Sperber & Wilson 1986/1995) or of obligatory processes of the linguistic structure (Stanley 2007). Using a conversation analytic approach, a corpus of narrative interviews can show that interlocutors use a wide range of turns to deal with these forms of semantic underdeterminacy. In narrative interviews, prototypically the communicative functions of these turns (e.g. repair, gaining time or developing the topic) are in relation to the accessibility of possible candidates within the given context. The empirical findings support the relevance theory account of free enrichment.

Keywords: semantic underdeterminacy, free enrichment, relevance theory, implicit-explicit communication, narrative interviews, retrospective utterance, communicative function, other-initiated repair, gaining time, developing the topic

1. Introduction

Within conversation, speakers leave a lot of information implicit. This information either is of no interest or has to be inferred. According to relevance theory, explicatures are inferences concerning the proposition or the logical forms, such as, ad hoc concept building or free enrichment (Carston 2004, 633). While there is a consensus that the meaning of words, such as *she* or *there*, must be inferred obligatorily to grasp the proper meaning of the utterance, there is no unanimity about the inferences concerning the lacking information in the form of implicit syntactic slots as in the following examples:

(1) 1.1 That employee is good/ talented [in what respect?].
     1.2 They are leaving, because they've had enough [of what?].
     1.3 I've had a very large breakfast [today?].
     1.4 Mary took out her key and opened the door [with / without the key].

<div align="right">(Recanati 2002: 301–312)</div>

In the theoretical discussion, the terms used to label the lacking information as "unarticulated constituents" (Recanati 2002, 2004), "hidden indexicals" (Stanley 2002, 2007) and "free enrichment" (Carston 2004; Carston & Hall 2017) already show that these processes are described within different semantic and pragmatic accounts, are analyzed as either obligatorily or facultative and are discussed in terms of whether or not the implicit information is part of the truth-conditional content. Research on the subject has been mostly restricted to theoretical and cognitive issues while minor attention has been given to conversational data.

In contrast, research within conversation analysis deals with vagueness or ambiguities from the point of view of the interlocutor within the core phenomenon of repair, first analyzed by Schegloff, Sacks and Jefferson (1977), which focuses on problems of speaking, hearing and understanding. More precisely, the question concerning how interlocutors react to these problems in interaction is described by the term "other-initiated repair", e.g., clarifying the meaning by asking questions through formulations such as *you mean x*, etc. (for a detailed overview of the considerable amount of literature published on the concept of other-initiated, see Mazeland & Benjamin 2013). The following sequence provides an example: Speaker b describes himself driving a VW-bus in a crowded street inviting some people along without specifying to what or where they are going (literally "*einladen/ invite*"):

(2) B71w, 39, 1:28 min

```
1    B: hatt=n we n VAU we bus; °hh
        We had a VW bus
2       und dann kam die leutn und STROEMten da alle; =ne?
        and then the people came, and all flocked there, you know?
3       und dann ham wir da n paar LEUte EINjeladen;
        And then we invited some of the people
4 -> I: mit nach HAUse; oder?
        At home, did you?
5    b: NÖ:.
        NO:.
6       erstmal ins auto; […]
        First to the car
```

From the conversation analysis perspective, with turns such as *At home, did you?* (l.4), the speaker – B – realizes the other-initiated repair dealing with a problem of understanding: Since the information is lacking to what or where the speaker invited the people (line 3), the interlocutor – I – formulates an appendor question

(Sacks 1998: 660) proposing a possible argument for the verb *invite* that is *inviting* "*at home*" (line 4). Speaker B corrects the false inference and formulates the correct information *inviting "first to the car"* (line 6).

Research on the theoretical discussion and within conversation analysis concentrates on two different levels of the same phenomenon. Whereas within the theoretical discussion, the focus lies on sentences, within conversation, there is always contextual information available. Conversation analysis is a strictly empirical endeavor, explicitly not centering on cognitive processes (Deppermann 2015: 60). Therefore, conversation analysis focuses on the sequential organization, for instance, on the turn taking and utterance formats and on the conversation signals, such as, pauses to analyze how speakers deal with problems in understanding. On the contrary, relevance theory is a cognitive approach to explain communication, focusing on comprehension processes, e.g., distinguishing cognitive processes and different inference types.

Parting from the perspective that communication and understanding are joint labours in conversation (Clark 1996), the purpose of this paper is to join these two points together to bring new insights to the theoretical discussion. Since the cognitive processes are not automatically verbalized and the conversation analytic standard question is *Why that now?*, the analysis concentrates on the communicative function of elaborating the meaning of these implicit syntactic slots of the prior turn.

Within the specific research traditions, different terms are used; therefore, it is necessary to clarify the respective terms here. Throughout this paper, I term the lacking information the 'implicit syntactic slot', since the source of the inferences or the lacking information can be described within the syntactic structure, e.g., as the argument of the verb *invite* (Example (2), l.3–4). For an offered interpretation of the lexical level or for possible meanings available through the language context, I use the concept "candidate" from conversation analysis (Antaki 2012).

The findings of the study reveal that in narrative interviews, prototypically, the communicative functions of these turns serve to elaborate the meaning of the implicit syntactic slots related to the accessibility of the candidates within the given language context.

a. Interlocutors formulate clearly accessible candidates with strategic function, e.g., to show empathy or to gain time.
b. In the case of two or a restricted set of possible candidates in a given language context, interlocutors refer to the lacking information with a repair function.
c. Having no candidate accessible or endless possible candidates for the lacking information in this context, interlocutors refer to the lacking information to elaborate the topic or to continue the conversation.

The paper is organized as follows: After presenting the theoretical discussion about the inferences regarding the implicit utterance meaning in the form of syntactic slots (Section 2), I present the method and the data of the study (Section 3). The main body of the paper presents the findings concerning the conversational elaboration of these implicit syntactic slots focusing on the source (4.1), the utterance format that serves to construct the meaning of the prior utterance and (4.2) the communicative goals of the sequences (4.3). The article ends with a short concluding section.

2.   The Relevance theoretical perspective

From a Relevance Theoretic perspective (Sperber & Wilson 1986/95: 65), communication is based on inferences, whereby interlocutors construct interpretative assumptions of the communicative intentions of an utterance or a gesture based on the presumption of optimal relevance and the relevance principle.[1] Hence, inferences about certain schemata depend on their presumed relevance in a given context. There are different levels of inferencing, for instance, regarding deictic expressions (*she; here)*, concerning lexical items (e.g., understanding of metaphors and ad-hoc concept-building) or pertaining to implicit linguistic slots, such as, unarticulated constituents (Recanati 2002, 2004) or hidden indexicals (Stanley 2002, 2007). On the relevance-theoretic side, Carston (2004); Carston & Hall (2017) argues that inferences as in (1.1–1.4) are facultative pragmatic inference processes. While the linguistic structure of indexical expressions – such as, *she* – makes the inference processes obligatory, in sentences as in (1), the facultative inference processes are driven by the context and the principle of relevance. The pragmatically driven processes are called "free enrichment", in the sense of not being linguistically controlled, as is the case with indexical expressions.

   On the contrary, according to Stanley (2002, 2007), in sentences as in (1.1–1.4), the information in brackets is part of the linguistic structure of the expression, hence named "hidden indexicals". According to semanticists like Stanley, free pragmatic processes are unsystematic and unconstrained; hence, the implicit utterance meaning must be traced back to saturation processes of the logical form or the hidden indexical element of the utterance (Carston & Hall 2017).

---

1.   Presumption of optimal relevance:

a.   The utterance is relevant enough to be worth processing.
b.   It is the most relevant one compatible with the communicator's abilities and preferences.
     (Sperber & Wilson 2005: 474)

Furthermore, Recanati (2002, 2004) distinguishes different types of so-called "unarticulated constituents. In sentences such as (1.3; 1.4), without understanding the unarticulated constituent, the proposition is less specific: The hearer understands merely that *The speaker had a very large breakfast (at some (unspecified) time in the past)* and that *Mary opened the door (not necessarily with the key in question)*. On the contrary, in sentences like *It is raining*, or *They are leaving because they've had enough* the information must be contextually supplied to get a definite proposition. According to Recanati, in a wider sense, expressions like *enough* or *small* work just as indexicals as "*I*" or "*here*", where one needs to fill in the contextual variable (Recanati 2002: 301–312).

The theoretical discussion whether the cognitive processes are facultative or obligatory cannot be answered by conversational data. Nevertheless, the question remains whether these theoretical distinctions have a communicative relevance in interaction.

2.1   The conversational analytic perspective: Other-initiated repair

On the empirical side, conversation analysis parts from the assumptions that interlocutors display their understanding of the prior turns with the next one (e. g. Heritage 2009: 3). Second turns can be conditionally relevant, as in the case of question-answer sequences or not conditionally relevant, as in laughter, comments, appendor questions (Sacks 1998: 647) or other-initiated expansions (Sidnell 2012). Schegloff (2007: 217) calls them retro-sequences, since neither the structure nor the action of the first turn makes one expect the second turn. The most famous case of retro-sequences are sequences where another speaker initiated a repair that is solving problems of speaking, hearing and understanding (e.g., Drew 1997; Egbert 2009; Enfield et al. 2013; Hayashi, Sidnell & Raymond 2013; Schegloff et al. 1977; Schegloff 2000; Selting 1995; Svennevig 2008). For this paper, especially interesting are the other-continuations (or turn-expansions, Auer 2006; Sidnell 2012) and appendor questions. While often both forms are described as opposites realizing different action types, whereby interlocutors demonstrate understanding with other-continuations and with appendor question misunderstandings (Rost-Roth 2006: 277; Sacks 1998: 647; Szczepek 2000b), recent research suggests rather a difference of the epistemic attitude, since both utterance formats are grammatically coherent with the prior turn and share common characteristics within a sequence (Sidnell 2012: 326; Paul 2019: 123). For the empirical analysis, sometimes the more formal term "retrospective utterances" is preferred to the functional term of "other-initiated repair", since these utterance formats share characteristics in their sequential environment (second turns, semantically coherent, addressed to the speaker of the prior turn, getting a feedback) but have another function like initiating repair as well (Paul 2014).

3.   Method and data

3.1   Narrative interviews: The corpus of the fall of the wall

In this section, I take a conversation analysis approach and analyze a corpus of narrative interviews of around 60 hours of spoken German. The interviews were held between East and West Germans living in Berlin. Since the interviewees are acquaintances, friends and unknown interlocutors, the interviews are partly formal and informal. The purely audio data were transcribed using GAT II (Selting et al. 2009).

Muylaert et al. (2014: 185) defines narrative interviews as "unstructured tools", similar to everyday conversation. Typically, the interviewer asks open-ended questions, encouraging the interlocutors to speak and develop their answer sequences by listening and reducing their own speaking parts (Paulson 2011). "Based on the idea of reconstructing social events from the point of view of informants, the influence of the interviewer in narrative should be minimal" (Muylaert et al. 2014: 185).

The broad question of the interviews focused on how inhabitants of Berlin talk about the fall of the Berlin Wall and their experiences of a united Germany. Since telling stories is one of the central text-type of narrative interviews (Lucius-Hoene & Deppermann 2004) and communities construct their shared identities by telling stories (Murray 2018: 264), the interviewer made mostly comments or asked questions while the interviewees constructed their personal stories around the historical event.[2]

3.2   Data

In my data, I found 35 sequences where interlocutors used different utterance formats to elaborate syntactically implicit slots or what relevance theorist call "free enrichment" (Carston 2004). The analysis of the sequences concentrates on the following levels:

1. **The implicit source linguistically**: Within the theoretical literature, the linguistic structure of the implicit linguistic slot is not clearly defined. Therefore, the question is about what kind of information, within the linguistic structure of the source utterance, elicits more information for the interlocutor? How can the implicit syntactic slot be defined more exactly in linguistic terms?

---

2.   For a more detailed introduction to the corpus and a research overviews see Dittmar & Paul (2019).

2. **The interpretative assumption**: By the second turn, the interlocutor marks information from the first turn that is relevant to him or her, by eliciting more information or by formulating none, one or several candidate meanings. How does the interlocutor present the candidate, for instance, is the utterance declarative or interrogative?
3. **The communicative goal**: Verbalizing an interpretative assumption has certain communicative goals, such as meaning construction, demonstrating understanding or gaining time. The third turn shows the reaction in interaction and, therefore, provides insights about the communicative goals of the second turn.

According to these questions and levels of analysis, the empirical analysis has been divided into these three parts.

4. Empirical analysis

4.1 The implicit source term linguistically

Regarding the syntactic structures, the lacking information can be partly described as valence structures, since verbalized candidates or the element an interlocutor is asking for can be analyzed as a complement of a verb (3a), an adjective (3b) or a noun (3c) of the source term:

(3a) (B71w, 268, 10: 45 min) (*invited -> at home*)

```
1    A:    und dann ham wir da n paar LEUte EINjeladen;
           And then we invited some of the people
2 -> i:    mit nach HAUse; oder?
           At home, did you?
```

(3b) (B68W, 26:29 min) (*getting aggressive -> towards the sellers*)

```
1  in:    aber eh (.) also da hab ich gemerkt dass ich doch AGgressiv
          werde ja,
          ah well then I realized I am getting aggressive, you know,
2 ->i:    DENen;
          Towards them;
3 ->      [den verKÄUferinnen gegenüber;]
          Towards the sellers.
```

(3c) B51WF, 19:50 min (*perspective -> on what?*)

```
1   Eb51: hm; hm; und dann (.) wenn ihr so dann geSEHen habt eh
           gerREdet habt (.)
           Hm, hm and then when you see them or when you talked to them
2          konntet ihr da ihre SICHTweise nachvollziehen?
           Could you understand their perspectives?
3          (.)  [oder kam]
                 Or was
4 ->Alex: über [WAT jetzt?]
           On[what right now?]
```

Further, interlocutors refer to implicit syntactic slots of open comparative structures such as when A says *something is better / easier / more* and B proposes a candidate in the form of *then x*, filling in the compared element or asking for it. Additionally, I found sequences with prepositional phrases of perspectives and directions, for examples when A says: *from the coronal on* and B continues with *to the top?* or when A says something like *from the right-hand side* and B asks: *from where?*, elaborating the position of the speaker not verbalized in the utterance.

Notwithstanding, interlocutors can verbalize the candidate for the implicit syntactic slot within an independent structure, not fitting in the syntactic structure of the source turn:

```
(3)   B100ON, 482, 29:20 min (go -> by car)
1   a:    und dann NACHTS SPÄT alleine dann in richtung marzahn fahre;
           And then when I go out late at night in direction of marzahn
           [district of Berlin]
           […]
3 ->i:    fahrn sie AUto?
           Do you go by car?
```

In this sequence, the interlocutor refers to the implicit syntactic slot of the valence structure of the verb to *go [by bus/ car]*. Even though the linguistic structure of B's utterance does not fit in in the implicit syntactic slot of the valence structure of *go out* because it is not a prepositional phrase, the speaker does refer to *go out* through the question *Do you go by car* thereby verbalizing a possible candidate within a completely independent structure. With the question, B does refer to the broader context, implying that when one has a car it is used to go out late at night alone. Therefore, the speaker can refer to these implicit syntactic slots using utterance formats which do not fit grammatically into the slot, for instance, *Wh*-questions or *Do you mean X*? questions.

4.2   Possible meanings: Candidates and utterance formats

Within conversation, interlocutors use a wide range of declarative and interrogative formats to elaborate the meaning of the implicit syntactic slots, eliciting more information by formulating none, one or various candidates. In my data, interlocutors used the following utterance formats:

Table 1. Utterance formats and candidates

Verbalizing candidate meanings	Utterance format	35
No candidate	Wh- questions (*From where?*)	11
	Other-initiated increment (connectors, prepositions that project a continuation such as *Between…?*, cf. Lerner 2004))	2

Table 1. (*continued*)

Verbalizing candidate meanings	Utterance format	35
One candidate	Polar question with the finite verb on 1. Position (*Do you mean X?*)	2
	Appendor questions or utterance expansions with tag questions (e.g., *At home?*)	3
	Utterance expansions (e.g., *At home.*)	14
Various candidates	Alternative-questions *(at home or to the car?)*	1
Mixed forms	Various utterance formats in one turn (e.g., *Towards whom? Towards the sellers?*)	2

Based on the view that different interrogative formats serve to finetune the epistemic stance (Heritage & Raymond 2012), the different interrogative and declarative utterance formats are used to express interpretative assumptions by varying epistemic support (Sidnell 2012: 323; Paul 2014). Contrary to prior research that declares appendor questions and utterance expansions as opposites (Sacks 1998: 660; Szczepek 2000a; b), the data supports the analysis that both are part of an epistemic continuum (Sidnell 2012: 323; Paul 2014). Besides the interrogative and declarative formats, for instance, interlocutors vary the epistemic stance by epistemic modal adverbs (*maybe, probably*). To illustrate the different epistemic stances, have a look again at example (cited before in a shorter sequence as 3b):

```
(4)   B68W, 26:19 min
1  in:    aber eh (.) also da hab ich gemerkt dass ich doch AGgressiv
          werde ja,
          ah well then I realized I am getting aggressive, you know,
2 ->i:    DENen;
          Towards them;
3 ->      [den verKÄUferinnen gegenüber;]
          Towards the sellers.
4 ->in:   [den verkäuferinnen gegenüber.]
          Towards the sellers.
5         ja.
          yes
```

Obviously, the interlocutor could use different syntactic formats to realize other-initiated repair, as illustrated with the simplified and fictive Example (5):

(5)   A.   *I am getting aggressive.*
      B.   (1)   *Towards whom?*
           (2)   *Towards the sellers or towards the other clients?*
           (3)   *Towards the sellers?*
           (4)   *Towards the sellers, probably.*

Moreover, the utterance formats can be distinguished regarding the relationship to the prior turn: Turns are either formulated in a grammatically coherent way with the prior turn (appendor questions, utterance expansions, cf. 6. B1) or they are structurally independent (6. B2). Again, for the sake of simplicity, a simplified and fictive example:

(6) A. *I am getting aggressive*
    B. (1) *Towards the sellers.*
        (2) *You mean towards the sellers.*

While by the grammatically coherent continuation (6.B1) the interlocutor presents a candidate as a continuation of the prior turn – either on the lexical or grammatical level – no linguistic element indicates a speaker shift. On the contrary, in 6.B2, the speaker clearly indicates by *you mean* that he is proposing a candidate interpretation. According to Paul (2019, 123), formulating grammatically coherent continuations (e.g., utterance expansions, appendor questions) serve as an evidential strategy. Aikenvald (2004: 149) defines an evidential strategy as a grammatical technique that gets additional semantic properties regarding the inferential justification. In other words, by a grammatically coherent continuation, the interlocutor presents his own turn as the continuation of the prior, indicating the candidate meaning as implied. By expressing a stance marker of the speakership (*you mean*), the candidate might be inferred by the knowledge about the world and not by the language context.

    To sum up, different linguistic means can be specified regarding the question concerning how interlocutors verbalize candidates and construct the meaning of the implicit slots that remain in an interaction. On the one hand, on the lexical level, interlocutors verbalize none, one or several candidates. Additionally, by the interrogative and declarative utterance formats and by epistemic modal adverbs, interlocutors present a candidate by varying certainty indicating the epistemic stance. Besides, interlocutors mark the prior turn as an evidential source by grammatically coherent continuations marking the candidate as implied by the prior turn.

4.3 The communicative function of elaborating implicit syntactic slots

It is a standard conversation analytical claim that second turns serve to establish joined understanding (e.g. Heritage 2009: 3). The linguistic means analyzed in Section 4.2 serve to express the cognitive side of the inferential process. Nonetheless, conversation analysis parts from the assumption that everything, including for instance laughter or pauses, is meaningful in conversation, hence the standard

question of conversational analysis is "Why that now" (Heritage & Claymann 2010: 14). Therefore, within this section, the analysis concentrates on the communicative goals of the turns dealing with the implicit syntactic slots.

A huge amount of research describes the communicative goal of a turn depending on either the characteristics of the utterance or concentrating on the following turn. Thus, prototypically, appendor questions and utterance expansions are described as different action types, demonstrating misunderstanding in the case of appendor questions and understanding in the case of utterance expansions (Sacks 1998: 660; Szczepek 2000b). By concentrating on conversations between two interlocutors and on the verbal reaction of these turns, some authors argue for a gradual difference, showing that both utterance formats receive a reaction of the first speaker (Sidnell 2012: 326; Paul 2019: 99). Within this paper, I concentrate on the lexical level and the question concerning in what ways the candidate meaning, made explicit by the interlocutor, was accessible by the communicative context.

Offering a candidate interpretation can have different communicative goals in interaction. Within narrative interviews, the communicative goals of the turns seem to be dependent on the accessibility of the candidate interpretation. Theoretically, there are three possibilities regarding the contextual accessibility of the implicit syntactic slot of the utterance. Within a given context, there can be (a) only one possible candidate, (b) there are two or a restricted set of possible candidates or (c) there is none or endless possible candidates.[3]

To make clear that the function of the turn cannot only be traced back to the characteristics of the utterance format but on the accessibility of the candidate in a given context, I will demonstrate the communicative function through examples of utterance expansions and another utterance format. I will start in Section (4.3.1) by showing that interlocutors refer to slots with clearly accessible candidates with strategic function, e.g., to show empathy or to gain time. In Section (4.3.2), I will give examples illustrating that sometimes two or a restricted set of possible candidates are accessible in a given language context. In these cases, interlocutors use the turns to refer back to the lacking information with a repair function. Furthermore, in (4.3.3), I demonstrate that sometimes within the conversation no candidate is accessible or endless candidates are possible for the lacking information. In narrative interviews, prototypically, interlocutors refer to these slots in order to develop the topic or to continue the conversation.

---

3. Notwithstanding, one should have in mind that categorizing the discourse within these categories is an analytical decision not always as clear as it seems to be.

4.3.1 *Presenting accessible candidates with strategic function*

In conversation, I define candidates for implicit syntactical slots as accessible when the missing information is easy to infer, for instance, the topic of the conversation and the principles of coherence or the knowledge about the world. In the corpus of narrative interviews, I found 10 sequences where interlocutors formulate accessible interpretative assumptions about the implicit syntactical slots. Since the information is inferable, interlocutors formulate this information mainly by utterance expansions or declarative questions with strategic functions, for instance, demonstrating understanding and/or knowledge (Szczepek 2000b), showing active attention or gaining time. Since declarative utterance formats, such as utterance expansions or declarative questions, are commonly analyzed as presenting understanding, knowledge or certainty (Labov & Fanshel 1977: 100; Szczepek 2000b), the formal characteristics of the utterance illustrates the accessibility of the implicit meaning of the syntactical slot on a further level. To provide an illustration, see the following sequence, where the West Berlin Speaker S describes the people in East Berlin as different:

(7) B113WN(A), 91, 4:13 min

```
01   s:  und das: sieht man sofort (.) die BLIcke und die reaktionen;
         And immediately, you see the glances and the reactions
02       (---)
03       und da hab ich dann KEEne lust drauf.
         And I don't like that.
04       !also es passiert schon wenn wir in! den TIERpark in ostberlin
         gehn ne,
         actually, yet it happens when we go to the zoo in East Berlin,
         you know
05   i:  hm,
         hm
06   s:  die spaziergänger die da ein=m DA entgegenkommen die gucken
         anders.
         The people who walk around there look around differently.
07       (---)
08 ->i:  ALS hier.
         From here.
09   s:  ALS hier. hmhm;
         From here. hmhm
10       (2.8)
11   i:  ham sie irgendwie beKANnnte oder verWANDte irgendwie sowas
         im osten?
         Do you have any acquaintances or relatives or somebody like
         that in the East?
12       auch keine.
         You don't.
```

The interlocutor I offers a comparative class to the expression *differently* from the prior turn through the utterance expansion *form here* (line 8). Since in the discourse

before, speaker S constructs the East- and the West Germans as clear opposites, the candidate meaning *form here* is easily to infer because of the discourse topic and coherence principles. While the speaker refers with *da (there*, l.6) to East Berlin or to East Germany generally, speaker I refers with the opposite deictic local pronoun *here* (l.8) to West Berlin or West Germany.

Additionally, the high level of certainty indicated by the declarative utterance format demonstrates further how obvious the verbalized utterance meaning for the interlocutor is. According to Schegloff (1996), speakers confirm the candidate as implied in the prior turn by literal repetitions, exactly as one can see in line 8 and 9. Furthermore, regarding the communicative functions of the utterance, the local deictic pronoun "*here*" is too unprecise to serve to concretize the region and therefore the understanding. More plausible and further supported by the communicative signals that speaker S has finished her turn and there is already a small pause, speaker I reacts to the interactive needs and formulates a turn by herself. By formulating the supposed meaning of the utterance, she actively shows attention and interest, more than by using just a particle such as *hm* or *yes*. The continuation of the conversation supports the thesis that the utterance's only strategic function is showing empathy and active attention as well. After the confirmation of the candidate, there follows another pause, after which the interviewer starts with a new question about a new topic.

Formulating an accessible candidate can also be a strategy for gaining time, for instance, when a speaker does not know how to continue the conversation or how to answer a question, as in the following sequence. Teacher b is describing a twinning of his West Berlin school and an East Berlin school, speaking about the differences in the education based on the different political systems before the fall of the Berlin wall. B is talking exclusively about the East German teachers; hence, it is rather clear to whom the interviewer is referring to when asking for the differences in language:

(8)   B71w, 268, 10: 35 min

```
01   i:    hmhm; (.)
02         und eh sind IHNen grade wenn se so in der SCHUle arbeiten;
           And eh since you are working at school
03         auch in der SPRAche unterschiede aufgefallen?
           Did you notice difference in language, too?
04         (3)
05 ->b:    eh von/ bei LEUten die eben (.) aus: (.) [aus ostberlin-]
           Eh from/ between people that come well (.) from (.) [from
           East Berlin]
06   i:                                     aus  [OSTberlin] oder
           umgebung komm.
                                             Coming   from    [East
           Berlin] or surrounding.
07         hm.
           Hm.
```

```
08   b:    JA.
           Yes.
09         na se ham/ sie SPREchen eben noch das was ich auch spreche eben.
           Well, they do/ they speak just what I am speaking as well.
10         sie BERlinern öfter noch;=ne?
           They are talking more often Berlin dialect, y'know?
```

Through the turn *Eh of/ from people that come well (.) from (.) [from East Berlin]* (l.5), b starts to formulate a candidate meaning of the implicit syntactic slot of *difference [between x and y]*, projecting a continuation by the pauses and stretching. Ferrara (1992: 220) describes these declarative turns as hidden questions, since the speaker suggests knowledge or understanding by the declarative characteristics of the utterances, but through speaking slowly as well as stretches and pauses, the speaker elicits a continuation. Notwithstanding, since the topic of the discourse is the differences between East and West Berlin teachers and no other referent is introduced before, according to the principles of discourse coherence it is redundant to formulate that the interviewer refers to the difference between the people of the East and West. Both, that the prior speaker understands the turn as a question and that the candidate is inferable is supported by the matter of fact that the continuation *from East Berlin* is produced in overlap with the speaker of the prior turn (cf.l. 5–6).

The large pause of three seconds before formulating the turn shows that speaker b has problems answering the question properly. By formulating an accessible candidate meaning, speaker i is realizing repair and therefore gaining time in an elegant way. The continuation of the sequence illustrates further the strategic communicative function of gaining time by formulating a possible meaning, since even afterwards, speaker b does not answer immediately, rather only after several conversational particles (*hm, yes, well* l.7–9) and then changes his construction *se ham/ sie SPREchen* (*Well, they do/ they speak*).

Formulating an accessible candidate is redundant on the content side but allows interactive cooperation, by taking the turn and giving it back again and, therefore, continuing the conversation. Short answers, slow speech, stretching and pausing further support the strategic functions of these turns, such as, gaining time and/or demonstrating active attention, understanding or empathy.

4.3.2   *Realizing repair in the case of two or a restricted set of possible candidates*
Within a given language context, sometimes there are two or a restricted set of possible candidates accessible for the implicit syntactic slot. In these cases, interlocutors refer to the lacking information to realize the repair and deal with a problem upon hearing or understanding (Mazeland and Benjamin 2013). Within my corpus, I found ten sequences, whereby the short turns, often inserted or formulated in overlap, are prototypically for turns dealing with problems of understanding. Interlocutors use different utterance formats to elaborate the meaning of the prior

turn such as wh-questions, polar questions and utterance expansions, by which the
use of the different utterance formats serve to indicate the continuum of epistemic
support of the possible utterance meaning.

The following sequence illustrates how a speaker deals with a problem of un-
derstanding since for the implicit syntactic slot the language context offers several
possibilities. The West Berlin brothers, Norm and Alex, describe their conversations
with their East Berlin acquaintance discussing different subjects such as econom-
ics or cultural events. When the interviewer EB51 asks if they can understand the
perspectives of their acquaintance, it is not clear what he is referring to:

(9)   B51WF, 423, 19:42 min

```
01   Eb51:   woRÜBer redet ihr da zum beispiel?
             What do you talk about for example?
02   Norm:   WIRTschaftliche probleme [xx]
             Economic problems
03   Alex:                            [ja]
                                       Yes
04           oder auch kulTUR viel; =ne; (.)
             or a lot about cultural events, too, y'know
05           also hier THEater oder (.) KONzerte die da stattjefunden
             haben oder (.) –
             Well, theater or concerts that took place here or
06           also !WE!nich noch diese diesen alten probleme.
             Well, seldom these old problems
07   Eb51:   hm; hm; und dann (.) wenn ihr so dann geSEHen habt eh
             gerREdet habt (.)
             Hm, hm and then when you see them or when you talked to them
08           konntet ihr da ihre SICHTweise nachvollziehen?
             Could you understand their perspectives?
09           (.) [oder kam]
                  Or  was
10 ->Alex:   über [WAT jetzt?]
             On[what right now?]
11   Eb51:   JA: zum beispiel über WIRTschaftliche probleme am anfang –
             Well: on economic problems at the beginning for example.
```

Before answering the question whether Alex and Norm could understand the per-
spectives and opinions of their East Berlin acquaintance, Alex asks the interviewer
Eb51 to specify the utterance meaning. The implicit syntactic slot *perspectives [on
what]* (l.8–10) is essential to answer the question, since from the language context,
Eb 51 could refer to perspectives about *economic problems* (l.2), *cultural events*
(l.4) such as *theater or concerts* (l.5) or *these old problems* (l.6). The meaning of the
implicit syntactic slot changes the answer expectations, hence the turn *On [what
right now?]* (l.10) deals with a problem in understanding. In addition, the question
About what right now is inserted and produced in overlap with the continuation
of the prior turn, a signal for realizing repair, since the problems must be solved
immediately to continue the conversation. By the wh-question, *About [what right*

now?] (l. 10), Alex does not propose any candidate, since there are various candidates accessible in the prior discourse.

Just as by Wh-question, interlocutors can realize other-initiated repair by utterance expansions, as in the following sequence. The West Berlin speaker In talks about the joy at the beginning when people of East- and West Berlin met after the fall of the wall and how the shops in West Berlin where full of people from East Berlin. In describes it as a very touching moment to see how the people of West Berlin reacted, helped and cared about the people of East Berlin:

(10) B68W, 453, 25:40

```
01   in:   es war schon sehr sehr (.) beRUEHrend auch zu sehn -
           It was very touching to see
02         wie: wie hier die leute aus=m WESten reagiert haben.
           How the people here in the West reacted
03   i:    hmhm;
04   in:   und und eh geHOLfen habn;
           And and eh helped
05         und ihnen ZUgewandt warn;
           And how they cared
06         und sich MITgefreut haben;
           And were happy for them
07         und bei REICHelt war es auf einmal !VOLL!;
           And the supermarket Reichelt was suddenly full
08         UEBerall liefen die leute rum [mit kleinen] kindern;
           Everywhere people were running around with small children
09   i:                              [aber da fingn]  fingen die
           MEISten ja
           an^ geNERVT zu werden.
           But then the majority started getting fed up
10   in:   also da HAETte ich ihnen wirklich- hehehe
           Well, then I had really….
11   i:    ja: haha
           Yes haha
12   in:   DOCH das hab ich auch mitgekriegt.
           Yes, I noticed that, too.
13         das WAR auch so: eh von von verKÄUferinnen-
           This was - well the sellers
14         GAB es eh gab es eh SOLche kommentare;
           Made some comments
15         die ich geHÖRT hab.
           I've heard
16         aber eh (.) also da hab ich gemerkt dass ich doch AGgressiv
           werde;=ja?
           But eh well, then I noticed I got aggressive, you know?
17 ->i:    DENen;
           them
18 ->      [den verKÄUferinnen gegenüber;]
           Towards the sellers
19 ->in:   [den verKÄUferinnen gegenüber.] ja.
           Towards the sellers, yes.
```

```
20          und hab es auch geSAGT;
            And I said
21          dass ich eh/ das es doch SELBSTverständlich ist dass die
            erstmal gucken;
            That is absolutely normal that they look first
22   i:     ja.
            Yes.
23   in:    und nicht viel KAUfen KÖNnen.
            And cannot buy so much.
```

On the formal side, the sequence where the interlocutor proposes a candidate for the implicit left slot *I got aggressive* [*towards x*] (line 16), has the same features as in the sequence ((7), l.8 *"als hier"*) in Section 4.3.1. For the open complement slot of *aggressive*, speaker I formulates a candidate in the form of a prepositional phrase *towards the sellers*. As in sequence (7), the interlocutor In indicates, with final falling intonation, high epistemic support and the candidate is confirmed as implied by the literal repetition. Notwithstanding, regarding the accessibility of information, the sequence is similar to the sequence (9). Within the given language context, there are two possibilities *towards whom* speaker In could get *aggressive*, on the one hand *towards the sellers*, on the other *towards the people with the small children* (l.9), whereby the meaning is essentially different: would speaker In get aggressive towards the people, she would make a statement against the people from the East, would she get aggressive towards the seller, she would make a statement in favor of the people from the East. Consequently, on the functional side, speaker I realizes repair by proposing a candidate interpretation, since it is essential to understand the complement to get the proper meaning of the utterance. The declarative utterance format and the, therefore, expressed high epistemic support of the candidate results from the before expressed positive attitude against the people of East Germany.

Distinguishing between the epistemic support and the communicative function can explain why utterance expansions are analyzed as demonstrating understanding (Szczepek 2000b) and realizing other-initiated repair, simultaneously. The format of an utterance expansion that expresses an implied candidate by high epistemic support can be explained by the very positive attitude of speaker In towards the people of the East which is expressed earlier in the discourse. The continuation of the sequence supports the argument that the speaker realizes repair with the turn, since the repair-sequence is an insertion and, after the confirmation of the candidate, speaker In continues immediately with her description.

That the utterance format does not determine the communicative function but does indicate the epistemic support of a possible interpretation is supported by the different utterance formats that serve to deal with problems of understanding within a language context where two or a restricted set of possible candidates are available.

4.3.3 *Endless possible candidates or no candidate accessible*

On some occasions, neither the knowledge about the topic nor the principles of coherence can be used to infer the implicit syntactic slot; consequently, there is simply no candidate or endless possible candidates accessible for the lacking information. In my data, I found 15 sequences where the interlocutors refer to these implicit left slots that were difficult to infer, mostly, but not exclusively by wh-questions. In narrative interviews, interviewers typically refer to these slots with no accessible candidate to develop the theme of the interview and thus to continue the conversation. Developing a theme consists of elaborating on a subtheme, for instance, in specifying or giving reasons (Brinker & Hagemann 2001: 1254):

```
(11)  B55W, 103, 7:45 min
01   m:   aber ehm ich perSÖNlich hätte NIE geglaubt-
          But ehm I, personally, would never have believed
02        dass nach fünf jahren DOCH noch (.) DOCH noch so=n
          großer unterschied
          besteht.
          That they differ so so much after five years.
03 ->i:   WOrin besteht dieser unterschied;
          How do they differ?
04        (---)
05   m:   worin beSTEHT dieser unterschied;
          How do they differ….
06        (12)
07        also ich/ ich weiß nich ich DENK mal so: (.)
          Well, I/I do not know I thinks just
08        in FEHlender souveräniTÄT. (xx)
          lacking confidence.
```

The interviewer asks speaker m to specify her statement concerning the huge difference between the Germans of the East and the West five years after the fall of the Wall and what the difference consists of. Structurally, in German, the noun *Unterschied (difference)* has a slot for a prepositional phrase (*Unterschied in x – the difference consists in x, differ in x*). From the language context, it is not possible to specify what kind of difference speaker m is referring to. Several studies investigating wh-questions have pointed out that by wh-questions, interlocutors identify a trouble source or a problem of understanding, indicating the lack of information without expressing interpretative assumptions (or "candidates", "hypothesis", etc.) contrary to polar questions (e.g., Egbert 2009: 101; Heritage & Raymond 2012; Rost-Roth 2006: 123). Nonetheless, within narrative interviews, it is a communicative goal of the interviewers to make the interlocutors talk. Therefore, asking for the meaning of the precise utterance is not always dealing with problems of understanding, but it can also be a strategy for developing a subtheme, in this case, the described differences used to continue the conversation. That speaking about

the changes in the life of the East- and West Germans and their living together in the united Berlin is one of the main topics in the corpus supports the interpretation that the interviewer asks for the difference mainly in order to develop the topic differences between the East and the West Germans. The communicative signals of the sequence show that the interlocutors are rather developing the theme of *differences between East and West Germans* than realizing repair. Speaker m needs time to specify the meaning, since she repeats the question, and after the very large pause of 12 seconds (l. 6), she still takes time using phrases such as *I don't know, I think* (l. 7). The difficulties speaker m has in specifying the differences shows, moreover, that meaning is not only constructed by the interlocutor afterwards, but by the speaker also.

Frequently, interlocutors use wh-question to refer to slots that are difficult to infer to continue the conversation or to develop the topic (12 of 15 sequences in my corpus). Notwithstanding, interlocutors can develop the topic by referring to these slots with utterance expansions too, using a list format, as one can see in the following sequences. Similarly, on the topic of the sequence before, the West Berlin teacher P is answering the question concerning how the East Berlin teachers differ from the West Berlin teachers:

(12)   B75W, 286, 12:48 min

```
01   P:   Ansonsten kann ich also auch von ihrer grundsätzlichen
          EINstellung
          und so weiter
          Otherwise, I cannot determine any difference in any sense
02        jetzt NACH der Wende,
          now after the fall of the Wall,
03        also jetzt nach FÜNF Jahren,
          five years after,
04        GEnerell (.) keine Unterschiede (.) in dem Sinne feststellen.
          generally (.)regarding their basic attitude and so on.
05 ->I:   Also im Umgang mit SCHÜlern,
          Also dealing with students,
06 ->     [im Umgang mit dem ganzen] LEHRstoff und
          dealing with the whole learning material and
07   P:   [im Umgang mit SCHÜlern.]
          dealing with students
08   I:   und so das is-
          that is
09   P:   das WAR am Anfang SCHWIErig,
          this was difficult at the beginning,
10   I:   hm,
          hm,
11   P:   mit dem Lehrstoff,
          with the learning material,
```

Since teacher P generally negates any difference between the teachers from both sides, there is no need to specify what kind of difference I is referring to, hence, there are endless possible (or no accessible) candidates for the slot *difference [in the class x*]. The interviewer specifies the differences by formulating possible candidate meanings – that is, differences regarding the students and regarding the whole of the learning materials. In German, if the candidate has the form of a prepositional phrase, which specifies the expression "*difference*", it fits in the syntactic structure as utterance expansions in list format. On the functional side, since P negates the differences generally, it is difficult to continue this topic of the conversation. Notwithstanding, by verbalizing the possible candidates of the negated differences, the interviewer finds the possibility of continuing the topic elegantly, since discussing the difference between East and West is one of the crucial topics of the Berliner Wendekorpus. By making the meaning of the prior turns more precise, she asks P if his statement counts for the verbalized candidates as well. Consequently, P starts speaking again about the difference which he shortly before negotiated generally. This sequence is only one example to show that the speaker can develop the topic by an utterance expansion as well.

5.   Summary

This study set out with the aim of analyzing how interlocutors deal with lacking information in the form of implicit syntactic slots. Since in interaction there are endless implicit syntactic slots, the presented study was designed to determine the communicative functions and the linguistic means interlocutors use to elaborate the missing information. The examples analyzed in this study highlight the importance of the contextually given information for verbalizing inference processes. To sum up the empirical analysis, within narrative interviews, the communicative functions of these turns dealing with implicit syntactic slots depend on the accessibility of the information that is being inferred. Speakers formulate accessible candidates with a strategic function. In the case of two or more possible candidates, they refer to the slot to realize repair. Moreover, within narrative interviews, interlocutors refer to slots where no candidate or endless candidates are accessible to develop the topic and to continue the conversation. That the decisive factor is the accessibility of the information and not the utterance format is supported by the fact that utterance expansions, for instance, are used with all these communicative functions. Consequently, the interlocutor indicates the epistemic support of the candidate by the utterance format, whereby, the communicative function depends on the contextual information.

Thus, does the conversational data shed any light on the theoretical discussion? Taking a conversational approach, one can see that within conversation there are endless cases of slots which interlocutors cannot easily infer. Notwithstanding – as the analysis shows – in conversation, not all implicit slots are of equal importance, since some inferences give relevant information and others not. The findings support a Relevance Theorist account of free enrichment. Furthermore, the analysis shows the empirical relevance of Recanatis' (2002, 2004) distinction between the implicit meaning of syntactic slots that can result either in vagueness or in misinterpretation. However, in conversation, both the candidates and the possible misinterpretations do not only depend on the structural characteristics of the utterance but on the given context as well.

It is important to bear in mind that the discourse type is decisive for certain communicative functions. Within the narrative interviews of the Berliner Wendekorpus, maintaining the conversation and developing themes, such as differences between the East and West Germans, are overarching communicative goals. Hence, the vagueness of language can be helpful for the interviewers to continue conversations by encouraging interviewees to specify meanings and develop subthemes. Therefore, in order to describe the communicative functions of turns that serve interlocutors to make explicit the further implicit syntactic slots within narrative interviews, the accessibility of information plays an important role. Further work is required to analyze how interlocutors elaborate the meaning of implicit syntactic slots in other discourse types.

6.   Transcript conventions GAT II (Selting et al. 2009, selection)

[]	overlap and simultaneous talk
=	fast, immediate continuation with a new turn or segment (latching)
(.)	micro pause, estimated, up to 0.2 sec. duration appr.
(-)	short estimated pause of appr. 0.2–0.5 sec. duration
(--)	intermediary estimated pause of appr. 0.5–0.8 sec. duration
(---)	longer estimated pause of appr. 0.8–1.0 sec. duration
(2.0)	measured pause of appr. 2.0 sec. duration
:; ::; :::;	lengthening (02.–1.sec.)
'	cut-off by glottal closure
akZENT	focus accent
ak!ZENT!	extra strong accent
?	rising to high
,	rising to mid
;	falling to mid

.	falling to low
–	level
↑	smaller pitch upstep
↓	smaller pitch downstep
()/(xx)	unintelligible passage
(das/was)	assumed wording, possible alternatives
<<smiling>>	actions and events with indication of scope
((wheeze))	para- and extralinguistic actions and events
°h, °hh, °hhh	inbreaths of appr. 0.2–0.5, 0.5–0.8, 0.8–1 sec. duration
h°, hh°, hhh°	outbreaths of appr. 0.2–0.5, 0.5–0.8, 0.8–1 sec. duration
<<f>>	forte, loud
<<ff>>	fortissimo, very loud
<<p>>	piano, soft
<<pp>>	pianissimo, very soft
<<all>>	allegro, fast
<<len>>	lento, slow
<<cresc>>	crescendo, increasingly louder
<<dim>>	diminuendo, increasingly softer
<<acc>>	accelerando, increasingly faster
<<rall>>	rallentando, increasingly slower

References

Aikhenvald, Alexandra Y. 2004. *Evidentiality*. Oxford: OUP.

Antaki, Charles. 2012. Affiliative and disaffiliative candidate understandings. *Discourse Studies* 14: 531–547. https://doi.org/10.1177/1461445612454074

Auer, Peter. 2006. Increments and more. Anmerkungen zur augenblicklichen Diskussion über die Erweiterbarkeit von Turnkonstruktionseinheiten. In *Grammatik und Interaktion*, Arnulf Deppermann, Reinhard Fiehler & Thomas Spranz-Fogasy (eds), 279–294. Radolfzell: Verlag für Gesprächsforschung.

Brinker, Klaus & Hagemann, Jörg. 2001. Themenstruktur und Themenentfaltung in Gesprächen. In *Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung*, Vol. 2 [Handbücher zur Sprach- und Kommunikationswissenschaft 16.2]), Klaus Brinker, Gerd Antos, Wolfgang Heinemann & Sven F. Sager (eds), 1252–1263. Berlin: De Gruyter.

Carston, Robyn. 2004. Relevance theory and the saying/implicating distinction. In *Handbook of Pragmatics*, Laurence Horn & Gergory Ward (eds), 633–656. Oxford: Blackwell.

Carston, Robyn & Hall, Alison. 2017. Contextual effects on explicature: Optional pragmatics or optional syntax? *International Review of Pragmatics* 9(1): 51–81. https://doi.org/10.1163/18773109-00901002

Clark, Herbert. 1996. *Using Language*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511620539

Deppermann, Arnulf. 2015. Retrospection and understanding in interaction. In *Temporality in Interaction* [Studies in Language and Social Interaction 27], Arnulf Deppermann & Susanne Günthner (eds), 57–94. Amsterdam: John Benjamins. https://doi.org/10.1075/slsi.27.02dep

Norbert, Dittmar & Paul, Christine. 2019. *Sprechen im Umbruch. Zeitzeugen erzählen und argumentieren rund um den Fall der Mauer im Wendekorpus*. Mannheim: Leibnitz-Institut für Deutsche Sprache.

Drew, Paul. 1997. ‚Open‘ class repair initiators in response to sequential sources of troubles in conversation. *Journal of Pragmatics* 28(1): 69–101. https://doi.org/10.1016/S0378-2166(97)89759-7

Egbert, Maria. 2009. *Der Reparaturmechanismus in deutschen Gesprächen*. Mannheim. <http://www.verlag-gespraechsforschung.de/2009/egbert.htm> (5 March 2020).

Enfield, Nicholas J., Mark Dingemanse, Julija Baranova, Joe Blythe, Penelope Brown, Tyko Dirksmeyer, Paul Drew, Simeon Floyd, Sonja Gipper, Rósa Gísladóttir, Gertie Hoymann, Kobin H. Kendrick, Stephen C. Levinson, Lilla Magyari, Elizabeth Manrique, Giovanni Rossi, Lila San Roque & Francisco Torreira. 2013. Huh? What? – A first survey in 21 languages. In *Conversational Repair and Human Understanding*, Makoto Hayashi, Geoffrey Raymond & Jack Sidnell (eds), 343–380. Cambridge: CUP. https://doi.org/10.26530/OAPEN_630828

Ferrara, Kathleen. 1992. The interactive achievement of a sentence: Joint productions in therapeutic discourse. *Discourse Processes* 15(2): 207–229.
https://doi.org/10.1080/01638539209544809

Hayashi, Makoto, Raymond, Geoffrey & Sidnell, Jack. 2013. Conversational repair and human understanding. An introduction. In *Conversational Repair and Human Understanding*, Makoto Hayashi, Geoffrey Raymond & Jack Sidnell (eds), 1–40. Cambridge: CUP.

Heritage, John. 2009. Conversation analysis and instituional talk: Analysing data. In *Conversation Analysis* [Sage Benchmarks in Social Research Methods 1], Paul Drew & John Heritage (eds), 1–26. London: Sage.

Heritage, John & Claymann, Steven. 2010. *Talk in Action. Interactions, Identities and Institutions*. Malden MA: Wiley-Blackwell. https://doi.org/10.1002/9781444318135

Heritage, John & Raymond, Geoffrey. 2012. Navigating epistemic landscapes: Acquiescence, agency and resistance in responses to polar questions. In *Questions: Formal, Functional and Interactional Perspectives*, Jan P. de Ruiter (ed.), 179–192. Cambridge: CUP.
https://doi.org/10.1017/CBO9781139045414.013

Labov, William & Fanshel, David. 1977. *Therapeutic Discourse: Psychotherapy as Conversation*. New York NY: Academic Press.

Lerner, Gene. 2004. On the place of linguistic resources in the organization of talk-in-interaction: Grammar as action in prompting a speaker to elaborate. *Research on Language and Social Interaction* 37(2): 151–184. https://doi.org/10.1207/s15327973rlsi3702_3

Lucius-Hoene, Gabriele & Deppermann, Arnulf. 2004. *Rekonstruktion narrativer Identität: Ein Arbeitsbuch zur Analyse narrativer Interviews*. Wiesbaden: VS Verlag für Sozialwissenschaften.
https://doi.org/10.1007/978-3-663-11291-4

Mazeland, Harrie & Benjamin, Trevor. 2013. Conversation analysis and other-initiated repair. In *The Encyclopedia of Applied Linguistics*, Carol A. Chapelle (ed.), 1068–1075. Malden MA: Wiley-Blackwell.

Muylaert, Camila, Sarubbi , Jr, Vicente, Rogério Gallo, Paulo, Rolim Neto, Modesto Leite & Advincula Reis, Alberto Olavo. 2014. Narrative interviews: An important resource in qualitative research. *Revista da Escola de Enfermagem da U S P* 48(2): 184–189. <http://www.scielo.br/pdf/reeusp/v48nspe2/0080-6234-reeusp-48-nspe2-00184.pdf> (5 March 2020).
https://doi.org/10.1590/S0080-623420140000800027

Murray, Michael. 2018. Narrative data. In *Sage Handbook of Qualitative Data Collection*, Uwe Flick (ed.), 264–279. London: Sage. https://doi.org/10.4135/9781526416070.n17

Paul, Christine. 2019. *Retrospektive Äußerungen. Nachfragen und Erweiterungen in narrativen Interviews am Beispiel des Berliner Wendekorpus*. Mannheim: IDS.

Paul, Christine. 2014. The epistemic side of retrospective utterances. *Language and Dialogue* 4(1): 24–42. https://doi.org/10.1075/ld.4.1.02pau

Paulson, Susan. 2011. The use of ethnography and narrative interviews in a study of 'cultures of dance'. *Journal of Health Psychology* 16(1): 148–157. https://doi.org/10.1177/1359105310370500

Recanati, François. 2004. *Literal Meaning*. Cambridge: CUP.

Recanati, François. 2002. Unarticulated constituents. *Linguistics and Philosophy* 25: 299–345. https://doi.org/10.1023/A:1015267930510

Rost-Roth, Martina. 2006. *Nachfragen. Formen und Funktionen äußerungsbezogener Interrogationen*. Berlin: De Gruyter. https://doi.org/10.1515/9783110912630

Sacks, Harvey. 1998. *Lectures on Conversation*, Vol. I. Oxford: Blackwell.

Schegloff, Emanuel A. 2007. *Sequence Organization in Interaction. A Primer in Conversation Analysis*, Vol. 1. Cambridge: CUP. https://doi.org/10.1017/CBO9780511791208

Schegloff, Emanuel A. 2000. When others initiate repair. *Applied Linguistics* 2(2): 205–243. https://doi.org/10.1093/applin/21.2.205

Schegloff, Emanuel A. 1996. Confirming allusions: Toward an empirical account of action. *The American Journal of Sociology* 102(1): 161–216. https://doi.org/10.1086/230911

Schegloff, Emanuel A., Jefferson, Gail & Sacks, Harvey. 1977. The preference for self-correction in the organization of repair in conversation. *Language* 53: 361–382. https://doi.org/10.1353/lan.1977.0041

Selting, Margret, Auer, Peter, Barth-Weingarten, Dagmar et al. 2009. Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 10: 353–402. <www.gespraechsforschung-ozs.de/heft2009/px-gat2.pdf> (5 March 2020).

Selting, Margret. 1995. *Prosodie im Gespräch. Aspekte einer interaktionalen Phonologie der Konversation*. Tübingen: Niemeyer. https://doi.org/10.1515/9783110934717

Sidnell, Jack. 2012. Turn continuation by self and by other. *Discourse Processes* 49(3–4): 314–337. https://doi.org/10.1080/0163853X.2012.654760

Sperber, Dan & Wilson, Deidre. 1986 [1995]. *Relevance. Communication and Cognition*. Oxford: Blackwell.

Sperber, Dan & Wilson, Deirdre. 2005. Pragmatics. In *The Oxford Handbook of Contemporary Philosophy*, Frank Jackson & Michael Smith (eds), 468–501. Oxford: OUP.

Stanley, Jason. 2007. *Language in Context. Selected Essays*. Oxford: OUP.

Stanley, Jason. 2002. Context and logical form. *Linguistics and Philosophy* 23: 391–434. https://doi.org/10.1023/A:1005599312747

Svennevig, Jan. 2008. Trying the easiest solution first in other-initiation of repair. *Journal of Pragmatics* 40(2): 333–348. https://doi.org/10.1016/j.pragma.2007.11.007

Szczepek, Beatrice. 2000a. Formal aspects of collaborative productions in English conversation. *InLiSt – Interaction and Linguistic Structures* 17, June 2000. <http://www.uni-potsdam.de/u/inlist/issues/17/index.htm> (5 March 2021).

Szczepek, Beatrice. 2000b. Functional aspects of collaborative productions in English conversation. *InLiSt – Interaction and Linguistic Structures* 21, December 2000. <http://www.uni-potsdam.de/u/inlist/issues/21/index.htm> (31 May 2018).

Online text mapping

The contribution of verbless constructions in spoken Italian and French

Carmela Sammarco

University of Salerno

This is a corpus driven study on verbless constructions (Cx$_s$[ØVP]) in spoken Italian and French that have a textual naming function, namely verbless constructions that speakers use to name what they are going to talk about. This paper suggests that these constructions are tools that speakers use as structure's mapping in the discourse to orient the listener and themselves and that build meta-discursive and meta-textual *ad hoc* categories. These structures are frequent in very high informative texts (i.e. meetings and conferences) and their occurrence correlates with the information's density in texts.

Keywords: verbless constructions, spontaneous speech, meta-textual, meta-discursive categories

1. Introduction

The present paper deals with a group of verbless constructions (hence Cx$_s$[ØVP][1]) that have the function of naming and organizing parts of texts (meta-textual Cx$_s$[ØVP]). In this study Cx[ØVP] are defined as every syntactically and prosodically autonomous structure without any finite verb, either predicative or not, and to which speakers associate a meaning (De Mauro 1974). Four macro-categories of Cx$_s$[ØVP] have been identified (see points I–IV).

---

[1]   The abbreviation Cx[ØVP] is composed of the symbols Cx that means Construction (Masini 2017), ØVP that stands for 'without any verbal finite form'. The label *verbless* is used not because it is assumed that the verb could be crucial for identifying syntactic constructions, but because it is almost an idiomatic expression used in linguistics to refer to some specific constructions that have in common the absence of the verbal form, despite their heterogeneity.

I. Dirhematic predicative $Cx_s[ØVP]$ like *altro furbacchione Costanzo* ('another sly one Costanzo') in (1) and *les steaks saignants? à point?* (lit. 'the steaks bloody? just right?')[2] in (2), in which two or three informational units are connected by a relationship that is presented as new by speakers (De Mauro & Thornton 1985).

> (1) It. VoLIP [MA15][3]
>
> > A: *anche lì // **altro furbacchione Costanzo** // altro furbacchione che appena vede uno che fa un pochino il vulcano lo acchiappa*
> > 'there too // *another sly one Costanzo* // another sly one that as soon as someone gets a little bit angry he catches him'

> (2) Fr. ESLO 2 [REPAS 1255]
>
> > A: *ils ont dû voir monsieur NPERS*
> > 'they had to see Mr NPERS'
> >
> > B: *bon **les steaks***
> > 'well *steaks*'
> >
> > A: *ils sont à Londres*
> > 'they are in London'
> >
> > B: ***saignants ? à point ?***
> > '*bloody? well done?*'
> >
> > A: *saignant*
> > 'bloody'
> >
> > B: *Killian tu veux le steak saignant ?*
> > 'Killian do you want bloody steaks?'

II. Non-dirhematic predicative $Cx_s[ØVP]$ such as the ones in (3) and (4) in which the linguistic sequence establishes a predicative relation with an entity not verbally expressed but that belongs to the extra-linguistic reality.

> (3) It. VoLIP [FA4]
>
> > A: *ecco # capito? che secondo me lei l'e' l'e' stronza // ma io son quando se voglio son piu' stronza di lei capito?*
> > 'so you know? like I think she's the bitch // but if I am more a bitch than her if I want I'm bitch you know?'
> >
> > B: ***che vipera***
> > '*what a viper*'

---

2. In Example (2) predication is distributed over several turns (Giordano & Voghera 2009).

3. For the transcription of examples the following symbols have been used: // to indicate prosodic boundaries as such as changes of intonational units; # to indicate pauses; <?> to indicate one incomprehensible word; *NPERS* for a name of person; [*pi:pronounce:instantaneous*] to indicate that there is overlapping of two or more turns.

(4) Fr. ESLO 1 [REPAS 272]

DM95FEM: *[pi:pronounce:instantaneous] y avait tout un // un article sur ce film*

'there was a whole an // an article about this movie'

DM95: *bien un article*

'even an article'

272LOC1: *non mais c'est très bien en couleurs // tu vois il y a des très jolies couleurs // très belles tenues // mais alors sinon c'est c'est atroce quoi c'est vraiment la décadence //* **horreur**

'no but it's very good in colors // you see there are very pretty beautiful colors // very well done // but then otherwise it's really atrocious it's really decadence // *horror*'

III. Phrases constituting circumstantial arguments of verbal clauses occurring in other turns (Giordano & Voghera 2009; Voghera et al. 2010), as *nei confronti della Lazio* ('towards Lazio') in (5) which is syntactically anchored to the verbal clause of the previous turn.

(5) It. CLIPS TV_it 03R

A: *questo è il fuori gioco di rientro*

'this is the return offside'

B: *nei confronti della Lazio*

'towards Lazio'

IV. Non-dirhematic and non-predicative $Cx_s[ØVP]$. This label includes very different constructions that often do not find a clear place in the grammatical descriptions. Within this group there are structures of different constitution (for example we distinguish constructions composed of one or more isolated phrase) and with different functions. On one hand, non-dirhematic and non-predicative $Cx_s[ØVP]$ includes discursive markers (i.e. *insomma* 'in short'), vocational NPs (see (6)), formulas, such as *merci* ('thank you') in (7), or greetings as such as *bonsoir* ('good evening') in (7).

(6) It. VoLIP [FA1]

A: *eh Paolo?*

'ehm Paolo?'

(7) Fr. ESLO1 [REPAS 272]

DM95: *allo? // oui merci madame // allo? // oui bonsoir monsieur*

'hello? // yes thanks madam // hello? // yes good evening sir'

On the other hand, in the group of non-dirhematic and non-predicative $Cx_s[ØVP]$ there are $Cx_s[ØVP]$ whose function is to add new information to the discourse. In the dialogical speech, frequently these $Cx_s[ØVP]$ add information blocks to the communicative exchange, like the complex NP *quattro chili seicentocinquanta*

grammi di figlia ('four kilos six hundred and fifty grams of daughter') in (8), through which the speaker informs the interlocutor about the weight of the newborn baby; and *la mauvaise habitude* ('the bad habit') in (9), through which the speaker comments the fact of leaving the keys in the lock.

(8) It. VoLIP [MA23]

    A: *non c'e' Marco?*
       'Isn't Marco here?'
    E: *e_ sara' sara' in ospedale // e' diventato papa' oggi*
       'and he will will be in the hospital // he became a father today'
    A: *e' diventato papa'?*
       'did he became a father?'
    E: *per la seconda volta # oggi alle dodici e mezzo e non lo so se_ ahah se e' a casa*
       'for the second time # today at half past twelve and I do not know if ahah if he is at home'
    A: *caspita*
       'good heavens'
    E: **quattro chili seicentocinquanta grammi di figlia**
       'four kilos six hundred and fifty grams of daughter'

(9) Fr. ESLO 2 [REPAS 1256]

    DS952MER: *c'est de pas laisser les clés dans la sur la porte*
                'it's not to leave the keys in the door'
    DS952: *bah oui // parce que ça sinon on pourra jamais rentrer*
          'well it is // because otherwise we will never be able to enter'
    DS952MER: **parce que l'habitude euh // la mauvaise habitude**
                'because the habit uh // the bad habit'
    DS952SOE: *on pourra pas rentrer*
           'we will not be able to enter'

In Table 1 the types of Cx$_s$[ØVP] are shown.

Table 1. Types of Cx$_s$[ØVP]

Types of Cx$_s$[ØVP]				
Dirhematic predicative	Non-Dirhematic predicative	Circumstantial NPs	Non-Dirhematic non-predicative	
			Chunks of information	DM, greetings, vocatives

For a long time, there has been discussion on the theoretical status of $Cx_s[\emptyset VP]$.[4] Traditionally only dirhematic and predicative $Cx_s[\emptyset VP]$ were considered by syntactic theories because they are closer to the verbal sentence model. As it is well known, the classical studies of the first half of the twentieth century were mainly concerned with non-verbal predication (Meillet 1906; Bally 1922; Hjelmslev 1948; Benveniste 1950). There were very few exceptions as Jespersen (1924) and more recently De Mauro (1974). However, $Cx_s[\emptyset VP]$ have received more attention only in the last thirty years thanks to the diffusion of corpus-based works that have get light on the frequency and on the diversity of these structures especially in speech (Voghera 1992; Biber et al. 1999; Cresti & Moneglia 2005; Voghera et al. 2010).[5]

As far as the theoretical status of non-dirhematic and non-predicative $Cx_s[\emptyset VP]$ is concerned, the different theories do not converge. On the one hand among works that refer to the model of verbal sentence, there are studies that consider the $Cx_s[\emptyset VP]$ as utterances (Cresti 1998; Cresti 2005; Blanche-Benveniste 2008) in which the subject and predicate relationships are not identifiable and consequently would be better described only through pragmatic criteria. Other works define non-dirhematic and non-predicative $Cx_s[\emptyset VP]$ as non-sentential units, that is syntactic structures different than the sentences (Barton 1990; Barton & Progovac 2005; Progovac 2006, 2013), or fragments (Merchant 2004, 2006), depending on whether they consider them autonomous or derived structures from verbal sentences. On the other hand, there are studies (De Mauro 1974; De Mauro & Thornton 1985; Voghera 1992) maintaining that the verb is not the decisive element to distinguish the different syntactic structures and include non-dirhematic and non-predicative $Cx_s[\emptyset VP]$ in the set of meaningful structures. In this analysis the latter approach has been adopted, according to which $Cx_s[\emptyset VP]$ can be considered syntactic autonomous structures, without having to assume an implied verb.

---

4.   For a review of definitions of $Cx_s[\emptyset VP]$ within the various theoretical frameworks of the history of linguistics see Graffi (2001).

5.   Obviously non-dirhematic and non-predicative $Cx_s[\emptyset VP]$ are not exclusive of spoken texts, as evidenced by studies conducted on literary and non-literary texts (Mortara Garavelli 1971; Lefeuvre 1999).

2. The corpus

The Meta-textual Cx$_s$[ØVP] analysed are taken from a corpus that consist of 1336 Cx$_s$[ØVP] with at least one noun (Cx$_s$NP[ØVP]).[6] They are distributed in spoken Italian and French as illustrated in Table 2.

Table 2. The number of Cx$_s$NP[ØVP] in spoken Italian and French

	Italian	French
Cx$_s$NP[ØVP] in dialogues	608	383
Cx$_s$NP[ØVP] in monologues	185	160
Total	**793**	**543**
Total	**1336**	

For this study face-to-face conversations, in which speakers take part to the conversation freely (it will be used the more generally term "dialogues" to refer to them) and one-way conversations, where communication takes place between a speaker and several listeners (it will be also used the term "monologues" to refer to them) have been analysed. Both for face-to-face conversations and for one-way conversations, it has been selected samples of speech as spontaneous as possible that could reproduce the language used in everyday life faithfully.

As far as spoken Italian, dialogues and monologues of the LIP corpus (De Mauro et al. 1993) have been analysed, in its audio version VoLIP (Voghera et al. 2014),[7] while for spoken French some subsections of corpus Eslo (*Enquête SocioLinguistique à Orléans*) (Serpollet et al. 2007)[8] have been used. In particular, in Eslo it has been considered the sections *Repas* and *Interview* for face-to-face speech, and *Conference* and *Media* for monologues. In Table 3, the hours of talks of the sections of corpora consulted are reported.

Table 3. The amount of hours of dialogues and monologues of spoken

	Italian	French	Total
Hours of speech in dialogues	11h:10′	12h:23′	23h:33′
Hours of speech in monologues	8h:36′	6h:37′	15h:12′
Total	**19h:46′**	**19h:00′**	**38h:46′**

---

6. The choice of analysing Cx$_s$NP[ØVP] is related to a larger study of Cx$_s$[ØVP] that focuses on how names can express grammatical relations in absence of verb (Sammarco 2017). The collection of Cx$_s$NP[ØVP] was mainly functional to this study but turned out to be very useful for aspects that do not properly concern the grammatical relations.

7. <http://www.parlaritaliano.it/index.php/it/visualizza-corpus>

8. <http://eslo.huma-num.fr/>

It must be said that Italian conversations present a broader diaphasic variation than French ones. In fact, Italian face to face conversations include very informal dialogues (i.e. between family members) and less informal dialogues (as such as dialogues between strangers in public places). Instead, those French ones present only one situational dimension, they are just conversations between friends in familiar places. Also, Italian monologues have been collected in different communicative situations and they represent more textual variation (university lectures, political meetings, lessons), while French monologues are exclusively taken from conferences. As we will see, when we talk about data, the major diaphasic variation in Italian dialogues is related to a higher density of some types of meta-textual $Cx_s[ØVP]$.

For detecting $Cx_sNP[ØVP]$ two basic criteria that belong to two different linguistic levels have been used: a) syntactic autonomy; b) prosody. The syntactic autonomy is the possibility of a construction to occur alone (Jespersen 1924; Bloomfield 1933; Barton 1990; Voghera 1992; Barton & Progovac 2005; Stainton 2006). Any $Cx[ØVP]$ is autonomous if it is not linked to other syntactic structures and it occurs in isolation in the speech text. Hence it has been considered as autonomous every $CxNP[ØVP]$ that is not part of the argumental structure of a verbal form occurring in the co-text. For example, the NPs *i soffitti* ('ceilings') and *deux steaks* ('two steaks'), in (10) and (11), are not arguments of the verbs *trattenere* ('to retain') and *mettre* ('to put') respectively, but they make up two new constructions.

(10)   It. VoLIP [MD 7]
A: *memorizzate questa immagine // non qua // un pennello normale // va a sporcare dappertutto // no // la nostra <?> trattiene la vernice // trattiene la tempera //* **e con un normale bastone //** **anche i soffitti //** *senza prendere scale // guardate*
'memorize this image // not here // a normal brush // goes to get dirty everywhere // no // our <?> holds the paint // retains the tempera // *and with a normal stick* // *also ceilings* // without taking stairs // look'

(11)   Fr. Eslo2 [REPAS 1262]
KR001:   *attends je vais mettre ma pizza au frais //* **pour demain //** **deux steaks** *// ça a rétréci?*
'wait I'll put my pizza in the fridge // *for tomorrow* // *two steaks* this has narrowed?'
PC261:   *je suis au régime*
'I am on a diet'
KR001:   *ah ouais ? // Nico au régime*
'are you? // Nico on a diet'

Therefore, the NPs that occur in question – answer pair structures have been not considered as autonomous $Cx_sNP[\emptyset VP]$ when they are arguments of the verbal form occurring in the co-text as *carta da acquerello* ('watercolour paper') in (12) and gapping structures as *moi aussi* ('me too') in (13).

(12) It. VoLIP [ND13]
      *perche' la carta ci ha la proprieta' di assorbire il sale // che cos'e' // questa carta salata? // **carta da acquerello** // quindi molto <paro> molto porosa bencollata // cioe' ricca di colla // con un bagno di cloruro di sodio // ne basta cinquanta grammi al litro*
      'because the paper has the property to absorb the salt // what is // this salt-paper? // *watercolour paper* // so very <paro> very porous well glued // which is rich of glue // with a sodium chloride bath // it suffices fifty grams per litre'

(13) Fr. ESLO2 [REPAS 1264]
RN488FRE:	*c'est vrai qu'il est mal foutu cet appartement-là en plus*
	'it is true that it is badly fucked this apartment there and more'
RN488 :	*bah oui mais attends // il fallait qu'elle fasse faire une cuisine*
	'well yes but wait // she had to make the kitchen'
RN488FRE:	*mais moi en tant en tant que propriétaire // non mais en moi en tant que propriétaire // j'aurais cassé le mur*
	'but me as an owner // no but in me as an owner // I would have broken the wall'
RN488:	*oui bah **moi aussi***
	'yes bah *me too*'

Afterwards, it has been considered as autonomous $Cx_sNP[\emptyset VP]$ when they did not contain arguments of a verbal form and they may form a turn alone in dialogic conversation (Voghera 1992). Even though there is no bi-univocal correlation between the fact of constituting a turn and that of being able to be syntactically autonomous, the sequence of turns is a constitutive principle of spoken texts and it is related to the syntactic structure (Sacks et al. 1974; Schegloff 1982). See $Cx_sNP[\emptyset VP]$ in bold in (14) and (15) that are structurally autonomous, because NPs *Marina* and *peperoni* are not arguments of a verbal form, moreover they constitute turn alone.

(14) It. VoLIP [NA1]
      E: *Marina peperoni?*
          'Marina peppers?'
      C: *no grazie*
          'no thanks'

(15)   Fr. Eslo2 [REPAS 1268]

LD860:	*ça fait moins cher hein*
	'it's cheaper huh'
LD860FRE1:	*à Disney // ils ont de l'argent hein*
	'at Disney // they have money huh'
LD860:	*bah si non tu en as pour quin- enfin en tant que si tu as la carte*
	euh tu sais jeunes là pour le train
	'well if you do not have it finally as if you have the card uh'
	'you know young there for the train'
LD860FRE1:	*moi je l'ai moi*
	'I have it'
LD860:	*tu en as pour euh quinze euros euh*
	'you have for uh fifteen euros uh'
LD860FRE1:	*en plus elle est encore bonne hein*
	'it is still good huh'
LD860SOE:	**aller-retour ?**
	'*round trip?*'
LD860FRE2:	**non pas quinze euros l'aller-retour**
	'*no fifteen euros round trip*'
LD860:	*l'aller*
	'the go'
LD860FRE2:	*l'aller-retour // tu as juste l'aller*
	'the round trip // you just have to go'
LD860:	**quinze euros l'aller**
	'*fifteen euros one way*'
LD860SOE:	*c'est ça va hein*
	'that's okay huh'

The second criteria used is the prosodic autonomy. In speech, prosody is crucial for the identification of syntactic and textual units (Voghera 1992, 2017). Hence, on the basis of a perceptual analysis, it has been considered as autonomous $Cx_sNP[ØVP]$ that correspond to one or more intonational units (Voghera 1992; Giordano & Voghera 2009). By intonational unit it is meant a portion of speech that is bounded by two intonational marks that may correspond to a pause, as in (16) and (17), to vowel lengthening or vocalization, as in (18) and (19), or to a tonal change. In the dialogical context when $Cx_sNP[ØVP]$ occur in isolation more frequently, they can correspond to one intonation unit. In monologic text, instead, $Cx_sNP[ØVP]$ could correspond to one or more intonational unit, due to the larger extension of the turns. In these cases, there is a prosodic cohesion among the parts that can form a one unique $Cx_sNP[ØVP]$ (Giordano & Voghera 2009).

(16)   It. VoLIP [FA2]

B: *mamma // quand'ero piccino // come me le lavavo le mani?*
'mum // when I was a child // how did I wash my hands?'

C: *quand'eri piccino // se vuoi te le faccio vedere? //*
'when you were a child // if you want I'll show you?'

B: *col sapone <?>*
'with soap <?>'

C: **davanti al lavandino // davanti a me // le mani sotto** # *cosi' si fa coi bambini*
'*in front of the sink // in front of me // the hands under //* so we do with the children'

(17)   Fr. ESLO1 [REPAS 273]

FA192AMI2:   *elle est très pratique // parce qu'elle se met dans un coin puis quand nous voulons // euh parce que // notre cuisine est si petite //* **nous deux mon mari et moi** *// nous nous comme nous sommes seuls // souvent // on dîne à dans la cuisine*
'it's very handy // because it gets in a corner and then when we want // uh because // our kitchen is so small // *the two of us my husband and I //* we as we are alone // often // we dine in the kitchen'

(18)   It. VoLIP [MA1]

A: *eh beh descrivi le cose significative // per cui per esempio // non so noi se io dovessi descrivere voi come sceneggiatura direi //* **eh salotto_ di casa eh // a Milano_ di appartamento a Milano eh // presenti sei persone**
'eh well you describe the meaningful things // so for example // I do not know if I had to describe you as a screenplay I would say // eh *living room_ home eh // in Milan_ of apartment in Milan eh // six people present*'

B: *ma e' assurdo <??> fatto bene comunque*
'but it is absurd <??> well done anyway'

(19)   Fr. ESLO2 [CONF 1244]

CBruneau: *et euh sans plus attendre donc je vais laisser la place à nos intervenants //* **alors euh donc euh peut-être une petite précision euh enfin** *// euh avant // euh hm // l'après-midi va va durer euh un certain temps hein* […]
'and uh without waiting further so I'll leave for our speakers // *so uh so uh maybe a little clarification uh finally //* uh before // uh hm // the afternoon will last uh some time huh […]'

Syntactic autonomy and prosodic autonomy are closely related to each other. Although there is no one-to-one relationship between the units of these two levels,

in fact there may be two or more units syntactically bounded but intonationally autonomous, syntactic and prosodic bounds tend to coincide, mainly in smaller syntactic units[9] (Voghera 1992).

3. Meta-textual Cx$_s$NP[ØVP]

As already mentioned in § 1, this paper focuses on a group of non-dirhematic and non-predicative Cx$_s$[ØVP] that are defined as meta-textual Cx$_s$NP[ØVP]. They are composed of at least one noun, and their function is to give a name not to an extra-linguistic entity but to a part of the spoken text, such as the structures in bold in the following examples.

(20) It. VoLIP [FA12]
    A: *va bene via senta **eh e il discorso** # **dei noleggi** come stanno andando ora quei noleggi?*
    'well listen eh and *the rental matter* # how are they working these rentals?'

(21) Fr. ESLO 2 [CONF 1074]
    A: *alors **donc peut-être une petite précision** // **euh enfin** avant l'après-midi va durer // euh un certain temps*
    'Well so perhaps *just a little clarification* // ehm finally before the afternoon will take some time'

(22) It. VoLIP [FA12]
    A: *ahah **ultima cosa** e poi la mando via # il telefono delle informazioni?*
    'ahah *one last thing* and then I let you go # the information phone number?'.

(23) Fr. ESLO 2 [REPAS 1270]
    TG634:   ***autre chose** // tu as fait ta ba- ta physique?*
                 '*another thing* // have you finished your your physics?
    TG634FIL2: *oui // comme tu l'as vu*
                  'yeah as you can see'

The Cx$_s$NP[ØVP] in (20), (21), (22) and (23) have three main functions.[10] The first function is to define the argument of the discourse by naming it or categorizing it into a textual macro-category more or less vague. They give a title to what is going to be said, they express a general idea of the theme about which something

---

9. Syntactic and prosodic bounds tend to coincide in syntactic units smaller than a sentence, such as clauses (Voghera 1992: 121; Giordano & Voghera 2009).

10. For a more detailed classification from a functional point of view, see Sammarco (2020).

more specific and more particular is going to be affirmed (Scarano 2003). In the Example (20) with *il discorso dei noleggi* ('the matter of rental'), the speaker is naming the topic that they want to talk about. In (21) the speaker gives a meta-discursive definition of the information ('just a little clarification') that they are going to introduce.

The second function is to maintain the textual continuity in dialogicity (Du Bois 2014). Meta-textual $Cx_sNP[ØVP]$ concern the process of linking utterances that speakers are going to produce and they are often considered as structures with rhetorical artifact that establish anaphoric or cataphoric relations (Scarano 2003). Constructions of this type also anticipate in the discourse a topic that the speakers probably already share. In the Example (21) the speaker affirms that what is following concerns the matter of rentals as a familiar topic. In the Examples (21) – (23) even if the $Cx_sNP[ØVP]$ do not really anticipate the content, they create a link with what it follows. In (21) the speaker introduces a new topic without saying what he is going to deal with, in the Examples (22) and (23) moreover through $Cx_sNP[ØVP]$ *ultima cosa* ('one last thing') and *autre chose* ('another thing') the speaker organizes the discourse that follows giving it a place within the spoken text.

The third function is to interact with the listener. Meta-textual $Cx_sNP[ØVP]$ can be considered as expressions of discursive dynamic in which participants are involved to build the discourse in real time (see Auer 2009; Du Bois 2014). These constructions provide an infrastructure that helps interlocutor's understanding (Depperman 2015). They are strategies useful to make easier the communicative exchange between speaker and hearer in face-to-face situations. In this sense they are phenomena that have a function of discursive support (Voghera 2017). Through meta-textual $Cx_sNP[ØVP]$, the speaker is warning the listener that they are going to mention they want to stand out as a well-defined piece of information so the listener gets ready to receive the chunk of information that that the speaker is going to tell them. In spoken text this effect is given also thanks to a very defined prosody. These structures in fact constitute a well-defined prosodic unit that contributes to localize the scope of the focus. In (20) the speaker is warning to the listener that he wants to deal with a specific topic. In Examples (21) – (23) he is warning to the listener about the fact that he is adding something new. In fact, these structures often co-occur with formulaic $Cx_sNP[ØVP]$ that the speaker uses to catch the listener's attention (see (24)).

(24)   It. VoLIP [MD2]
   allora // **attenzione** # [**promemoria**] // **piccolo salto indietro** e // eh andiamo <?> quello che v'ho detto adesso #
   'well // *attention* # *[reminder]* // *little jump back* and // eh let's go <?> what I told you now #'

In (24) the speaker is explaining a lesson. She calls the attention of the audience through the formulaic expression *attention* and she warns the listeners that what she will say later is an important topic and at the same time that, from the textual point of view, it represents a digression from the discourse she was carrying out.

$Cx_sNP[ØVP]$ such as the ones in the Examples (20) – (24) are considered typical phenomena of spoken texts (Voghera 1992, 2017; Bazzanella 1994, 2005; Auer 2009; Du Bois 2014; Voghera 2017). They have textual and syntactic features that belong to the most frequent phenomena whose frequency correlates with spoken modality (Voghera 2017) and their theoretical status is halfway between that the one of the discursive markers and the one of information structuring phenomena (Ferrari 2003; Voghera 2017).

Like discourse markers, meta-textual $Cx_sNP[ØVP]$ are semantically redundant elements, since they add nothing to the content of the discourse. They can be compared to structures that Bazzanella (1994, 2005) defines as meta-textual discourse markers, that is elements or phenomena that usually occur to the left of the structure and that marks the text and organize it from the point of view of the argumentative structure (Bazzanella 1994: 160). According to Bazzanella, meta-textual discourse markers represent around 10% of the discourse markers and perform different functions such as highlighting the matter's referent or object, indicating reformulations of the text through paraphrases, corrections, exemplifications or digressions, or simply pointing a changing in speech structures.

From a syntactic point of view, meta-textual $Cx_sNP[ØVP]$ are discontinuous structures because they are syntactically separated from the rest of the sentence with a suspended ascending intonation. They are very functional in spoken texts because of their brevity, like all verbless constructions they adapt well to the dialogic shifts (Voghera 2017). Traditionally, because of their nominal constituency, they are defined as anacoluthons (Cresti 2003, 2016; Scarano 2003) or hanging topics (Benincà 1988: 131).

Finally, from a very strict pragmatic point of view meta-textual $Cx_sNP[ØVP]$ are defined as syntactic islands that have the function of anchoring the comment of the utterance (Cresti 2003). In works on task-oriented dialogues, meta-textual $Cx_sNP[ØVP]$ are included within the pragmatic moves labelled as Info-request (Savy & Alfano 2016), that are a type of construction usually made up of isolated NPs, used by the speakers use to require an informative contribution, eliciting a Discourse Topic (Savy & Alfano 2016). Speakers use these structures just as a request to their interlocutor to talk about a specific topic. In the schemas of pragmatic annotation of Map-tasks, these structures are defined as beginnings of transaction (Sinclair & Coulthard 1975); in other words, they represent utterances or pragmatic moves that speakers use to open a transaction, a term used in task oriented dialogues to define a type of information's macrostructure that participants make as a first step for get the task (De Leo 2008).

The underlying hypothesis of this work is that meta-textual $Cx_sNP[ØVP]$ can be considered as strategies that the participants of the conversation use with the main purpose to guide the interaction. Moreover, at the same time, through this strategy, speakers create new discursive, thematic and semantic categories, as it usually happens in spontaneous communicative interaction. These structures can be considered as mechanism of categorization at textual level in the same way as constructions that have the function of building sematic categories. They can be analysed as strategies that speakers use to find solutions gradually to the communication needs that emerge during the communicative interaction, like special plurals, derivational strategies, reduplication, non-exhaustive connectives, general extenders (Mauri 2017).

3.1   Types of meta-textual CxsNP[ØVP]

A very general classification of meta-textual $Cx_sNP[ØVP]$ is proposed, based on their meaning and functions. Meta-textual $Cx_sNP[ØVP]$ could be divided into three sets. The first group consists of meta-textual $Cx_sNP[ØVP]$ that have the function to give a title to the topic of the discourse. $Cx_sNP[ØVP]$ with this function are usually noun phrases that express the topic that speakers want to deal with, as structures in bold in the Example (25) – (27).

(25)   It. VoLIP [NA13]
    C: *allora //* **tutta la questione del //** *eh* **la moda il prestigio linguistico** *# no? lei si e' fatta un'idea del prestigio linguistico dal capitolo precedente*
        'so // *the matter about fashion linguistic prestige* # no? you have got any idea about linguistic prestige from the previous chapter'

(26)   It. VoLIP [FA10]
    A: *mamma mia*
        'my goodness'
    B: **contratti d'affitto //** *allora lei s'era cautela<ta> // io qui c'ho tutta una cosa che ti te la faccio solo vedere solo da lontano #*
        '*contracts of rent* // then she was precautionary // I've got a whole thing here that I just let you see only from a distance'

(27)   Fr. ESLO [CONF 1244]
    FLegouy: *bien entendu quelques définitions* **le vin dans la mondialisation** *// euh je vais pas tout reprendre je n'ai absolument pas le temps // euh on vous définit ce que représente le vin*
        'well some definitions *the wine in the globalization* // I do not want to repeat everything I have absolutely no time // I will give you the definition of what the wine represents'

In the Examples (25) – (27), the $Cx_sNP[ØVP]$ consist of the complex noun phrases *tutta la questione della moda il prestigio linguistico* ('the matter about trend linguistic

prestige'), *contratti d'affitto* ('contracts of rent'), *le vin dans la mondialisation* ('the wine in the globalization') that have the merely function of naming the topic that speaker wants to talk about. These structures represent the key word, and they could be repeated in the discourse that follows, as it happens in the Example (25), where the speaker makes an answer about the linguistic prestige that represents the topic of the following conversation. They can be followed by a paraphrase, as in the example of spoken French (27), where during a conference the speaker introduces the new step of the discourse: he gives some definitions of wine and then he explains what he really wants to say. Finally, as they clearly separate the discourse that follows from what it has just been said, these constructions could merely open a new step of the discourse as in the Example (26) with *contratti d'affitto*, where the speaker specifies that she starts a new topic that have been planned before and she does not repeat the words of the construction but resumes the story that concerns the question of contracts of rent. She shows some documents to the interlocutor that are useful to understand what she is going to say about the matter.

The second group includes Cx$_s$NP[ØVP] that consist of more general nouns such as the Italian '*cosa*' ('thing') or the French *type de problème* ('type of problem') that may co-occur with numerals or indefinites, as in the Examples (28), (29).

(28)   It. VoLIP [NA3]

   B: *no che me ne frega del lavoro // **seconda cosa** // eh c'e' un collega che avrebbe bisogno di mettersi in contatto con gli amici dell'<?> [DI MUSEI]*

   B: 'no I don't care about the work // *second thing* // eh there is a colleague who needs to get in contact with the Museums Friends'

(29)   Fr. ESLO [CONF 1242]

   ch_GB9:   *on travaille sur un élément qui est pour nous essentiel qui est la parole // euh le discours tel qu'on peut l'entendre // et vous le savez son énorme problème // c'est que c'est bien notre matériau // mais ce matériau-là a un inconvénient terrible // c'est que dès qu'on a fini de parler // on a tout oublié // nous vérifions chaque année au moment des partiels avec nos étudiants // donc **premier type de problème** // on a bien un matériau qui est facilement accessible // mais euh qui en quantité à peu près illimité*

   'we work on an element that is for us essential which is the speech // uh the speech as we can hear it // and you know its huge problem // it's our material // but this material has a terrible disadvantage // it is that as soon as we have finished speaking we have forgotten everything // we verify every year when our students take examinations // so *first type of problem* // we have enough material that is easy to reach // but that is in unlimited quantities'

Meta-textual Cx$_s$NP[ØVP] with general nouns such as these ones in the Examples (28) – (29) have the same function to encapsulate the text (Voghera 2017: 169),

dividing it in thematic chunks and preparing the listener for the discourse that follows. As the structures of the first group, meta-textual $Cx_sNP[ØVP]$ of this group informs the listener that the speaker is going to talk about a new topic and that he is closing what he has already said. Anyway, through this second type of meta-textual $Cx_sNP[ØVP]$ the speaker does not mention the topic, he only informs the listener that he is going to introduce a new chunk of discourse. $Cx_sNP[ØVP]$ of the second group make their meta-textual feature more clear. In fact, these structures organize the parts of the text and their meaning may not interfere with the content of the text. On the one hand they can link two parts of the discourse that may be not semantically related and introduce a discourse that is completely detached from the previous talk, as in the Example (28). On the other hand, they can connect two parts of text that are semantically related, as it happens in the Example (29), where the text that follows and the one that precedes the $CxNP[ØVP]$ are in general-specific semantic relation. The second part, in fact, specifies what have been said before.

Finally, the third group consist of meta-textual $Cx_sNP[ØVP]$ composed of nouns that belong to different semantic areas that make clear the type of texts that speakers are going to introduce. For example, the $CxNP[ØVP]$ *promemoria* ('reminder') in (24) or *definition* ('definition') in (27). Meta-textual $Cx_sNP[ØVP]$ of this group define the textual category which the discourse that they introduce belongs to. In other words, with this kind of meta-textual $Cx_sNP[ØVP]$, speakers organize the discourse and define it naming the discursive category. In the Example (24) the teacher is telling that what they are going to say is something that the pupils have to remind, so they specify that it is a memorandum; in the Example (27) the speaker is saying that the following text belongs to the textual category of definition.

Three semantic subclasses of nouns that can form meta-textual $Cx_sNP[ØVP]$ of this group have been identified. The first subclass is represented by meta-textual $Cx_sNP[ØVP]$ composed of nouns that belong to the semantic area of 'saying'. As the terms *domanda* and *question* ('question'), in (30) and (31), *comunicazione di servizio* ('public service announcement') in (32) and *messaggio della segreteria*, in (33).

(30) It. VoLIP [ND12]

A: *allora // si tratta adesso di siste<mare> di sistemare di risistemare tutte queste informazioni // sotto forma // di eh elaborazione scritta // per scritto // **allora prima domanda** // che forma vogliamo dare a tutto questo materiale? come lo vogliamo organizzare?*

'So // now we have to organize and reorganize all these data // writing them down // writing // *well first question* // what order will we give to all this material? // how do we want to organize it?'

(31)   Fr. ESLO2 [CONF 1241]

JGuarrigues: ***deuxième question*** // *comment* // *je dirais* // *ce contexte de crise se se transmue en recours* // *qu'est-ce qu'il se passe? comment est-ce qu'on fabrique l'homme providentiel ?*

'*second question* // how // I should say // this context of crisis changes in support of // what does it happen? how do we get to the providential man?'

(32)   It. VoLIP [MA13]

A: ***<comunicazione> di servizio*** *fare attenzione al binario tredici*

'*public service announcement* pay attention to platform thirteen'.

(33)   It. VoLIP [RA1]

A: *salgo* // *faccio pipi'* ***messaggio della segreteria*** // *ah sai non mi andava di prendere il treno* // *ho preso una macchina in affitto*

'I go upstairs // I go pee *message on the answering machine* // well you know I did not want to take the train // I rented a car'

Cx$_s$NP[ØVP] in (30), (31) and (32) divide the text into sections but at the same time they give a sort of labelling that explains the type of text and linguistic act that is introduced: an answer in (30) and (31) an important notice that is addressed to staff, in (32). The latter is a type of statements that is usually heard in public places, in this case a train station. Referring to Schegloff (2007), it is assumed that these structures are pre-sequences, that is preliminary sequences in the conversational action that precede linguistic acts, such as requests or suggestions or assertions. These elements do not correspond to real anticipations of the linguistic act but to the recognition of the object to which the following discourse refers (Schegloff 2007: 45–46). In (33) the meta-textual Cx$_s$NP[ØVP] *messaggio della segreteria* ('message on the answering machine') label the discourse that follows with the name of its discursive category. Furthermore, *messaggio della segreteria* has the function of introducing a reported speech, namely, it reproduces a text of a different speaker and it frames the content of the cited speech (Ferrari 2014: 233; Calaresu 2004). The speaker A is going to tell what the message said and she announces that what she is going to tell is the text of the message itself. The second subclass is represented by structures composed of nouns of text's portion. These structures consist of meta-textual nouns such as *piccola parentesi* or (34) or *act un* in the French Example (35).

(34)   It. VoLIP [ND13]

A: *questa e' una macchina fotografica* // *in effetti un piccolo* // *e anzi* ***una piccola parentesi*** // *gli americani durante l'ultima guerra # agli americani fotografi venivano insegnate che cosa?*

A: 'this is a camera // a very little actually // and rather *a very little parenthesis* // American people // during the last war # what did they teach to the American photographers?'

(35)   Fr. ESLO 2 [REPAS 1247]
       RN166: *ça // enregistre là?*
               'it is recording?'
       DR381: *nous mangeons*
               'we are eating'
       BV647: *oui*
               'yeah'
       DR381: **acte un** *// je fais les didascalies tu sais*
               '*act one* // I do the subtitles you know'

In (34) the meta-textual Cx$_s$NP[ØVP] specifies that the text introduced by it represents a discontinuity to the previous discourse. Through *piccola parentesi* ('little parenthesis') the speaker explains that what is going to say is a parenthetical element and he tells the listeners that the content is isolated from the rest of the text. In the Example (35) the NP *act un* ('act one') is used by the speaker to announce that what follows is included in the *act un*. Finally, the third subclass of meta-textual Cx$_s$NP[ØVP] with less general nouns include NP$_s$ composed of metaphors, such as *morale della favola* in (36) or *cerise sur le gateau* in (37).

(36)   It. VoLIP [RA1]
       B: *manco e' arrivato* [*quest'uomo*] *subito rompe le palle // insomma //* **morale della favola**
            'He has just arrived this man and he has already annoyed us // *moral of the story*'
       A: *morale della favola eh #*
            '*moral of the story*'.

(37)   Fr. ESLO2 [CONF 1241]
       BGratuze: *la preuve par la composition de ces verres // que // Bernard Perrot travaillait du verre au plomb avec des compositions // dans la gamme // exactement dans la gamme de composition qu'on avait trouvée pour les // l'ensemble des pièces qui lui sont attribuées // voilà //* **et cerise sur la gateau** *// on a eu la chance de travailler sur différents verres rouges alors le verre rouge c'est un verre très classique hein euh chez Perrot*
                   'The test through the composition of these glasses // that Bernard Perrot worked lead glass in some compositions // in the range // exactly in the range of composition that we have found in the set of items that are associated to him // here is *and cherry on the top* // they have the opportunity of working on different red glasses so red glass is a very classic glass'

In the Example (36) when the speaker says *morale della favola* ('moral of the story') she tells us that, considering the previous discourse, she has concluded that this man is very annoying. The meta-textual $Cx_sNP[ØVP]$ gives a textual definition of the text that follows through a discursive metaphor denoting that it is going to be introduced the conclusions of the discourse. In the Example (37), the meta-textual $Cx_sNP[ØVP]$ *cerise sur le gateau* ('cherry on the top') is used by the speaker during a conference to explains all the discoveries about the way of working glasses and it introduces a notion that could be considered the final point of the discourse. There could be also spatial metaphors that show the clear point to which the speaker wants to lead the addressee, as *piccolo salto indietro* in (24) that it will be repeated in (38).

(38)   It. VoLIP [MD2]
allora // attenzione # [promemoria] // **piccolo salto indietro** e // eh andiamo <?> quello che v'ho detto adesso #
'well // attention # [reminder] // *little jump back* // eh let's go <?> what I told you now #'.

Besides giving a name to the parts of the discourse, meta-textual $Cx_sNP[ØVP]$ give an order to the spontaneous discourse. They can be considered as strategies that speakers use to organize the discourse and, in some way, they can be part of the semantic phenomena of a broader range attested in spontaneous speech, that is the building of semantic *ad-hoc* categories (Mauri 2017). Although traditionally the study of *ad-hoc* categories has been at the centre of the interests of cognitive and psychological studies, it proved to be of great linguistic interest. In fact, in spontaneous speech speakers use linguistic tools to create semantic *ad-hoc* categories to achieve their communicative goals. As corpus driven studies show (Mauri & Sansò 2018), these categories are context-dependent, that is they are presented as a set of entities whose membership is established on-line, in the interaction at the same time speakers communicate in the speech flow. Even if the creation of a category does not necessarily take place through a clear label, speakers are able to interpret the relevant properties of the elements that compose them through inferential processes, encyclopaedic knowledge and information from the extra-linguistic context.

As other linguistic strategies, meta-textual $Cx_sNP[ØVP]$ build categories of the text itself. Meta-textual $Cx_sNP[ØVP]$ have a prospective function to create and build meta-textual anchorage for the discourse that is going to be produced. They represent a real container that speaker creates to insert the text, as the PP *dans le premier point* ('in the first point') in the Example (39), used by the speakers as a textual box.

(39)   Fr. ESLO1 [CONF 503]

503CONF: *euh euh se préoccupant de l'évolution psychologique de l'individu // au cours du temps // **dans le premier point** donc les conditions générales de perception de du message parlé ou écrit // donc les conditions de psychologie générale // nous euh allons*
'caring about the psychological evolution of the individual // over time // *in the first point* therefore the general conditions of perception of the spoken or written message // so the conditions of general psychology // we uh let's'

In (39) the NP *le première point* ('the first point') is used by the speaker as a big set that collect the parts of the discourse that concern the same topic ('the general conditions of perception of the spoken or written message and the conditions of general psychology').

Through meta-textual $Cx_sNP[ØVP]$, speakers build thematic and meta-textual categories, starting from a meta-textual label that could be semantically vague or specific. Through $Cx_sNP[ØVP]$ that belong to the first group, speakers give a name to the topic creating a new step of the discourse that is going to be built on line. Like mapping structures, they reproduce a relationship between utterances as parallelism does (Du Bois 2014). $Cx_sNP[ØVP]$ of the second and third groups introduce the topic as well, but they do not properly name the discourse but the textual categories to which the discourse belong to or could belong to. More specifically, $Cx_sNP[ØVP]$ belonging to the second group are used to build generic meta-textual categories that the speaker does not know or simply consider not important to know in this situation, and he uses to simply monitor what he wants to say. $Cx_sNP[ØVP]$ of the third group clearly define the textual category and speakers use them because they consider important to underline the type of text they are going to produce to achieve properly the communication goals.

In order to have a clearer picture of what have been said so far, Figure 1 summarises the different types of $Cx_s[ØVP]$ and meta-textual $Cx_sNP[ØVP]$.
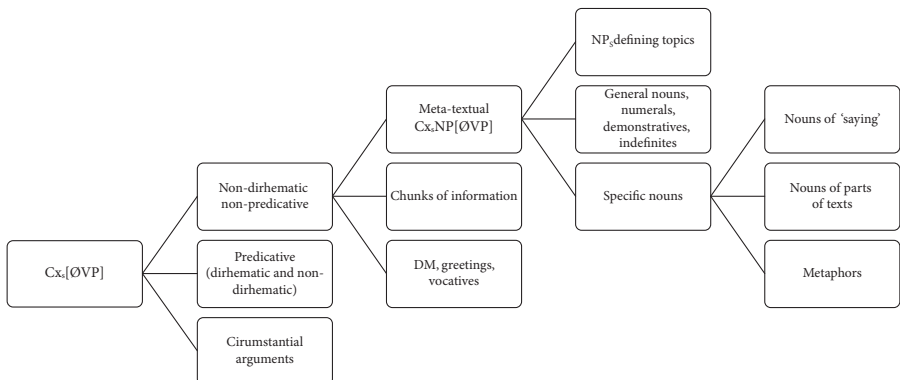
Figure 1. Types of Cxs[ØVP] and meta-textual CxsNP[ØVP]

4.   Some quantitative data

The corpus on which the analysis was made consists of 102 meta-textual $Cx_sNP[ØVP]$. They represent a small percentage (7.6%) of the total $Cx_sNP[ØVP]$. From a quantitative point of view, the frequency of these constructions is almost the same in Italian and French speech. However, it is interesting to note their different percentage in dialogic and monologic speech (see. Table 4).

Table 4. The percentage of meta-textual $Cx_sNP[ØVP]$ on the total amount of $Cx_sNP[ØVP]$ in Italian and French dialogues and monologues

	Italian	French
$Cx_sNP[ØVP]$ in dialogues	4.7%	2%
$Cx_sNP[ØVP]$ in monologues	14%	24%
Total	**6.9%**	**8.6%**

Predictably, the percentage of meta-textual $Cx_sNP[ØVP]$ is greater in monologic speech. Since the function of these structures is to organize the discourse, it is clear that they are used to a greater extent in a longer type of speech that requires a greater effort for the interlocutors to maintain the thread of the discourse.

The difference between dialogues and monologues is higher in French than in Italian. A closer look confirms that the constructions of this type are distributed in the dialogic and monologic speech in a different way for the two languages. In Table 5 the data concerning the percentages on the only meta-textual $Cx_sNP[ØVP]$ show how they are distributed among the dialogic and monologic texts in Italian and French.

Table 5. The percentage of the distribution of CxsNP[ØVP] with the textual naming function in spoken Italian and French dialogues and monologues

	Italian	French	
$Cx_sNP[ØVP]$ in dialogues	29 (28.5%)	8 (7.8%)	
$Cx_sNP[ØVP]$ in monologues	26 (25.5%)	39 (38.2%)	
Total	**55 (54%)**	**47 (46%)**	**102 (100%)**

The data in Table 5 show a substantial difference between the two languages in the distribution of the meta-textual $Cx_sNP[ØVP]$ in dialogues and monologues. Contrary to what one would expect, more than half of the Italian meta-textual $Cx_sNP[ØVP]$ occur in the dialogues (29), while in French only about one sixth of the meta-textual $Cx_sNP[ØVP]$ occur in dialogues (8). This difference may be due to the greater variation in Italian dialogues. As it has already been said, while French dialogues consist of only conversations between family members, Italian dialogues also include face-to-face conversations between people who do not know each

other or who are in work environments. In fact, giving a closer look at the Italian dialogues, we see that the distribution of meta-textual Cx$_s$NP[ØVP] is not uniform. They do not occur in informal dialogues among family members, like those that have been collected for French. They are all concentrated in the dialogues between co-workers (meetings, conversations between employer-employee). The higher frequency of the meta-textual Cx$_s$NP[ØVP] in these types of conversations is not related to a greater formality, however, but to a different textuality and informative density, in other words it correlates with the type of structure of the spoken text and the amount of information that have to be said. In fact, the number of meta-textual Cx$_s$NP[ØVP] is equally low in formal dialogues between strangers (for example at public counters). While they are concentrated in texts that present longer turns and higher amount of information that need a more planning control.

To get a complete picture on the distribution of the meta-textual Cx$_s$NP[ØVP] we need to look at how the three types of meta-textual Cx$_s$NP[ØVP] are distributed: NPs defining Topics, Cx$_s$NP[ØVP] composed of general nouns, and Cx$_s$NP[ØVP] composed of more specific meta-textual nouns (see Tables 6 and 7).

Table 6. The distribution of three types of meta-textual CxsNP[ØVP] in the whole corpus

	Corpus
NP$_S$ defining Topics	31 (30.3%)
General nouns, numerals, demonstratives, indefinites	34 (33.3%)
Specific nouns	37 (36.2%)

Table 6 shows that the three types of meta-textual Cx$_s$NP[ØVP] are equally frequent. Regarding the three subgroups of the meta-textual Cx$_s$NP[ØVP]composed of specific nouns they are distributed as shown in Table 7.

Table 7. The distribution of three types of meta-textual CxsNP[ØVP] composed of specific nouns

	Corpus
Nouns of 'saying'	**23 (62.2%)**
Nouns of parts of texts	5 (13.5%)
Metaphors	9 (24.3%)

As it can be seen from Table 7, most of the meta-textual Cx$_s$NP[ØVP] composed of specific nouns have NPs that have a name of saying as lexical head (i.e. *question; information*). Instead, about a quarter is represented by meta-textual Cx$_s$NP[ØVP] made of NPs whose lexical head is a metaphor relative to the semantic area of

the argumentation (i.e. *cheer on the top; little jump back*). The lesser part of the meta-textual Cx$_s$NP[ØVP] with specific nouns is composed of nouns referring to parts of text.

Now we focus on the distribution of the meta-textual Cx$_s$NP[ØVP] types in the dialogues and monologues of Italian and French (see Table 8).

Table 8. The distribution of the meta-textual Cx$_s$NP[ØVP] types in the dialogues and monologues of Italian and French

	Dialogues		Monologues	
	Italian	**French**	**Italian**	**French**
NP$_S$ defining Topics	7 (24.2%)	3 (37.5%)	**10 (38.4%)**	11 (28.2%)
General nouns, numerals, demonstratives, indefinites	**16 (55.2%)**	2 (25%)	8 (30.8%)	8 (20.5%)
Specific nouns	6 (20.6%)	3 (37.5%)	8 (30.8%)	**20 (51.2%)**
Total	**29 (100%)**	**8 (100%)**	**26 (100%)**	**39 (100%)**

Table 8 allows us to make qualitative considerations on the distribution of meta-textual Cx$_s$NP[ØVP] in dialogues and monologues. The first point seems worthy of attention is that most of the meta-textual Cx$_s$NP[ØVP] that occur in the dialogues is composed of vague names. In face-to-face conversations, 13 out of 16 Cx$_s$NP[ØVP] made up of general names have a NP whose head is the noun *cosa*, while in monologic texts only occur once. Moreover, the Cx$_s$NP[ØVP] of this type that occurr in dialogue have lighter NP, (i.e. phrases only made of lexical head, or whose head is a pronoun) (Voghera et al. 2004) than those that occur in monologues. The second observation is that in both languages most of meta-textual Cx$_s$NP[ØVP] that occur in the monologues have less vague nouns. In Italian they are represented by the meta-textual Cx$_s$NP[ØVP] which define the Topic, in French instead they belong to the group of the meta-textual Cx$_s$NP[ØVP] composed of more specific nouns which define the textual category of the speech (nouns of saying, nouns of editing and discursive metaphors).

The different lexical density and the heaviness of the phrase between meta-textual Cx$_s$NP[ØVP] of dialogues and meta-textual Cx$_s$NP[ØVP] of monologues is related to the different structure of the dialogic and monologic text, in particular to the length and alternation of the shifts, which in turn are related to the planning of the text. The meta-textual Cx$_s$NP[ØVP] made of general nouns and lighter phrases are more frequent in dialogues, because they are more suitable for shorter turns and for the less amount of planning time, which is in line with the peculiar characteristics of spoken dialogic texts (Voghera 2017).

Finally, also within the group of meta-textual Cx$_s$NP[ØVP] with specific nouns, it is possible to note that the structures with more phrasal and semantic complexity, such as discursive metaphors, mostly occur in monologues in both languages. The data concerning the distribution of the three subgroups of the meta-textual Cx$_s$NP[ØVP] with specific nouns in the two types of texts are reported in Table 9.

Table 9. The distribution of the three sub-groups of meta-textual CxsNP[ØVP] with specific nouns in the dialogues and monologues of Italian and French

	Dialogues		Monologues	
	Italian	French	Italian	French
Nouns of 'saying'	5 (71.4%)	3 (100%)	2 (25%)	13 (65%)
Nouns of parts of text	1 (14.3%)	0	2 (25%)	2 (10%)
Metaphors	1 (14.3%)	0	4 (50%)	5 (25%)
Total	**7 (100%)**	**3 (100%)**	**8 (100%)**	**20 (100%)**

As it can be seen from Table 9, most of the meta-textual Cx$_s$NP[ØVP] with specific nouns are composed of nouns that concern the semantic area of saying, and they are more in French monologues. The Cx$_s$NP[ØVP] whose lexical head is a noun of a part of text or a metaphor are mainly used only in monologues.

5. Conclusions

Meta-textual Cx$_s$NP[ØVP] represent a class of Cx$_s$NP[ØVP] that could be considered tools for textual naming. These constructions are among the phenomena of speech whose frequency is closely correlated with the syntactic and textual structure of spontaneous speech. The correlation of these structures with spontaneous speech concerns two aspects that are arranged on different descriptive levels: the interaction and textual structure. Regarding the interactive aspect, in the use of meta-textual Cx$_s$NP[ØVP] it is possible to find the process of building semantic categories in the interaction (Mauri 2017). Through these structures, speakers can start from the topic to map the discourse and create a sort of *ad-hoc* thematic and meta-textual categories that they can leave undefined, above all in dialogic texts. As it happens in the building process of *ad-hoc* semantic categories, speakers build meta-textual categories following two directions. On one hand, they can proceed from a bottom-up path: they build a lexical category through a part of the text itself that they use to define the discourse. This is the case, for example, of the meta-textual Cx$_s$NP [ØVP] that define the topic, which are made of NPs that are part of the topic itself. On the other hand, speakers can follow a top-down direction,

namely the meta-textual $Cx_sNP[ØVP]$ defines the textual category that belongs to what they will say. In this case, the category that defines the text can be more or less specific according to the lexical semantics of the head of NP. Through meta-textual $Cx_sNP[ØVP]$ composed of general nouns, speakers leave the category vague, while through more specific nouns they give well-defined meta-textual categories, giving new projections and expectation to the listeners and helping them to understand.

The second aspect concerns the correlation between types of meta-textual $Cx_sNP[ØVP]$ and dialogic and monological spoken texts. Data confirm the more general trend that, in dialogic texts, very light NPs composed of general nouns are more frequent, because more functional and more flexible at the temporal conditions that influence the greater frequency of turn-taking characterizing face to face conversations. On the contrary, in monologic texts, which are more planned, specific meta-textual nouns are more frequent. Moreover, from our data it emerges that the variety of these structures is not strictly related to the different communicative situation, but to the informative density of the text. Meta-textual $Cx_sNP[ØVP]$ occur in texts that present a higher degree of informative planning, both in dialogues and monologues: the higher the amount of information, the higher the probability that the meta-textual $Cx_sNP[ØVP]$ are used.

References

Auer, Peter. 2009. On-line syntax: Thoughts on the temporality of spoken language. *Language Sciences* 31(1): 1–13. https://doi.org/10.1016/j.langsci.2007.10.004

Bally, Charles. 1922. Copule zéro et faits connexes. *Bulletin de la Societé Linguistique de Paris* 23 : 1–6.

Barton, Ellen. 1990. *Nonsentential Constituents: A Theory of Grammatical Structure and Pragmatic Interpretation* [Pragmatics & Beyond New Series 2]. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.2

Barton, Ellen & Progovac, Ljiljana. 2005. Nonsententials in minimalism. In *Ellipsis and Non-Sentential Speech*, Reinaldo Elugardo & Robert Stainton (eds), 71–93. Dordrecht: Springer. https://doi.org/10.1007/1-4020-2301-4_4

Bazzanella, Carla. 1994. *Le facce del parlare. Un approccio pragmatico all'Italiano parlato*. Firenze: La Nuova Italia.

Bazzanella, Carla. 2005. Segnali discorsivi e sviluppi conversazionali. In *Italiano parlato. Analisi di un dialogo*, Federico Albano Leoni & Rosa Giordano (eds), 137–158. Napoli: Liguori.

Benincà, Paola. 1988. L'ordine degli elementi della frase e le costruzioni marcate. In *Grande grammatica Italiana di consultazione*, Vol. 1, Lorenzo Renzi (ed.), 129–194. Bologna: Il Mulino.

Benveniste, Émile. 1950. La phrase nominale. *Bulletin de la Societé Linguistique de Paris* 46: 19–36.

Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan & Finegan, Edward. 1999. *The Longman Grammar of Spoken and Written English*. London: Longman.

Blanche-Benveniste, Claire. 2008. Les énoncés sans verbe en français parlé. In *La comunicazione parlata. Atti del Congresso Internazionale, Napoli 23–25 febbraio 2006*, Massimo Pettorino, Antonella Giannini, Marianna Vallone & Renata Savy (eds), 1716–46. Napoli: Liguori.

Bloomfield, Leonard. 1933. *Language*. London: George Allen & Unwin.

Calaresu, Emilia. 2004. *Testuali parole. La dimensione pragmatica e testuale del discorso riportato*. Milano: FrancoAngeli.

Cresti, Emanuela & Moneglia, Massimo. 2005. *C-Oral-Rom. Integrated Reference Corpora for Spoken Languages* [Studies in Corpus Linguistics 15]. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.15

Cresti, Emanuela. 1998. Gli enunciati nominali. In *Atti del IV Convegno della Società Internazionale di Linguistica e Filologia Italiana (SILFI), Madrid 27–29 giugno*, María Teresa Navarro Salazar (ed.), 171–191. Pisa: Cesati Editori.

Cresti, Emanuela. 2003. Modalité et illocution dans le topic et le comment. In *Macro-syntaxe et pragmatique*, Antonietta Scarano (ed.), 133–182. Roma: Bulzoni.

Cresti, Emanuela. 2005. Enunciato e frase: Teoria e verifiche empiriche. In *Italia linguistica: Discorsi di scritto e di parlato. Scritti in onore di Giovanni Nencioni*, Marco Biffi, Omar Calabrese & Luciana Salibra (eds), 249–60. Siena: Prolagon.

Cresti, Emanuela. 2016. Dalla struttura informativa (alla prosodia) alla sintassi: Dati sulla subordinazione nell'italiano parlato. In *Livelli di analisi e fenomeni diiInterfaccia. Atti del XLVII Congresso Internazionale di Studi della Società di Linguistica Italiana. Salerno 26–28 settembre 2013*, Annibale Elia; Claudio Iacobini, Miriam Voghera (eds), 53–76. Roma: Bulzoni.

De Leo, Simona. 2008. La struttura topicale in dialoghi task-oriented. In *Lingua e testi: Verso una grammatica comune* [Testi e Linguaggi 2], Miriam Voghera (ed.), 105–129. Roma: Carocci.

De Mauro, Tullio & Thornton, Anna Maria. 1985. La predicazione: Teoria e applicazioni all'italiano. In *Sintassi e morfologia della lingua Italiana d'uso. Atti del XVII Congresso Internazionale di Studi della SLI*, Annalisa Franchi De Bellis & Leonardo M. Savoia (eds), 407–419. Roma: Bulzoni.

De Mauro, Tullio. 1974. Premesse a una raccolta di tipi sintattici. In *Fenomeni morfologici e sintattici dell'Italiano contemporaneo. Atti del VI Congresso Internazionale di Studi della SLI. Roma 4–6 settembre* 1972, Mario Medici & Antonella Sangregorio (eds), 551–554. Roma: Bulzoni.

De Mauro, Tullio, Mancini, Federico, Vedovelli, Massimo & Voghera, Miriam. 1993. *Lessico di frequenza dell'Italiano parlato (LIP)*. Milano: Etaslibri.

Deppermann, Arnulf. 2015. Retrospection and understanding in interaction. In *Temporality in Interaction* [Studies in Language and Social Interaction 27], Arnulf Deppermann & Susanne Günthner (eds), 57–93. Amsterdam: John Benjamins. https://doi.org/10.1075/slsi.27.02dep

Du Bois, John W. 2014. Towards a dialogic syntax. *Cognitive Linguistics* 25(3): 351–410. https://doi.org/10.1515/cog-2014-0023

Ferrari, Angela. 2003. Pour une analyse informationnelle de l'écrit. In *Macro-syntaxe et pragmatique*, Antonietta Scarano (ed.), 213–241. Roma: Bulzoni.

Ferrari, Angela. 2014. *Linguistica del testo. Principi, fenomeni, strutture*. Roma: Carocci.

Giordano, Rosa & Voghera, Miriam. 2009. Frasi senza verbo: Il contributo della prosodia. In *Sintassi storica e sincronica dell'Italiano. subordinazione, coordinazione, giustapposizione. Atti del X Congresso della Società Internazionale di Linguistica e Filologia Italiana (SILFI)*, Agela Ferrari (ed.), 1005–1024. Firenze: Cesati.

Graffi, Giorgio. 2001. *200 Years of Syntax* [Studies in the History of the Language Sciences 98]. Amsterdam: John Benjamins. https://doi.org/10.1075/sihols.98

Hjelmslev, Louis. 1948 [1971]. Le verbe et la phrase nominale. In *Essais linguistiques*, Louis Hjelmslev, 174–200. Paris: Editions de Minuit.

Jespersen, Otto. 1924. *The Philosophy of Grammar*. London: George Allen & Unwin.

Lefeuvre, Florence. 1999. *La phrase averbale en Français*. Paris: L'Harmattan.

Masini, Francesca. 2017. *Grammatica delle costruzioni. Un'introduzione*. Roma: Carocci.

Mauri, Caterina & Sansò, Andrea. 2018. Strategie linguistiche per la costruzione on-line di categorie: un quadro tipologico. In *Tipologia e 'Dintorni': Il metodo tipologico alla intersezione di piani di analisi*, Giuseppe Brincat & Sandro Caruana (eds), 209–232. Roma: Bulzoni.

Mauri, Caterina. 2017. Building and interpreting ad hoc categories: A linguistic analysis. In *Formal Models in the Study of Language*, Johanna Blochowiak, Cristina Grisot, Stephanie Durrleman-Tame & Christopher Laenzlinger (eds), 297–326. Berlin: Springer. https://doi.org/10.1007/978-3-319-48832-5_16

Meillet, Antoine. 1906. La phrase nominale en Indo-européen. *Mémoires de la Societé Linguistique de Paris* 14: 1–26.

Merchant, Jason. 2004. Fragments and ellipsis. *Linguistics and Philosophy* 27: 661–738. https://doi.org/10.1007/s10988-005-7378-3

Merchant, Jason. 2006. Small structures a sentientialist perspective. In *The Syntax of Nonsententials. Multidisciplinary Perspectives*, Ljiljana Progovac [Linguistik Aktuell/Linguistics Today 93], Kate Paesani, Eugenia Casielles & Ellen Barton (eds), 73–91. Amsterdam: John Benjamins. https://doi.org/10.1075/la.93.05mer

Mortara Garavelli, Bice. 1971. Fra norma e invenzione: Lo stile nominale. *Studi di Grammatica Italiana* 1: 271–315.

Progovac, Ljiljana. 2006. The syntax of nonsententials: Small clauses and phrases at the root. In *The Syntax of Nonsententials. Multidisciplinary Perspectives* [Linguistik Aktuell/Linguistics Today 93], Ljiljana Progovac, Kate Paesani, Eugenia Casielles & Ellen Barton (eds), 33–71. Amsterdam: John Benjamins. https://doi.org/10.1075/la.93.04pro

Progovac, Ljiljana. 2013. Non-sentential vs. ellipsis approaches: Review and extensions. *Language and Linguistics Compass* 7: 597–617. https://doi.org/10.1111/lnc3.12044

Sacks, Harvey, Schegloff, Emanuel A. & Jefferson Gail. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50(4): 696–735. https://doi.org/10.1353/lan.1974.0010

Sammarco, Carmela. 2017. L'espressione delle relazioni grammaticali nelle costruzioni senza verbo dell'Italiano e del Francese parlati. PhD dissertation, Università degli Studi di Salerno.

Sammarco, Carmela. 2020. Le costruzioni senza verbo nell'organizzazione dei testi dialogici e monologici dell'italiano e del francese parlati. In *Per una prospettiva funzionale sulle costruzioni sintatticamente marcate / Pour une perspective fonctionnelle sur les constructions syntaxiquement marquées*, Anna-Maria De Cesare & Mervi Helkkula (ed.). *Neuphilologische Mitteilungen* 120: 269–292.

Savy, Renata & Alfano, Iolanda. 2016. La richiesta di informazione nei dialoghi task-oriented: Aspetti di interfaccia prosodia-pragmatica in prospettiva intra- e inter- linguistica. In *Livelli di analisi e fenomeni di interfaccia. Atti del XLVII Congresso Internazionale di Studi della Società di Linguistica Italiana. Salerno 26–28 settembre 2013*, Annibale Elia; Claudio Iacobini, Miriam Voghera (eds), 205–229. Roma: Bulzoni.

Scarano, Antonietta. 2003. Les constructions de syntaxe segmentée: Syntaxe, macro-syntaxe et articulation de l'information. In *Macro-syntaxe et pragmatique*, Antonietta Scarano (ed.), 134–183. Roma: Bulzoni.

Schegloff, Emanuel A. 1982. Discourse as an interactional achievement: Some uses of "uh huh" and other things that come between sentences. In *Georgetown University Roundtable on Languages and Linguistics*, Deborah Tannen (ed.), 71–93. Washington DC: Georgetown University Press.

Schegloff, Emanuel A. 2007. *Sequence Organization in Interaction. A Primer in Conversation Analysis*, Vol. 1. Cambridge: CUP. https://doi.org/10.1017/CBO9780511791208

Serpollet, Noëlle, Bergounioux, Gabriel, Chesneau, Annie & Walter, Richard. 2007. A large reference corpus for spoken French: ESLO 1 and 2 and its Variations. In *Proceedings of Corpus Linguistics Conference Series*, Matthew Davies, Paul Rayson, Susan Hunston & Pernilla Danielsson (eds). Birmingham: University of Birmingham.

Sinclair, John M. & Coulthard, Malcom. 1975. *Towards an Analysis of Discourse: The English Used by Teachers and Pupils*. Oxford: OUP.

Stainton, Reinaldo. J. 2006. *Words and Thoughts*. Oxford: OUP. https://doi.org/10.1093/acprof:oso/9780199250387.001.0001

Voghera, Miriam. 1992. *Sintassi e intonazione dell'Italiano parlato*. Bologna: Il Mulino.

Voghera, Miriam, Basile, Grazia, Cerbasi, Donato & Fiorentino, Giuliana. 2004. La sintassi della clausola nel dialogo. In *Il parlato Italiano. Atti del Convegno Nazionale sulla Comunicazione parlata*, Federico Albano Leoni, Francesco Cutugno, Massimo Pettorino, Renata Savy, (eds), cd-rom, B17. Napoli: D'Auria Editore.

Voghera, Miriam, Landolfi, Annamaria & Sammarco, Carmela. 2010. Verbless clauses in Italian, Spanish and English: A treebank annotation. In *Statistical Analysis of Textual Data. Proceedings of 10th International Conference Journée d'Analyse Statistique des Données Textuelles (JADT) 9–11 june 2010- Sapienza University of Rome*, Sergio Bolasco, Isabella Chiari & Luca Giuliano (eds), 1187–1194. Milano: LED Edizioni Universitarie.

Voghera, Miriam, Iacobini, Claudio, Savy, Renata, Cutugno, Franco, De Rosa, Aurelio & Alfano, Iolanda. 2014. VoLIP: A searchable Italian spoken corpus, in complex visibles out there. In *Language Use and Linguistic Structure, Proceedings of the Olomouc Linguistics Colloquium*, Ludmila Veselovská & Markéta Janebová (eds), 627–640. Olomouc: Palacký University.

Voghera, Miriam. 2017. *Dal parlato alla grammatica*. Roma: Carocci.

Exemplification in interaction

From reformulation to the creation of common ground

Alessandra Barotto and Maria Cristina Lo Baido

University of Bologna / University of Cagliari

The aim of this paper is to examine how exemplification is used in real-time interactions to make reference to conceptual categories. Based on real occurrences of spoken Italian, it is shown that in conversations exemplification is used cooperatively by different participants to perform several functions. Not only can exemplification be used to present or expand the reference to a category by both speaker and addressee, but it can also contribute to the creation of a mutually accepted common ground (examples can be employed to communicate that alignment has been reached or to communicate how such alignment could be reached). Finally, we argue that examples are part of a bigger picture of cooperative reference construction, where speakers collaborate reformulating the reference using different types of strategies.

Keywords: exemplification, collaborative model, reference, categorization, discourse analysis

1. Introduction

The aim of this paper is to examine how exemplification is used by both speaker and addressee while making reference to conceptual categories in spoken interactions. In this regard, we will propose that the cooperative model theorized by Clark and Wilkes-Gibbs (1986) can be fruitfully applied to the study of exemplification in interaction.

Traditionally, exemplification has been mainly studied as a means of reformulation used by a single speaker or writer (cf. Halliday & Hasan 1976; Bazzanella 1995). In our analysis, it will be shown that, in conversations, examples are used cooperatively by different participants to perform a wider range of functions. Crucially, not only do speakers use examples to introduce and expand their reference, but

even addressees exploit exemplification (*i*) to expand the speaker's reference and actively help the referring process, and (*ii*) to provide or ask for feedback regarding the reference. In the end, we will argue that exemplification can function also to reach conversational alignment showing the speaker's propositional attitude of acceptance with respect to the other speaker's move or act (i.e. intentions), on a par with discourse markers (e.g. *yeah, indeed, in fact*, and so on).

To achieve our goal, we will analyse real occurrences taken from two corpora of spoken Italian, namely the LIP Corpus and the KIParla Corpus. We will focus mainly on those occurrences where exemplifying processes span over more than one turn and among different speakers.

The paper is structured as follows: Section 2 outlines the theoretical background of the study starting from the discussion of the general issue of referring to conceptual categories (2.1), to the discussion of the premises of the collaborative model of referring (2.2), and to finally focusing on exemplification in the creation of reference (2.3). Some methodological remarks are made in Section 3, where we will provide a list of linguistic strategies that can be used to signal instances of exemplification in Italian, and we will briefly describe the parameters of analysis. Section 4 and 5 will focus on the discussion of the corpus data. More specifically, in Section 4 we will discuss the use of exemplification by the speaker to introduce and expand the reference. In Section 5 we will discuss the usage of exemplification by the addressee to expand the reference (5.1), to provide feedback and accept the speaker's reference (5.2) to ask for feedback when the reference is not clear (5.3). Section 6 contains some concluding remarks.

2. Exemplification in interaction and the creation of reference: Some theoretical premises

2.1 The issue of making reference to conceptual categories

Categorization is an essential cognitive mechanism allowing individuals to organize their knowledge of the world into processable chunks (cf. Bruner et al. 1956; Zillmann 2002). Despite the pervasiveness of categorization theories in many different academic fields (cf. Overstreet 1999: 34), comparatively less attention was paid to the possible ways in which speakers can make reference to categories in real life interactions. This fact is likely linked to the traditional assumptions behind the notion of categorization. For many centuries, the so-called 'classical view' has considered categories as discrete, context-independent and representing stable features. In the last decades of the twentieth century, such a conceptualization was called into

question, leading to the formulation of a new model by Rosch (1973, 1977, 1978). According to her 'Prototype theory', categories do not have clear-cut boundaries and there is no single feature that all category members must have. These two models differ in many aspects but one: they both tend to focus on 'natural categories' (e.g. BIRD, FRUIT, FURNITURE), assuming the existence of some kind of underlying representations of categories in the human mind. This shows some consequences also at the linguistic levels: natural categories can create fixed links with specific words (e.g. *birds*) or short expressions (e.g. *musical instruments*). While these models inspired works on semantics (cf. Cruse 1990), the idea that categories can be named through category labels has likely blocked further investigation on different types of linguistic strategies to refer to categories.

The conceptualization of ad hoc categories by Barsalou (1983) marks an important turn in the conversation around the reference of conceptual categories. According to Barsalou, individuals can create functional categories on the spot to achieve a specific goal (e.g. the category PLACES TO LOOK FOR ANTIQUE DESKS while looking for a new antique desk). These categories are volatile since they do not have well-established representation in the memory. They are also context-dependent since they are construed according to a specific situational context. Because of this, the reference to ad hoc categories represents a challenge for linguists. Ad hoc categories cannot hold stable associations with dedicated words or linguistic expressions, and their reference needs to be built (and maybe even negotiated with the addressee) any time they are used in a specific interaction. This issue becomes even more crucial when we consider that some scholars (see for instance Smith and Samuelson 1997) have proposed that all categories are *ad hoc*, denying the existence of stable representations.

In linguistics, some studies have addressed this issue by investigating linguistic strategies that can be used by speakers to make reference to ad hoc categories (see Channell 1994; Overstreet 1999 for their discussion on general extenders, Mauri & Sansò 2018 for their typological analysis of the linguistic strategies that can encode ad hoc categorization). Interestingly, all these studies identify strategies that rely on providing some exemplars of the category as a starting point of an inferential process (Mauri 2017). Barotto (2021) notes that speakers essentially have two main strategies to make reference to conceptual categories: lexicalization and exemplification. In the first case, speakers create and use abstract formulations (or labels), in an effort to name the category (cf. (1)). In the second case, speakers mention one or more items marking them with some linguistic strategies that signal their status as exemplars of the category (cf. (2)). These two strategies can also function together (cf. (3)).

(1)   *sono veramente molto interessanti come associazione pero' ci stiamo interessando anche di altre esperienze che ci sono in Italia e all' estero che tendono direi a prevenire **difficolta' che possono sorgere sia sul piano fisico che sul piano psichico eh da parte degli anziani***   [LIP Corpus – MD17]
      'they are really very interesting as an association but we are also interested in other experiences that exist in Italy and abroad that tend to prevent **difficulties that can arise both on the physical level and on the psychic level eh in elderly people**'

(2)   *non sono un'amante della **non so di discoteche e cose cosi'***
          [KIParla Corpus – BOD2001]
      'I'm not much into **dunno into nightclubs, and things like that**.'

(3)   *non abbiamo la possibilita' di fare **lotte tipo l'occupazione autogestione***
          [LIP Corpus – MC4]
      'we don't have the possibility of doing [political] **fights such as occupation, self-management**'

In our paper, we investigate the role of examples in the creation of reference to conceptual categories in spoken conversations, with a special focus on the interactional dimension. We argue that spoken interactions represent a privileged environment to study categorization processes because, contrary to writing, they are characterized by an intrinsic online dimension (cf. Auer 2009). As already noted, ad hoc categories are not retrieved from stable representation, but are created on the spot, in a specific context, whenever they are needed. We may therefore assume that (*i*) the cognitive construal happens simultaneously to the spoken interaction, and thus that (*ii*) the spoken interaction conserves traces of the categorization process as happens in the speaker's mind. For this reason, we believe that examining the way speakers build the reference to conceptual categories in spoken interactions can provide important insights into the way categories are created and used at the cognitive level.

2.2   The collaborative model of referring

Traditionally, referring has been considered as an ideational or subjective operation (cf. Redeker 1990), which pertains to the communication of abstract concepts as elaborated by the speaker. Some scholars theorized what can be called a *literary* model of reference construction (cf. Clark & Wilkes-Gibbs 1986). In such a model, the speaker is claimed to refer as if she were writing to a distant and absent reader. In this regard, the act of referring is cotemporal with the uttering of a specific expression and the speaker satisfies her intention to make reference simply by issuing the selected noun phrase (Clark & Wilkes-Gibbs 1986: 3). More importantly,

according to this model, the speaker holds complete responsibility and control over the referring process. The addressee has no power over the speaker's reference and can only hear the noun phrase and infer the identity of the referent. Finally, this model does not consider the role of the context, since the referential meaning is assumed as unreceptive of the effects of context.

This model shows clear shortcomings when real-life conversations are considered, because it starts from the assumption that spoken interactions and written texts are substantially similar. On the contrary, as noted by Auer (2009: 1ff), speaking and writing differ in many aspects, many of which can be summarized through the notion of *temporality*. Specifically, speaking shows a temporal structure which is the fundamental result of an online interactive process. This means that unlike writing – in which only the final product comes into contact with the addressee, "spoken language must (and can) reckon with interactive openness from the very outset" (Auer 2009: 1). Therefore, for instance, spoken language is characterized by irreversibility (i.e. what is said cannot be cancelled, Auer 2009: 3), whose consequences involve editing phenomena such as hesitations, self-repairs, and restarts.

To better describe the creation of reference in spoken interaction, Clark and Wilkes-Gibbs (1986) propose a different model, called the collaborative model (or conversational model). According to it, conversation should be considered as a joint activity, which can be defined as "the coordination of individual actions by two or more people" (Clark 1996: 59). Pivotal to this model is the notion of *grounding* (or *common ground*, cf. Clark & Brennan 1991; Clark 1996; Clark & Schaefer 1989), a process through which participants establish the mutual belief that a specific utterance has been understood as intended. Once this mutual belief is established, then the reference communicated by the speaker becomes part of the common ground (Clark & Wilkes-Gibbs 1986: 7–9).

Two aspects are thus crucial. First, grounding is intrinsically a joint responsibility of both participants: speakers and hearers must "go beyond autonomous actions and collaborate with each other, moment by moment, to try to ensure that what is said is also understood" (Schober & Clark 1989: 211). Second, the common ground is inherently dynamic since the grounding process constantly occurs during the conversation. This means that participants may start with certain assumptions about their common ground, but new beliefs can emerge over the conversation as speakers introduce new concepts or elaborate on previous ideas.

The grounding process consists of two basic moments: (1) presentation and (2) acceptance/rejection. In the first phase, the speaker presents some information to the addressee. This phase can be in turn divided into two sub-phases. First, the speaker initiates the reference usually by means of a noun phrase. If she is not sure about the expression that should be used, she can use some other expressions like dummy noun phrases (e.g. *whatchamacallit*) or proxy noun phrases (e.g. *what's*

the word, you know, see Clark & Wilkes-Gibbs 1986: 18). Then, if the initial expression is not considered acceptable, it can be refashioned in three main ways: (i) *repair*, that is, the speaker detects a problem in her referring process and thus she self-repairs it (cf. Levelt 1983; Schegloff et al. 1977); (ii) *expansion*, that is, the speaker expands the initial noun phrase which is deemed insufficient for the referring purposes; (iii) *replacement*, that is, the hearer can reject the noun phrase of the speaker and replace it with a different description (that, in turn, is accepted or rejected by the speaker, see Clark & Wilkes-Gibbs 1986: 20–23).

In the second phase, the addressee responds with an indication of understanding (or not understanding) of the previously presented information. This phase can occur explicitly or implicitly: addressees can use words, gestures, nodding, etc. Finally, the process can be concluded by a final phase of acknowledgment, in which the speaker acknowledges in some way the acceptance (or rejection) of the addressee.

To investigate this model, Clark and Wilkes-Gibbs (1986) used cards each showing one of the so-called Tangram figures. In other words, the participants were asked to make reference to abstract figures that do not have specific words or expressions to name them. In these cases, the reference needs to be negotiated by creating ad hoc expressions to name them, by reformulating these expressions, by expanding them, etc.

We argue that, in a way, this process is not very dissimilar to what people do when they try to make reference to ad hoc categories. As noted, ad hoc categories do no hold stable representation in the mind and do not have a link with specific words or expressions to name them. The reference needs to be created and negotiated in any conversation, according to the situational context. While the lexicalization process (i.e. the creation of a category label or names, see Barotto 2021) is likely consistent with the analysis provided by Clark and Wilkes-Gibbs (1986), how examples are employed in this process (and more generally, the role of exemplification) needs to be investigated further.

2.3   Exemplification in the creation of reference

Although exemplification was not explicitly addressed with regards to the collaborative model, there exist some studies that discuss the usage of exemplification in discourse, with a specific focus on discourse coherence. Even though some of these studies analyse written texts, they provide clues as to what we may expect investigating exemplification in spoken interactions.

In taxonomies on coherence relations (cf. Halliday & Hasan 1976; Hobbs 1979, 1985), exemplification is generally recognized as a subtype of elaboration relation, that is, relations in which a secondary part of the text or *satellite* (cf. Mann & Thompson 1988) contributes to the important point of the text or *nucleus* by

elaborating the material provided through the nucleus. For instance, Longacre (1983: 83–84) describes exemplification as a type of illustration that allows to elaborate an abstract formulation by means of some concrete instances. Similar classifications are also provided by Mann and Thompson (1988) and Hovy and Maier (1994).

Because of this elaborative function, exemplification has often been examined in relation to another important discursive phenomenon, namely reformulation. For instance, Hyland (2007: 268) notes that exemplification and reformulation have the same basic functions of clarifying the writer/speaker's communicative goal and negotiating the meaning in different contexts. He describes exemplification as a clarification process "through which meaning is clarified or supported by a second unit which illustrates the first by citing an example" (2007: 270). Studies on discourse markers[1] go even further recognizing exemplification as a specific type of reformulation (cf. Bazzanella 1995) or as an elaboration process signalling some sort of refinement on the preceding discourse (Fraser 1996: 187–188). Consider again (3), repeated here as (4).

(4)   *non abbiamo la possibilita' di fare **lotte tipo l'occupazione autogestione***

  [LIP Corpus – MC4]

  'we don't have the possibility of doing [political] **fights such as occupation, self-management**'

The occurrence in (4) represents what the literature considers a prototypical instance of exemplifying construction (cf. Rodríguez Abruñeiras 2015: 55). It consists of a general element or statement with a broad referent (i.e. *lotte* 'fights') which is reformulated by means of an exemplifying element which is more specific and whose referent is included within the referent of the general element (i.e. *occupazione autogestione* 'occupation, self-management').

Applying this to the collaborative model, we may think that exemplification is basically a form of expansion, that is, a way in which the speaker expands the initial name or label of the category by means of some concrete exemplars. Nevertheless, it is noteworthy that exemplification cannot always be recognized as a type of reformulation. For instance, in (2), the speaker provides only a list of exemplars to make reference to a conceptual category, without any general name or label. In such cases, it is hard to argue that some sort of reformulation takes actually place. More importantly, it appears that, in some cases, instead of using some sort of noun phrase, the speaker may initiate the reference directly by means of exemplification.

---

1.   Discourse markers are procedural strategies playing a threefold role contributing (*i*) to the discourse management (i.e. textual coherence); (*ii*) to the speaker/hearer interaction (i.e. interactional and interpersonal meanings), (*iii*) to speaker's attitudes (i.e. epistemic meaning), see Degand 2014: 151; cf. also Brinton 2006; Fischer 2006; Heine 2013, among many others.

Starting from this naïve observation, we would like to expand the discussion, analyzing the different ways in which exemplification can be exploited in the collaborative construction of reference. In doing so, we will not only monitor the initial phase of presentation usually performed by the speaker, but also the acceptance phase performed by the addressee. In other words, moving a step further from the idea of exemplification as a type of reformulation or expansion, we would like to investigate if exemplification can perform functions that are interactionally motivated such as accepting the reference.

3.   Methodological remarks: Corpora, objects of analysis and parameters

The aim of our paper is to provide an account of the different ways through which exemplification is used in construing the reference of conceptual categories. To achieve our aim, we adopt a qualitative corpus-based approach, examining data from spoken Italian gathered through the LIP Corpus and the KIParla Corpus.

The LIP (Lessico di Frequenza dell' Italiano Parlato) Corpus comprises 469 texts of spoken Italian collected between 1990 and 1992 in four cities (Milan, Florence, Rome, and Naples), for a total of approximately 490,000 words (see De Mauro et al. 1993). The texts are classified into five macro-types: (a) bi-directional exchange, face to face, with free turn-taking; (b) bi-directional exchange, not face to face, with free turn-taking; (c) bi-directional exchange, face to face, with regulated turn-taking; (d) unidirectional exchange, with the addressee being present; (e) distanced unidirectional exchange.

To complement our survey with more recent data, we used the KIParla Corpus (KIP module) (Goria & Mauri 2018), which was collected between 2016 and 2019 in two cities, Bologna and Torino. The corpus collects different types of interaction recorded in academic contexts, for a total of approximately 700.000 words. In particular, it comprises: (a) professor-student interactions during office hours; (b) random conversations recorded by in-group members (students and professors) without direct involving of the researcher; (c) professor-student interactions in oral examinations; (d) academic lessons; (e) semi-structured interviews collected by students within the peer-group and aimed at the elicitation of oral narratives.

Since our aim is to investigate the use of exemplification by both speakers and addressee (see Section 2.2), we focused our research mainly on bi-directional exchanges, where the hearer is present and actively participates to the conversation.

To monitor instances of exemplification, we selected several exemplifying strategies, ranging from single words to parenthetical constructions. These strategies can be used to signal that the word, phrase, sentence or text under their scope is an example of a given or potential set (see Manzotti 1998), even though their semantics

may apparently be quite distant from that of exemplification (e.g. *per dire* 'to say', or *magari* 'maybe').[2]

The exemplifying constructions under analysis are nominal, verbal, and adverbial strategies. The nominal strategies are: *ad/per esempio* ('for instance'), *tipo* ('like', see Voghera 2013), and *per ipotesi* ('to make a hypothesis'). The verbal strategies are *non so* ('I don't know'), *che (ne) so* (lit. 'what do I know (about it)', id. 'for instance', see Schneider 1999, 2007; Molinelli 2014), *per dire* ('so to speak', see Ghezzi 2013), *mettiamo/poniamo (il) caso* ('let us suppose', see Schneider 2007), *fai conto* ('consider') and *voglio dire* (lit. I want to say', id. 'I mean'). Finally, the adverbial expressions examined are *magari* ('maybe', see Manzotti 1998; Masini & Pietrandrea 2010) and *anche* ('also').[3] Some of these expressions are dedicated to exemplification in that their semantics is transparent to the process of exemplification (e.g. *ad/per esempio* see also Rodríguez Abruñeiras 2015), while others can be mobilized for this specific function, which thus emerges contextually (e.g. the parenthetical comment clauses *non so* or *che so* or the pseudo-purpose conditional *per dire* – Prandi 2006).[4]

In order to distinguish different types of uses of exemplification, some parameters are considered. Specifically, we monitor (*i*) whether the exemplification is part of the speaker's utterance or the addressee's utterance; (*ii*) in what stage of the co-operative model (i.e. presenting the reference, expanding the reference, acceptance, acknowledging) the act of exemplification occurs. Thus, we also monitor (*i*) the co-occurrence with reformulation markers (e.g., *cioè, volevo dire* 'I mean', *in altre parole* 'in other words'); (*ii*) the presence of explicit markers of acceptance (e.g. *sì* 'yes', *ho capito* 'I understand', *certo* 'of course' cf. Wang et al. 2010) and markers of disagreement or doubts (e.g., *non lo so* 'I don't know', *forse* 'maybe', *assolutamente, affatto* 'at all').

---

2. When these markers are used to encode exemplification, they can be paraphrased as the most transparent exemplifying marker *per esempio* 'for example'.

3. For a comprehensive discussion on Italian exemplifying markers and constructions, see Lo Baido (2018: 74–75).

4. Beyond the constructions listed above, we have also monitored the following strategies that emerge in spoken discourse and can sometimes be used also to convey exemplification: *tanto per buttar lì un valore* ('just to throw a value'), *faccio per dire* ('I manage to say'), *mettiamo un'ipotesi* ('let us make a hypothesis'), *ti faccio/facciamo un esempio* ('I show you an example'), *si pensi* (lit. 'one thinks', id. 'let's say; see, for instance'), *vedi* ('see'), *si prenda* (lit. 'one takes', id. 'let's say'; 'see, for instance'), *si veda* (lit. 'one sees', id. 'let's say'; 'see, for instance'), *ricordiamo* ('let us remember'), *prendiamo* ('let us take'), *si consideri* (lit. 'one considers', id. 'let's say'; 'see, for instance'), *facciamo un caso* (lit. 'let us make a case', id. 'let's consider'; 'let's say'), *per dirne una* (lit. 'just to say one thing', id. 'just to mention one example'), *altro esempio* ('further example'), *per fare un esempio* ('to make an example'), *io prendo* (lit. 'I take', id. 'I consider the example x'), *basti pensare* (lit. 'just think', id. 'let's say'; 'see, for instance'), *boh* (lit. 'dunno', id. 'for instance').

Following the model described in Section 2.2, our analysis will be divided into two stages. In the first one, we will investigate the use of exemplification by the speaker to present and refashion the reference. In this regard, we will address previous studies on the issue (cf. Barotto 2018, 2021; Lo Baido 2018; Barotto and Mauri 2018), while highlighting new aspects that emerged from our data. In the second stage, we will investigate the usage of exemplification by the addressee to (i) collaborate in the referring process by expanding the speaker's reference, (ii) accept the reference as the speaker has presented it, (iii) ask for feedback whenever the reference presented by the speaker is not clear enough.

4.   Exemplification by the speaker: Presenting and refashioning the reference

In this section, we briefly analyse how the speaker uses exemplification to create reference following the model theorized by Clark and Wilkes-Gibbs (1986). As already noted, the role of exemplification in refashioning (in the sense of reformulating) the reference is not a novelty in the literature and has been analysed from different perspectives. Still, our data provide some interesting insights that should be discussed.

The first interesting point is that examples are used also to introduce a referent. Traditionally, studies on reference have theorized that speakers introduce their referents by means of a noun phrase (see Section 2.2). The collaborative model has expanded this notion adding that, in some cases, speakers may use dummy or proxy noun phrases whenever they are unsure of the exact expression to use. Nevertheless, beyond the identification of a potential name or label for the referent, no other strategies are actually considered.

In our data, there are cases where the speaker initiates the act of referring not by formulating a category name, but by listing some exemplars of the category, as shown in the following occurrences:

(5)   *assistiamo sempre alle stesse immagini eh cambiano soltanto **magari i nomi delle battaglie o il numero dei dei feriti eccetera*** [LIP Corpus – ME8]
'we always see the same images and only **maybe/for example the names of the battles or the number of of causalities, and so on** change'

(6)   *c'e' anche una fatica reale **a leggere** cioe' lui per lui **leggere per esempio anche fare gli esercizi anche scrivere** non so* [LIP Corpus – FA3]
'there is also a real effort **to read** that is for him **to read for example even to do exercises even to write** dunno'

In (5) and (6), speakers simply introduce the target category by means of exemplification. In (5), the speaker makes reference to a category of small war details that change every time ('the name of the battles or the number of causalities'), standing in contrast to the idea that war images look always the same. In (6), the speaker wants to refer to a category of activities that another person finds extremely difficult to perform without getting tired or bored. To do so, he lists some verbal forms (i.e. 'read', 'do exercises', 'write') to exemplify some members of the category.

Of course, exemplification is also used to refashion the reference expressed by means of a previous noun phrase. Nevertheless, inside the very general macro-function of reformulating, it is possible to identify more fine-grained functions. In particular, examples can be used (i) to contextualize the reference, and (ii) to disambiguate the reference (see also Barotto 2018: 47–50). The first function involves those cases in which the speaker tries to make reference to a contextually relevant category using a label which denotes a broader referent, and then provides contextual information by means of exemplification, as in (7).

> (7) *io ho letto dei* **giornali** *oggi eh* **tipo il Corriere della sera o Repubblica** *che sono la stragrande tiratura e ci sono delle cose folli cioè $ $ secondo loro il movimento sta scemando […]*   [LIP Corpus – MC4]
> 'I have read some **newspapers** today eh **for example Il Corriere della sera or Repubblica** which have an overwhelming circulation and there are crazy things, that is, according to them the movement is falling off […]'

The speaker uses a very general label *giornali* 'newspapers' to make reference to a conceptual category. If we consider only the label provided, we might think that it well represents the category the speaker has in mind, and that he is making reference to all (Italian) newspapers in general. However, when we consider the actual broad context (e.g. *stragrande tiratura* 'overwhelming circulation'), we realize that the target category is likely more specific than the one denoted by the label. The contextualizing process is helped by the examples provided by the speaker, which – ideally – represent good exemplars of the target category: *Il Corriere della Sera* and *Repubblica* are indeed the most widely read newspapers in Italy. Therefore, in this case, exemplification allows to tailor the category to the context, signalling what types of newspapers are contextually relevant and should thus be included in the target category.

The second function involves those case in which the speaker tries to provide contextual information by creating and using more detailed labels. However, since there is an unstable link between these *ad hoc* labels to their *ad hoc* referents, the speaker is often compelled to disambiguate the reference by means of exemplification, as in (8).

(8) *l'obbiettivo generale degli approcci plurali è quello di eh portare avanti una poli-*
     tica di differenziazione linguistica al fine di superare **contesti e situazioni di**
     monolinguismo come ad esempio può essere per l'inglese

                                                    [KIParla Corpus – BOC1006]

     'the overall goal of pluralist approaches is the one of eh to pursue a policy of
     linguistic differentiation with the ultimate purpose of overcoming **contexts and**
     situations of monolingualism such as for instance it may happen for English'

The example in (8) is taken from an oral examination. The student is trying to
make reference to a specific type of policy which aims at overcoming cases of
monolingualism. To do so she makes reference to a category of situations where
monolingualism is present by using two relatively generic expressions, namely
'contexts' and 'situations'. Nevertheless, since this is not a natural category with a
stable connection with a specific label and since the expression at stake may include
various referents ranging from events to behaviours and practises, the process of
understanding the actual referent may be complex. This may be especially true in
the context of an examination, where unambiguity is required. For this reason, in
order to disambiguate her reference, the speaker provides a prototypical exemplar
of the category (i.e. English).

It is noteworthy that these two functions are not mutually exclusive since the
difference between general labels and complex detailed labels is more a matter of
degrees than a discrete distinction (Barotto 2021: 58ff). Therefore, in some case, we
may find middle-ground cases. Despite this, general labels tend to have more stable
representation and thus it is usually easier to retrieve their reference. This means
that they do not generally need disambiguation, but they do need to be tailored to
the situational context. On the other hand, complex detailed labels have already
undergone a contextualization process, but being more volatile and *ad hoc*, their
reference needs to be disambiguated.

More generally, noun phrases as category names and exemplifying construc-
tions are used one after the other, even more than once in the same referring
process, thus creating long and sometimes complex sequences. For instance, the
speaker can present or reformulate the reference by means of exemplification, and
then provide a noun phrase to name the category only in the end, as to verify the
entire previous referring process. Let us consider (9).

(9) A: *ehm com' e' il ristorante di tuo padre che tipo di ristorante e'*
     B: *e'* **un ristorante di cucina tipica** *quindi cucina tipica delle nostre parti*
         emiliana
     B: *quindi buoni primi e pero' comunque con anche*
     B: *con anche possibilita'* **di magari mangiare qualcosa di un po' piu' diciamo**
         raffinato non so delle carni
     B: *gnocco e tigelle e' comunque un* **un ristorante tipico non troppo rustico**
         diciamo   [KIParla Corpus – BOD2008]

> A: 'ehm how is your father's restaurant what kind of restaurant is it'
> B: 'it is a **restaurant with typical cuisine** and so typical cuisine from around here of Emilia'
> B: 'so good first courses but still there is also'
> B: 'also the possibility **maybe/for example of eating something a little more refined for example meat**'
> B: 'gnocco and tigelle still it is a **typical restaurant let's say not too rustic**'

B is asked to describe her father's restaurant. First, she tries to describe it using a very well-known definition, namely *ristorante di cucina tipica* 'restaurant with typical cuisine'. However, she realizes that this definition is too broad, and it might also be misleading. In particular, it might give the impression that the restaurant is more like a *trattoria* (and thus not very fancy). To address this issue, she uses exemplification to make reference to a category of refined dishes that can be ordered as well ('for example eating something a little more refined for example meat'), together with some exemplars of typical Emilian dishes ('first courses', 'gnocco', 'tigelle'). Interestingly, in the end, she rephrases her first definition by adding the crucial property that has emerged from the exemplifying construction: the restaurant is *tipico* 'typical' but *non troppo rustico* 'not too rustic'.

The analysis of these lists of more general (re)formulations and concrete examples in natural spoken occurrences provides interesting insights regarding the hypothesis that the cognitive construal of ad hoc categories likely happens simultaneously to the spoken utterance (cf. Section 2.2). Looking at (9), it appears that the speaker is construing the contextually relevant category of dishes in her mind along with the different attempts to build the reference in her speech. Some interesting consequences may thus follow. For instance, it seems that there are two ways in which individuals can build categories. The first way is the identification of the relevant property of the category, that is, that particular and contextually relevant characteristic that is shared by all the category members. At the speech level, this is mirrored by the construction of general formulations (in the form of noun phrases) that encode the property (e.g. the property of being SITUATIONS OF MONOLINGUALISM is the important feature shared by all the members of the category construed in (8)). The second method is the identification of some members of the category as a starting point of an inferential process (see Mauri 2017). As noted at the beginning of this section, speakers can introduce the intended referent by means of exemplifying constructions, suggesting some exemplars of the category.

Speakers can choose one way rather than the other, depending on the category and the context. However, they can also decide to use both, creating complex constructions as in (9), to verify the inferential process triggered by exemplification or to actualize an abstract formulation which may sound too vague or too specific.

This process can be observed also scattered across turns and different speakers. Let us consider (10).

(10) A: *no no no calma no calma io ho detto nel senso cose che noi consumiamo
          tipo tutti i giorni c' e' chi prende il caffe' chi **per esempio si lava il dentifricio
          chi si lava con l'acqua** capito * cose di questo tipo ma non cose queste cose
          qui capito *

      Y: si'

      A: per $ $

      Y: *le cose che fanno tutti cioe' tutti di sicuro

      A: *brava brava che fanno tutti e tutti i giorni

      Y: eh

      A: proviamola via

      Y: **boh mi ci butto posate le posate i piatti**

      A: *ecco   [LIP Corpus – FB14]

      A: 'I said in the sense things we use like every day there are people who drink
          coffee people' who **for example wash the toothpaste who wash with water**,
          get it? things like this but not things these things here, get it?

      Y: 'yes'

      A: 'for'

      Y: 'the things that everyone does that is everyone for sure'

      A: 'bravo bravo, everyone does and everyday'

      Y: 'eh'

      A: 'let's try it'

      Y: **'boh, I try, cutlery the cutlery plates'**

      A: 'well'

During a radio show, radio host A asks to his guest Y to guess the identity of a
specific category of objects. Again, we can see how the category and the reference
to the category are construed step-by-step, using different approaches and strat-
egies. At first, A makes reference to a category by means of a noun phrase and a
list of exemplars. The category label is then reformulated by both Y ('things that
everyone does that is everyone for sure') and A ('everyone does and everyday'), to
better highlight the property shared by the category members. In the end, Y tries
to provide her own list of exemplars ('cutlery plates').

     In (10), the construction of reference is performed cooperatively by two speak-
ers. In particular, we can see how Y not only accepts explicitly the reference of A
('yes'), but in the following turns she (i) tries to reformulate it to better encode and
highlight the pivotal property of the category, and then (ii) tries to provide some
potential exemplars of the category. In the next section, we will focus specifically
on the usage of exemplification by the addressee, in order to better understand the
role of exemplars in the collaborative model.

5.   Exemplification by the addressee: From collaborating to reference construction to interpersonal functions

In this section, we will focus on how exemplification can be used by the addressee in the construction of reference. Contrary to the use of exemplification by the speaker which has been studied in relation to different topics (e.g. discourse markers, textual coherence, categorization, cf. Section 2.3), no specific attention has been paid to how and why examples can be employed also by the addressee. In our analysis, we will focus on three functions: (1) when examples are used to actively collaborate in the construction of the reference, helping the speaker to highlight important aspect of the category, (2) when examples are used to provide an implicit or even explicit feedback on the referring process, and (3) when examples are used to ask for some feedback on the referring process.

5.1   Exemplifying to collaborate in the reference construction

Since Grice (1975), it has been argued that, in conversations, participants dynamically assume each other's cooperation. Conversation is thus claimed to be a co-operative work in which the participants seek to contribute by adding further, relevant propositions to the commonly shared set of propositions. The final aim is reaching a final relevant interpretation (cf. Relevance-theoretic account), which emerges from the cooperation between the participants to the speech event, even when their contribution is not explicitly asked.

In this regard, a crucial point of the collaborative model is that both speaker and addressee make an actual effort in signalling and making sure that all participants have correctly understood the reference (Clark and Wilkes-Gibbs 1986: 6). It is through such a process that speakers may communicate that the target reference has become part of their common ground (or, set of shared assumptions). Such alignment is achieved also by the addressee, who is not a passive recipient through the referring process. In fact, in this section, we will show that the addressee is often compelled to intervene with the purpose of collaborating in the creation and disambiguation of the reference.

As already noted in Section 2.2, a speaker may perform an operation of *expansion*, by expanding the initial noun phrase which is considered as insufficient for the referring purposes (cf. Clark & Wilkes-Gibbs 1986: 19–23). Generally, speakers tend "to make their own expansions unprompted" (Clark & Wilkes-Gibbs 1986: 27) providing either reformulation(s) of the first noun phrase or using exemplification. As far as the addressee is concerned, to maintain the collaborative process, she could in principle respond simply by showing her acceptance through a variety of

strategies (from nodding to explicitly confirming her acceptance). However, for collaborative efficiency in determining the common ground, the addressee can offer her own expansions, thus providing additional (which in some cases may be even more decisive) information regarding the referent at issue, often through constructions which are barely juxtaposed to previous speaker's utterances (cf. Calaresu 2018: 508 on the category of *giustapposizioni e combinazioni di enunciati (sintagmi e frasi)* 'juxtapositions and combinations of utterances (phrases and sentences)').[5] Exemplification can play a key role in this, since the addressee can use one or more examples of the target category to help the speaker, and ultimately optimize the mutual acceptance of the referring process (see Clark & Wilkes-Gibbs 1986: 27). Let us consider the following exchange:

(11)   B: *[…] fare compagnia a **persone che in questo momento possono essere sole** e sono tantissime eh la notte ci sono anche **persone negli ospedali per esempio** che non nominiamo mai abbastanza […] poi le **persone che lavorano che sono lì come noi a lavorare di notte** quindi ci sono categorie ben precise che lavorano ci sono **le signorine** eh **i ladri** e poi tutta un'altra categoria **pasticcieri fornai** eh*

   A: *e i centralinisti*
   B: *e i centralinisti*   [LIP Corpus – FE15]

   B: '[…] to entertain **people who can be lonely at this moment** and they are so many eh at night there are also **people in hospitals for example** that we never speak of enough […] then **people who work who are there like us to work at night** then there are specific categories that work there are **prostitutes** eh **thieves** and then a whole other category **confectioners bakers** eh'

   A: 'and telephone operators'
   B: 'and telephone operators'

In (11), B is trying to make reference to a very broad conceptual category and A intervenes to integrate such a process providing further information by means of exemplification. More specifically, B wants to make reference to the category of ALL

---

5.   We are referring to the existence of juxtaposed fragments which can give rise to what Calaresu (2018: 508, 509) argues with respect to 'polyphonic grammaticalization'. In spoken dialogue, speakers favour the arrangements of utterances, which are often not connected on a micro-syntactic level, as the ones she defines as hanging topics or left dislocations: *la tesi ci vorrà un annetto* ('the thesis, it will take one year'). Many of the examples we retrieved may be defined as online sequences which are produced on the spot, also through the cooperation of different speakers, without specific connecting strategies between the units at stake; they are simply produced through a process of addition and juxtaposition, not linearization and hierarchization, as the segment 'the thesis' in the abovementioned example (see also Haselow 2016: 90).

THOSE PEOPLE THAT ARE AWAKE AT NIGHT. In order to do so, he first splits up the category into different sub-categories (i.e. 'people who can be lonely at this moment [at night]', 'people who work […] at night') and then he utters some exemplars for each (e.g. 'people in hospitals for example', 'prostitutes, thieves […] confectioners, bakers').

Despite her intervention is not explicitly demanded by B in any way, A intervenes by providing a further example ('telephone operators'), thus expanding the list started by B through the addition (or juxtaposition) of a new segment (cf. Calaresu 2018: 508–509). This further example is functional to the construal of the category. In fact, 'telephone operators' is a different type of example from those previously mentioned by B, thus serving to highlight just how broad and heterogeneous the target category actually is (to the point of including people in hospitals, prostitutes and also telephone operators). In this sense, the intervention of A helps to complete and specify both the referring and the categorization process.

In other instances, the addressee seems compelled to intervene through exemplification because the speaker shows doubts regarding the referring process. Let us consider the following occurrence:

(12)   C: *nella guerra nella seconda guerra mondiale cioè una guerra civile erano eh diciamo erano eh # erano coinvolti anche i civili cioè le persone che non non c'entravano niente con la guerra infatti*

A: * *per esempio vittime dei bombardamenti*

C: *ahah infatti*   [LIP Corpus – NC9]

C: 'in the Second World War that is a civil war were eh let's say were eh involved also civilians I mean people who had nothing to do with the war in fact'

A: '**for example, victims of bombing attacks**'

C: 'ahah indeed'

In (12), C makes various attempts to find the right referring expression through an on-going process of replacement repairs (Schiffrin 1987: 300) and expansions, thus showing a certain degree of uncertainty regarding how to refer to the category. He starts the referring process by using the generic label 'civilians', which is broader than his target category. To avoid potential ambiguity (cf. the notion of *appropriateness-repair* in Levelt 1983: 52), C repairs such expression by means of the analytical reformulation 'people who had nothing to do with the war' introduced by the reformulation marker *cioè* 'that is'.

The referring process is then expanded by the addressee, who is likely compelled to intervene because of the various attempts made by C. Interestingly, she does so by using a different strategy. More specifically, instead of repairing the final formulation of the speaker (which in some situations may be considered inappropriate, cf. Schegloff et al. 1977), the addressee provides an actual exemplar of

the category ('victims of bombing attacks') as additional information. This fact is interesting for two reasons.

First, exemplars are much more specific than the general formulations that are used to name a specific category. For this reason, they can be used to disambiguate the reference, whenever there are some doubts over the right noun phrase. In (12), exemplification allows the addressee to disambiguate the speaker's reference without openly correcting him or repairing his formulations. The intrinsic paradigmatic nature of examples (i.e. being one arbitrarily chosen among many other alternatives, see Manzotti 1998) allows to attenuate the strength of the addressee's utterance, construing it as merely a suggestion, which the speaker can refuse or accept (see Caffi 2007: 272 on the notion of exemplification as an indirect strategy of reticence).

Secondly, the example is useful to the entire referring process because it effectively highlights the most important feature of the category, that is, that it refers to civilians who were deeply affected by the war. The addressee is thus playing an active role in the process of referring, expanding in a useful way the speaker's referring expression. In the end of the referring process, C accepts the example, giving a final feedback (i.e. *infatti* 'indeed', see Wang et al. 2010: 242, 245) to the entire cooperative process, and thus confirming the expansion suggested by the addressee (cf. the process of expansion acceptance in Clark & Wilkes-Gibbs 1986: 22).

Let us consider another example.

(13)   A: *chi vuole si porta un pubblico da casa eh puo' essere la **mamma piuttosto che eccetera non so***

D: *sì*

A: *una cosa di questo genere*

D: **il ragazzo *[…]***   [LIP corpus – MA23]

A: 'those who want can bring the audience from home eh it can be **the mother or etcetera I dunno**'

D: 'yes'

A: 'something like that'

D: '**the boyfriend** […]'

In (13), A aims at making reference to a specific category of AUDIENCE. To do so, he uses a label 'audience from home' and a potential exemplar, that is 'mother'. However, despite this apparent precision and the addressee's explicit acceptance (cf. the agreement marker *sì* 'yes'), the exemplar is still followed by a long list of exemplifying strategies, namely *non so* ('I don't know') and two general extenders,[6]

---

6.   General extenders are linguistic strategies used to signal the non-exhaustive status of what occurs in their scope. The label proposed by Overstreet (1999: 3) is motivated by the fact that they are *general* in the sense of being nonspecific, and *extenders* in the sense that they extend

piuttosto che and *eccetera*. This string of markers seems to indicate a certain amount of doubt with respect to the referring process itself. For instance, the exemplifying construction *non so* ('I don't know') explicitly removes speaker's responsibility in that it stems from the negation of the semifactive verb *sapere* ('to know') which refers to the speaker's epistemic positioning. Then, he also uses the approximation strategy *una cosa del genere* ('something like that') to further signal the existence of other potential (not really defined, quite vague) items, making his attempts to build a context-based category more explicit and clearer.

Like in (12), also in this case, the addressee is compelled to add a further exemplar ('boyfriend') to ease the speaker's difficulty in specifying the reference. In this regard, the addressee helps the speaker make explicit what type of items can be listed as potential members of the target category.

In the light of the above, we can argue that, when used by the addressee, exemplification performs functions similar to those described in Section 4 regarding the speaker (i.e. disambiguating or further contextualizing the reference). However, in this specific case, exemplification is not only semantically motivated by the need of clarifying the reference, but it is also pragmatically motivated by the interpersonal, mutual need of assisting the other participant to the speech event. The main goal is to construe the target reference in such a way that can be mutually accepted and negotiated by all the participants. This type of interpersonal function can only emerge when we consider dialogic interactions (see Calaresu 2016, 2018).

5.2   Exemplifying to provide feedback on the reference

As noted in Section 2.2, an important part of the collaborative model is providing constant feedback on the referring process while it is occurring. In particular, as participant to the speech event, the addressee is asked to accept (or in some cases refuse) the reference provided by the speaker. This is a pivotal moment to achieve that mutual acceptance on the reciprocal understanding which is the basis of the success of the referring process. In other words, it is essential to establish that the addressee has understood the speaker's reference before the conversation moves forward. This can be achieved implicitly by allowing the speaker to continue on to the next contribution, or explicitly with head nods, continuers like *yes, I see, okay* (cf. *presuppose acceptance* and *assert acceptance* in Clark & Wilkes-Gibbs 1986: 9).

---

otherwise grammatically complete utterances. As noted by the literature (cf. Channell 1994; Overstreet 1999, among others), general extenders are frequently used when speakers are not sure how to continue their statement or when they prefer (for different discourse-motivated reasons, e.g. politeness) to be vague.

In this section, we will see how exemplification can also be used to achieve this important function. Let us consider again Example (12), repeated here as (14).

(14) C: *nella guerra nella seconda guerra mondiale cioè una guerra civile erano eh diciamo erano eh # erano coinvolti anche i civili cioè le persone che non non c'entravano niente con la guerra infatti*

A: * *per esempio vittime dei bombardamenti*

C: *ahah infatti*   [LIP Corpus – NC9]

C: 'in the Second World War that is a civil war were eh let's say were eh involved also civilians I mean people who had nothing to do with the war in fact'

A: '**for example, victims of bombing attacks**'

C: 'ahah indeed'

In the previous section, we have seen how exemplification can be used by the addressee to expand the reference provided by the speaker. The addressee is thus directly collaborating to the referring process by giving further important information about the target category. However, we can argue that by providing an actual example of the category, the addressee is also implicitly providing feedback on the entire referring process. Indeed, the fact that the addressee is able to successfully identify an actual member of the target category is an important – albeit implicit – clue of her understanding. Exemplification can thus be seen as a specific type of asserting acceptance, since it can be seen as a form of feedback of the last contribution made by the speaker.

Interestingly, since the addressee (Speaker A) expresses her acceptance through an actual expansion of the reference, the speaker too is compelled to assert his own *acknowledgement* ('indeed') in the end of the exchange. This further shows how providing constant feedback on each step is essential in the referring process.

Looking again at the examples presented in the previous section, we can argue that exemplification can be seen as a collaborative process aimed at reaching agreement and symmetry. This function is mirrored by peculiar structural properties characterizing the internal organization of the interactions. Indeed, we can note a kind of *resonance* pertaining to the architecture of the exchanges, which can be detected in different turns or with respect to the very same speaker's utterance(s). For instance, in (11) B repeats A's example in order to communicate his agreement. In the exchange in (13), we can notice another type of resonance: a symmetrical couple, where feedback markers occupy the very same slot in their respective adjacency pair. The exemplifying progression seems to mirror the principles of dialogic syntax (Du Bois 2014), which looks beyond the single sentence in order to identify relations between pairs of sentences (see also Schiffrin 1987). More specifically, A produces two exemplifying utterances, whereas D utters two different types of

agreement strategies (the former is an explicit agreement marker – namely the holophrastic form *yes* –, whereas the latter is constituted by an act of exemplification). In dialogic syntax terms, the text shows a kind of resonance, understood as the "activation of affinities across utterances" (Du Bois 2014: 360). What is crucial for our analysis is the fact that speakers consider examples as a kind of feedback mechanism *on a par* with more transparent strategies and utter the two different strategies in the same position (what we may define as resonance) with respect to the system of turns. Such syntactic parallel would reveal a functional equivalence between exemplification and explicit agreement markers.

Those discussed above are cases where exemplification has a twofold function, (i) expanding the reference and (ii) asserting acceptance regarding the speaker's reference. Nevertheless, in our data, there are cases where examples are provided by addressees with the only purpose of proving that they have understood the speaker's intended reference and thus of establishing mutual acceptance. Consider the following exchange.

(15)   A: *dove ci sono gli ambulatori*
      B: *sì*
      A: *per fare le visite*
      B: **sì sì ho capito per esempio come l'ottica per gli occhi**
      A: *sì […]*   [LIP Corpus – RC8]
      A: 'where there are clinics'
      B: 'yes'
      A: 'to do the medical examination'
      B: **'yes, yes, I understand for example like the eye optics'**
      A: 'yes […]'

The exchange above shows well all the different stages of the cooperative process. First, A starts the referring process by uttering the noun phrase 'clinics'. B confirms her acceptance ('yes'), allowing the next contribution to occur. Then, A expands his own reference adding a new segment to specify what types of clinics are relevant in the current situation ('those that do medical examinations'). At this point, instead of using again a continuer, B decides to provide a concrete example of the category ('for example like the eye optics'), to better prove that she has understood the reference and has accepted the expansion provided by A.

In this case, it is unlikely that exemplification is used with the purpose of co-construing the reference, as shown in the previous exchanges, since it could be elided without changing the meaning of the exchange or the reference to the category. On the contrary, the actual aim of the addressee is to highlight her understanding and acceptance. This process is made explicit by the fact that B overtly verbalizes her specific intentions, uttering 'yes yes I understand' immediately before

the act of exemplification, and thus implicitly marking the example as a proof of the previous statement. In other words, the exemplar uttered by B functions as an agreement marker, highlighting the acceptance phase. This fact is noteworthy because if examples are not used as a means to reformulate the reference, but only as a means to assert acceptance, it follows that, in exchanges like (15), exemplification performs a purely intersubjective function, which is pivotal to the creation of a mutually accepted common ground.

5.3   Exemplifying to ask for feedback on the reference

In the previous section , we have seen how providing constant feedback is an important part of the collaborative process because it allows the reference to become part of the accepted common ground. However, there may be cases in which the reference is far from being mutually accepted and/or understood, and the addressee needs to ask for further clarification. This can be especially true in the case of ad hoc categories. As already noted in Section 2.1, these categories are built completely on the fly, according to the context. Therefore, the search for a suitable category label may be a tricky task. We have seen that this issue is perceived also by individual speakers, who frequently resort to exemplification in order to expand the reference to the category they have in mind. Nevertheless, it becomes even more pivotal in interactions: the ability to effectively catch the reference to the target category is crucial for the success of the entire communication. For this very reason, the addressee may resort to different strategies to ask for feedback regarding the entire inferential process.

The most obvious strategy is simply asking for clarification directly. However, in some cases, the addressees may want to suggest not only that the reference is not clear to them, but also to provide some clues regarding what they have actually understood. This can be very helpful to the entire process, because it makes explicit the potential misunderstanding and thus functions as a starting point to establish again a common ground.

For instance, the addressee can ask for feedback while suggesting a new formulation of the category, or an expansion of the label provided by the speaker. Hence, the addressee can make explicit how the reference has been interpreted.

(16)   A: *alla fine anche andare in una casa normale cioe'*
      B: *si*
      A: *ci starebbe*
      B: *ma cosa intendi per casa larga casa normale* **cioe' col giardino e cose cosi'**
      A: *eh*
      A: *no vabo' avere anche un living abbastanza abbastanza ampio cioe' cioe' non vorrei che ci fossero degli spazi angusti*   [KIParla Corpus – BOD2007]

A: 'in the end also going in a normal house that is'

B: 'yes'

A: 'it would be nice'

B: 'but what do you mean with large house normal house **that is with a garden or something like that**'

A: 'eh'

A: 'no well also having a quite quite large living that is that is I don't want any narrow space'

In (16), A is describing his favorite type of house. B is not sure about the actual reference and explicitly asks for clarification while proposing a possible reformulation ('house with a garden or something like that'). Basically, B is asking for feedback on what she has understood, that is, that the notion of having a large house expressed by A means having a house with a garden or some sort of large open space. At this point, A can accept or deny (like in this case, i.e. *no vabò* 'no, well') the proposed formulation, ultimately clarifying his reference.

A similar process can be achieved by means of exemplification. In other words, instead of suggesting a possible reformulation, the addressee provides one or more examples of category as it was understood by her, asking if they are indeed part of the target category. Let us consider (17).

(17)   A: *in cucina ricotta va bene tipo negli agnolotti nelle torte sala te pero' rigoro-samente cotta*

        […]

     B: ***e tipo nei cannoli siciliani***

     A: *io non li mangio*   [KIParla Corpus – TOA3004]

     A: 'ricotta is fine, like in agnolotti, in savory pies, but it must be cooked'

        […]

     B: **'and like/for example in Sicilian cannoli'**

     A: 'I don't eat them'

In (17), A, who does not like cheese, tries to describe food containing cheese that he actually likes. The underlying common property is the presence of cooked cheese (in this specific case, ricotta cheese), against raw cheese which he does not appreciate. In order to assess the boundaries of the category, B asks feedback by proposing an example of a very popular pastry containing (raw) ricotta (i.e. *cannoli siciliani* 'Sicilian cannoli') to better understand what is actually included in the target category. The final denial of A confirms that 'containing only cooking cheese' is the relevant property of the target category.

We can argue that there are at least two reasons to use exemplification to perform this checking function instead of providing abstract reformulations. The first and more obvious one concerns the very nature of examples, which represent the

more specific, context-dependent part of the category, that is, its members. Unlike labels and formulations which pertain to a more abstract level, examples are more straightforward and vivid (see the role of exemplification in communication studies, cf. Baesler and Burgoon 1994). Examples provide clear information regarding what type of items are relevant in a specific context and therefore should be included inside the category. This fact may be especially useful when there is no apparent agreement on how to interpret the initial reference provided by the speaker or when the addressee needs to test how clear and strong the category boundaries are (as in (17)).

Another reason for using examples over general formulation is that they prove to be a quite versatile tool to ask for feedback on specific elements or characteristics of the category, depending on the type of example used by the addressee. In particular, the addressee can use (i) good prototypical examples of the category to assess the important characteristics of the category and thus to better understand the overall reference, (ii) problematic examples of the category to understand where to place the boundaries of the category.

In the first case, we can find situations where the speaker and the addressee have different ideas on how to construe the reference in a given context, that is, the addressee does not understand what type of category members are relevant in the current situation. Let us consider the following example.

(18) C: *cercare di capire in quali situazioni in quali contesti quindi piu' che situazioni comunicative in senso lato in quali contesti*
      B: *okay*
      C: *mh*
      B: **ma tipo per strada o a scuola o**
      C: *eh si' o anche il diminutivo all'interno di una domanda*

            [KIParla corpus – BOA1003]

      C: 'to figure out in which situations in which contexts therefore rather than in which communicative situations in a loose way in which contexts'
      B: 'okay'
      C: 'mh'
      B: **'but like/for example in the street or at school or'**
      C: 'eh yes or also the [use of] diminutive inside an interrogative [clause]'

In (18), C makes an explicit reference to a sector-specific category using the label *situazioni/contesti comunicativi* 'communicative contexts'. B understands the literal meaning of the label used by C, but she does not understand the reference in the specific context of linguistics studies. Instead of explicitly indicating her confusion, B asks for feedback by providing some good examples of the category she can construe through the mentioned label (i.e. 'in the street', 'at the school'). In this

case, the use of examples makes clear to C that his addressee is uncertain on how to contextualize the target category. Thus, C politely refuses by giving in turn an actual exemplar of his target category (i.e. 'interrogative (clause)') to successfully direct the inferential process of the addressee.

Here we can see a function similar to that described in Section 4 and called 'contextualizing the reference'. In the framework of real interactions, we can see that this function is also pivotal to negotiate the contextualization of the reference, making clear potential misunderstandings without being too explicit. This fact is quite important in those types of interactions where stating bluntly that the reference is not clear may be problematic or where repairing the speaker's utterance is not considered appropriate (cf. Schegloff et al. 1977). For instance, (18) is taken from a conversation between a student and a university professor: a type of situation in which an explicit statement of misunderstanding or repairing the formulation of the speaker can be perceived as rude or embarrassing. The mediation of examples allows to repair the reference while avoiding an explicit disagreement and thus protecting the speakers' faces (cf. Goffman 1959 on the notion of *face*, and Brown and Levinson 1987 regarding the politeness theory) of all participants.

Good prototypical examples can also be used to ask for disambiguation of problematic or unclear formulations. In other words, the addressee does not understand the reference the way the speaker formulates it. Instead of repairing it, she suggests some potential examples of the reference asking for feedback. Consider (19).

(19) A: *ok allora abito a*
     B: *okay*
     A: *in una in una casa a due piani*
     A: *sopra abito io con la mia famiglia e sotto abita mia zia*
     B: *okay*
     A: *e mh*
     B: *e' una casa grande*
     A: *si' e' una casa abbastanza grande*
     A: *ma e' **tipo un condominio oppure una villetta** che tipo di casa*
     A: *mmh no siamo io e mia zia*
     B: *okay quindi sono due appartamenti*
     A: *si mh diciamo intorno agli ottanta metri quadrati*   [KIParla – BOD2007]
     A: 'OK well I live in'
     B: 'OK'
     A: 'in a two-floor house'
     A: 'I live upstairs with my family and my aunt lives downstairs'
     B: 'OK'
     A: 'and mh'
     B: 'is it a big house?'

A: 'yeah it is quite big'
B: 'but is it **like a condo or a cottage** which kind of house'
A: 'mh no it is only me and my aunt'
B: 'OK so there are two independent flats'
A: 'yeah mh it's around eighty square metres'

In (19), A is describing the type of house he lives in. He construes the reference using a very general formulation (i.e. 'big house'). Although B seems to have a vague idea of what he likely means, she still asks for feedback by creating a list of big houses (i.e. 'condo' and 'cottage') followed by an explicit request (i.e. *che tipo di casa* 'which kind of house?') to clarify the type of house to which A is referring. This list of examples ultimately allows A to clarify his reference, highlighting the important feature of the target category, that is, the fact that his house is not as big as an entire condo or a cottage.

Beyond suggesting prototypical good examples of the category, the addressee can ask for feedback by providing *problematic* examples. With the label 'problematic' we mean items that for different reasons (e.g. cultural reasons) may have a controversial status inside a specific category. Asking about their membership allows the addressee to assess the boundaries of the category (cf. Croft & Cruse 2004: 93-ff.). Consider (20).

(20)   B: *a tu non piace il forma a te non piace il form*
       A: *io solo mozzarella*
            […]
       B: *ma **tipo ricotta zero***
       A: *nelle torte salate*   [KIParla Corpus – TOA3004]
       B: 'you do not like chee[se] you do not like chee[se]'
       A: 'I only [like] mozzarella'
            […]
       B: 'but **for example ricotta, nothing?**'
       A: 'in savory pies'

The occurrence in (20) is taken from the same interaction of (18). As already described, A does not like raw cheese. B tries to assess the boundaries of this category by asking about the membership of a cheese with a peculiar status in Italian cuisine, that is, *ricotta* 'ricotta', which is frequently found in traditional and very popular Italian recipes (e.g. it is a prominent ingredient in Sicilian cannoli, as noted for example in (18)). The underlying assumption is that even those people who do not like cheese frequently make an exception to ricotta or mozzarella. Therefore, by using an example with such a peculiar status, B wants to know to what extent (and how firmly) the category boundaries are located and better understand A's reference.

6. Concluding remarks: Exemplification and the online creation of reference

In this paper, we have shown how the reference to conceptual categories is construed online through a collaborative process, and not just through an operation performed by a single speaker. The aim of this paper was to investigate the use of exemplification in such a process, showing that its role goes beyond the simple act of reformulating or refashioning a previous explicit noun phrase.

Using the collaborative model by Clark and Wilkes-Gibbs (1986) as our theoretical framework, we have analysed instances of exemplification in real occurrences of spoken Italian, using a corpus-based approach. According to this model, both speaker and addressee make an effort in signalling and guaranteeing that all participants have correctly understood a given target reference. For this reason, special attention has been devoted to exchanges where exemplification is scattered across different turns and different speakers, to better investigate different potential functions also at the intersubjective level.

First, we have observed how examples are used by the speaker to introduce a referent and the different ways in which exemplification can expand a previously mentioned referent to provide further important information about the target category. Then, we moved a step forward, noticing that these functions can also be performed by the addressee. In particular, we observed that also the addressee can directly collaborate to the referring process by providing one or more examples in order to highlight crucial information about the target category, even when such an operation is not explicitly required. We underlined that dialogic conditions play a crucial role in such a collaborative operation of reference construction, due to semiotic and semantic properties of spoken dialogue (e.g. temporal constraints, co-presence between interlocutors, context-anchoring).

Beyond being used to build the reference itself, exemplification can also play an important part in the underlying construction of a mutually accepted common ground. More specifically, examples can be used by the addressee with the primary goal of providing feedback on the referring process started by the speaker, that is, as a means to communicate that understanding has been reached and thus allow the conversation to move forward. Exemplification can thus be seen as a specific type of asserting acceptance process, since, in these cases, examples are not used as tools to specify the reference, but only as a means to express agreement. It follows that examples can perform a genuine intersubjective function aimed at ensuring symmetry and cooperation on social ground.

Moreover, exemplification can also be used by the addressee to explicitly ask for feedback about her interpretation. Whenever the reference is not clear or problematic, instead of suggesting a possible reformulation, the addressee can provide one

or more examples of category as it was understood by her, asking if such examples actually belong to the target category. Prototypical examples can be used to ask for disambiguation of problematic or unclear formulations. However, the addressee can also ask for feedback by providing problematic examples in order to set the boundaries of the target reference. More generally, we argued that the act of exemplification in dialogic interactions is crucial to negotiate the reference, by resolving possible misunderstanding without sounding too blunt.

In our paper, we focused our attention on exemplification due to the fact that despite being a very important strategy to make reference to categories, also in dialogic interactions, its role has not been addressed systematically by the existent literature. However, it is important to note that exemplification is just one strategy among others. As noted for several occurrences throughout the paper, we can observe that exemplifying constructions frequently occur within larger and more complex processes of reference elaboration, where both speakers and addressees participate by creating more general formulations, exemplifying, reformulating again and so forth. Consider (21).

(21)   B: *e quindi abbiamo diciamo organizzato Buona Idea fornendo tutta una gamma di servizi che vanno dalla dall' organizzazione del tempo libero con eh # giochi animazione sia per un target di clientela adulta*

A: ** mh*

B: *che bambini e eh una serie di servizi che fossero così una risposta anche a eh non so ecco **una serie di attività che la le persone che lavorano non la possono sviluppare quindi non so dall' andare a fare le file alla posta eh piuttosto che avere una baby-sitter all' ultimo momento che non si trova piuttosto che non so organizzare ecco una festa per una mamma che lavora al suo bambino***

A: *quindi insomma vi rivolgete non solo al tempo libero ma all' **uso del tempo in genere*** [LIP Corpus – RE8]

B: 'so we have, let's say, organized "Buona Idea" providing a whole range of **services ranging from the organization of free time with eh games animation for both adult clients**'

A: '* hm'

B: 'and children e eh a range of services that can be an answer also to eh I don't know **a series of activities that people who work cannot undertake, so for example, from staying in a cue at the Post Office eh, rather than/ for example having a last-minute babysitter you can't find rather than/ for example I don't know organizing yeah, a party on behalf of a working mother for his child**'

A: 'so you deal not only with leisure time but with **the use of time in general**'

In (21), different strategies are used to construe the reference and to disambiguate possible misunderstandings. The speaker firstly reformulates the category label because it leaves out some activities that are not strictly related to leisure time (i.e. 'activities that people who work cannot undertake'). After such a reformulation, concrete exemplars (i.e. staying in a cue at the Post Office, having a last-minute babysitter) are employed to actualize the category through a top-down strategy. In the end, the addressee concludes the process by employing a bottom-up mechanism through which the abstract concept ('the use of time in general') is uttered with the final purpose of verifying the top-down approach carried out by the speaker, and to highlight the pivotal property of the category. Again, reference construction appears as an incremental, additive and tentative process which is cooperatively performed.

This apparent redundancy is not without purpose. As already noted, ad hoc categories are volatile and created online whenever they are needed. Making reference to them in a way that resonates with the other speech participants is not an easy task. More importantly, also the creation of these categories at the cognitive level may require different approaches and strategies that are likely mirrored at the speech level. Although a psycholinguistics analysis is beyond the scope of this paper, we can still observe how exemplification is a fundamental part of these processes. While some cognitive studies have used the elicitation of exemplars as a way to analyse categorization processes, the fact that in spontaneous speech speakers frequently rely on exemplification can reveal something about the way these ad hoc categories are built. Specifically, it suggests that the creation of umbrella terms is not always the best way to approach these categories. On the contrary, the identification of one or more exemplars to assess specific internal details of the category (e.g. the prototype, the boundaries) can be a good solution in many cases.

Acknowledgements

References

Auer, Peter. 2009. On-line syntax. Thoughts on the temporality of spoken language. *Language Sciences* 31(1): 1–13. https://doi.org/10.1016/j.langsci.2007.10.004

Baesler, James E. & Burgoon, Judee K. 1994. The temporal effects of story and statistical evidence on belief change. *Communication Research* 21: 582–602.
https://doi.org/10.1177/009365094021005002

Barotto, Alessandra. 2021. *Exemplification and Categorization. The Case of Japanese*. Berlin: Mouton de Gruyter.

Barotto, Alessandra. 2018. The role of exemplification in the construction of categories: The case of Japanese. In *Linguistic Strategies for the Construction of Ad Hoc Categories: Synchronic and Diachronic Perspectives*, Caterina Mauri & Andrea Sansò (eds). Special issue of *Folia Linguistica* 52(s39–1): 37–68. https://doi.org/10.1515/flih-2018-0002

Barotto, Alessandra & Mauri, Caterina. 2018. Constructing lists to construct categories. *Italian Journal of Linguistics* 30: 95–134.

Barsalou, Lawrence W. 1983. Ad hoc categories. *Memory and Cognition* 11(3): 211–227.
https://doi.org/10.3758/BF03196968

Bazzanella, Carla. 1995. I segnali discorsivi. In *Grande Grammatica italiana di consultazione. III Tipi di frase, deissi, formazione delle parole*, Lorenzo Renzi, Giampaolo Salvi & Anna Cardinaletti (eds), 225–257. Bologna: Il Mulino.

Brinton, Laurel J. 1996. *Pragmatic Markers in English: Grammaticalization and Discourse Functions*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110907582

Brown, Penelope & Levinson, Stephen C. 1987. *Politeness: Some Universals in Language Usage*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511813085

Bruner, Jerome S., Goodnow, Jacqueline J. & Austin, George A. 1956. *A study of thinking*. New York NY: Wiley.

Caffi, Claudia. 2007. *Mitigation*. Amsterdam: Elsevier.

Calaresu, Emilia M. 2016. Dialogicità e grammatica. In *Dinamiche dell'interazione: Testo, dialogo, applicazioni educative*, Cecilia Andorno & Roberta Grassi (eds), 13–27. Milano: Collana "Studi AItLA".

Calaresu, Emilia M. 2018. Grammaticalizzazioni polifoniche o "verticali" e sintassi dialogica. Dagli enunciati-eco ai temi sospesi: l'infinito anteposto in strutture del tipo "mangiare, mangio". In *Strutture e dinamismi della variazione e del cambiamento linguistico. Atti del Convegno DIA III, Napoli, 24–27 novembre 2014*, Paolo Greco, Cesarina Vecchia & Rosanna Sornicola (eds), 505–521. Napoli: Accademia di Archeologia, Lettere e Belle Arti, Società Nazionale di Scienze, Lettere ed Arti in Napoli: Giannini editore.

Channell, Joanna. 1994. *Vague Language*. Oxford: OUP.

Clark, Herbert H. 1996. *Using language*. Cambridge: CUP.
https://doi.org/10.1017/CBO9780511620539

Clark, Herbert H. & Brennan, Susan E. 1991. Grounding in communication. In *Perspectives on socially shared cognition*, Lauren B. Resnick, John M. Levine, Stephanie D. Teasley (eds), 127–149. Washington, DC: American Psychological Association.
https://doi.org/10.1037/10096-006

Clark, Herbert H. & Wilkes-Gibbs, Deanna. 1986. Referring as a collaborative process. *Cognition* 22(1): 1–39. https://doi.org/10.1016/0010-0277(86)90010-7

Clark, Herbert H. & Schaefer, Edward F. 1989. Contributing to discourse. *Cognitive Science* 13(2): 259–294. https://doi.org/10.1207/s15516709cog1302_7

Croft, William & Cruse, Alan D. 2004. *Cognitive Linguistics*. Cambridge: CUP.
  https://doi.org/10.1017/CBO9780511803864

Cruse, Alan D. 1990. Prototype theory and lexical semantics. In *Meaning and Prototypes: Studies in Linguistic Categorization*, Savas L. Tsohatzidis (ed.), 382–402. London: Routledge.

Degand, Liesbeth. 2014. 'So very fast then'. Discourse markers at left and right periphery in spoken French. In *Discourse Functions at the Left and Right Periphery: Crosslinguistic Investigations of Language Use and Language Change*, Kate Beeching & Ulrich Detges (eds), 151–178. Leiden: Brill.

De Mauro, Tullio, Mancini, Federico, Vedovelli, Massimo & Voghera, Miriam. 1993. *Lessico di frequenza dell'italiano parlato*. Milan: Etaslibri.

Du Bois, John W. 2014. Towards a dialogic syntax. *Cognitive Linguistics* 25(3): 359–410.
  https://doi.org/10.1515/cog-2014-0024

Fischer, Kerstin. 2006. *Approaches to Discourse Particles*. Amsterdam: Elsevier.

Fraser, Bruce. 1996. Pragmatic markers. *Pragmatics* 6(2): 167–190.
  https://doi.org/10.1075/prag.6.2.03fra

Ghezzi, Chiara. 2013. Vagueness Markers in Contemporary Italian: Intergenerational Variation and Pragmatic Change. PhD dissertation, Bergamo University.

Goffman, Erving. 1959. *The Presentation of Self in Everyday Life*. New York NY: Doubleday.

Goria, Eugenio & Mauri, Caterina. 2018. Il corpus KIParla: Una nuova risorsa per lo studio dell'italiano parlato. In *CLUB Working Papers in Linguistics*, Vol. 2, Francesca Masini & Fabio Tamburini (eds), 96–116. Bologna: Circolo Linguistico dell'Università di Bologna (CLUB).

Grice, Paul. 1975. Logic and conversation. In *Syntax and Semantics*, Vol. 3: *Speech Acts*, Peter Cole & Jerry L. Morgan (eds), 41–58. New York NY: Academic Press.

Halliday, Michael A. K. & Hasan, Ruqaia. 1976. *Cohesion in English*. London: Longman.

Haselow, Alexander. 2016. A processual view on grammar: Macrogrammar and the final field in spoken syntax. *Language Sciences* (54): 77–101. https://doi.org/10.1016/j.langsci.2015.12.001

Heine, Bernd. 2013. On discourse markers: Grammaticalization, pragmaticalization, or something else? *Linguistics* 51(6): 1205–1247. https://doi.org/10.1515/ling-2013-0048

Hobbs, Jerry R. 1979. Coherence and coreference. *Cognitive Science* 3(1): 67–90.
  https://doi.org/10.1207/s15516709cog0301_4

Hobbs, Jerry R. 1985. *On the Coherence and Structure of Discourse*. Stanford CA: CSLI.

Hovy, Eduard & Maier, Elisabeth. 1994. Parsimonious and profligate: How many and which discourse structure relations? Ms. <https://bit.ly/2UCA7Xf> (4 March 2019).

Hyland, Ken. 2007. Applying a gloss: Exemplifying and reformulating in academia discourse. *Applied Linguistics* 28 (2): 266–285. https://doi.org/10.1093/applin/amm011

Levelt, Willem J. M. 1983. Monitoring and self-repair in speech. *Cognition* 14: 41–104.
  https://doi.org/10.1016/0010-0277(83)90026-4

Lo Baido, Maria Cristina. 2018. Categorization via exemplification: Evidence from Italian. In *Linguistic Strategies for the Construction of Ad Hoc Categories: Synchronic and Diachronic Perspectives*, Caterina Mauri & Andrea Sansò (eds). Special issue of *Folia Linguistica* 52(s39–1): 69–95. https://doi.org/10.1515/flih-2018-0007

Longacre, Robert E. 1983. *The Grammar of Discourse*. New York NY: Plenum.

Mann, William C. & Thompson, Sandra A. 1998. Rhetorical structure theory: Toward a functional theory of text organization. *Text-Interdisciplinary Journal for the Study of Discourse* 8(3): 243–281.

Manzotti, Emilio. 1998. L'esempio. Natura, definizioni, problemi. *Cuadernos de Filología Italiana* 5: 99–123.

Masini, Francesca & Pietrandrea, Paola. 2010. Magari. *Cognitive Linguistics* 21: 75–121.
  https://doi.org/10.1515/cogl.2010.003

Mauri, Caterina. 2017. Building and interpreting ad hoc categories: A linguistic analysis. In *Formal Models in the Study of Language*, Joanna Blochowiak, Cristina Grisot, Stephanie Durrleman-Tame & Cristopher Laenzlinger (eds), 297–326. Berlin: Springer. https://doi.org/10.1007/978-3-319-48832-5_16

Molinelli, Piera. 2014. "Sai cosa ti dico? Non lo so, se non me lo dici". Sapere come segnale pragmatico nell'italiano parlato contemporaneo. In *Dall'architettura della lingua italiana all'architettura linguistica dell'Italia. Saggi in omaggio a Heidi Siller-Runggaldier*, Paul Danler & Christine Konecny (eds), 487–501. Frankfurt: Peter Lang.

Overstreet, Maryann. 1999. *Whales, Candlelight, and Stuff Like That: General Extenders in English Discourse*. Oxford: OUP.

Prandi, Michele. 2006. *Le regole e le scelte. Introduzione alla grammatica italiana*. UTET Universtà.

Redeker, Gisela. 1990. Ideational and pragmatic markers of discourse structure. *Journal of Pragmatics* 14: 367–381. https://doi.org/10.1016/0378-2166(90)90095-U

Rodríguez Abruñeiras, Paula. 2015. Exemplifying Markers in English: Synchronic and Diachronic Considerations. PhD dissertation, Santiago de Compostela University.

Rosch, Eleanor. 1973. Natural categories. *Cognitive Psychology* 4: 328–50. https://doi.org/10.1016/0010-0285(73)90017-0

Rosch, Eleanor. 1977. Human Categorization. In *Advances in Cross-Cultural Psychology*, Neil Warren (ed.), 1–72. London: Academic Press.

Rosch, Eleanor. 1978. Principles of categorization. In *Cognition and Categorization*, Eleanor Rosch & Barbara B. Lloyd (eds), 27–48. Hillsdale NJ: Lawrence Erlbaum Associates.

Schegloff, Emanuel A., Jefferson, Gail & Sacks, Harvey. 1977. The preference for self-correction in the organization of repair in conversation. *Language* 53: 361–382. https://doi.org/10.1353/lan.1977.0041

Schiffrin, Deborah. 1987. *Discourse Markers*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511611841

Schober, Michael F. & Clark, Herbert H. 1989. Understanding by addressees and overhearers. *Cognitive Psychology* 21(2): 211–232. https://doi.org/10.1016/0010-0285(89)90008-X

Schneider, Stefan. 1999. *Il congiuntivo tra modalità e subordinazione. Uno studio sull'italiano parlato*. Roma: Carocci.

Schneider, Stefan. 2007. Reduced parenthetical clauses in Romance languages – A pragmatic typology. In *Parentheticals* [Linguistik Aktuell/Linguistics Today 106], Nicole Dehé & Yordanka Kavalova (eds), 237–258. Amsterdam: John Benjamins. https://doi.org/10.1075/la.106.13sch

Smith, Linda B. & Samuelson, Larissa K. 1997. Perceiving and remembering: Category stability, variability and development. In *Knowledge, Concepts and Categories*, Koen Lamberts & David Shanks (eds), 161–195. Hove: Psychology Press.

Voghera, Miriam. 2013. A case study on the relationship between grammatical change and synchronic variation: The emergence of tipo[-N] in Italian. In *Synchrony and Diachrony. A Dynamic Interface* [Studies in Language Companion Series 133], Anna Giacalone Ramat, Caterina Mauri & Piera Molinelli (eds), 283–312. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.133.12vog

Wang, Yu-Fang, Tsai Pi-Hua, Goodman, David & Lin, Meng-Ying. 2010. Agreement, acknowledgment, and alignment: The discourse-pragmatic functions of hao and dui in Taiwan Mandarin conversation. *Discourse Studies* 12(2): 241–267. https://doi.org/10.1177/1461445609346922

Zillmann, Dolf. 2002. Exemplification theory of media influence. In *Media Effects: Advances in Theory and Research*, Jennings Bryant & Dolf Zillmann (eds), 19–41. London: Routledge.

The on-line construction of meaning in Mandarin Chinese

Focus on relative clauses

Giorgio Francesco Arcodia
Ca' Foscari University of Venice

Mandarin Chinese employs a fairly wide range of constructions to encode categories, and specifically *ad hoc* categories. These include, for instance, a general extender as 等等 *děng(děng)* 'etc., and so on', non-exhaustive connectives as 啊 *ā…*啊 *ā* (see Zhang 2008), exemplifying constructions, and so on. In this paper, I focus on the use of a specific strategy of on-line category construction in Chinese, namely postnominal relative clauses (RCs). Postnominal RCs are particularly interesting since in Mandarin Chinese, as in nearly every Sinitic language, normally all modifiers (including RCs) appear before the head noun; it has been proposed that postnominal RCs are always added as afterthoughts, to resolve a potentially ambiguous reference, or just to narrow down the scope of predication (see Wang & Wu 2020). I conduct a manual search of postnominal RCs in excerpts of transcribed spoken dialogue from the National Broadcast Media Language Corpus, and I propose an analysis of the use of postnominal RCs as devices for on-line categorization, focussing on their interaction with other strategies for category-building. I also discuss the pragmatic and functional correlates of the use of postnominal RCs, as opposed to canonical, prenominal RCs.

Keywords: Chinese, categorization, relative clause, headedness, word order

1. Introduction

Mandarin Chinese employs a fairly wide range of constructions to encode categories, and specifically *ad hoc* categories. Some of the most common are, for instance, the general extender 等(等) *děng(děng)* 'etc., and so on', and the non-exhaustive connectives 啊 *ā…*啊 *ā* (Zhang 2008), exemplifying constructions, and so on. As is

the case for many other languages, different strategies for the on-line construction of meaning and categorization may coexist in the same utterance, as in the following example (from the National Broadcast Media Language Corpus):[1]

(1) 眼前的北方高温, 南方暴雨, 前一段的
     yǎnqián de běifāng gāo-wēn,   nánfāng bàoyǔ,   qián
     at.present SP north   high-temperature south   cloudburst before
     yīduàn de
     time   SP
     西南干旱等等, 极端天气接二连三
     *xī-nán   gānhàn **děngděng, jíduān tiānqì   jiē'èrliánsān***
     west-south drought **etc.   extreme weather** one.after.another
     'At present, **extreme weather events like** the high temperatures in the north, the cloudbursts in the south, and the droughts in the southwest, just happen one after another'

In Example (1), a list of exemplars of the category [EXTREME WEATHER EVENTS] is first presented and followed by the general extender 等等 *děng*(*děng*); after a pause, the category is explicitly named, for further identification of the set evoked by the exemplars. Such strategies, aimed at clarifying the reference which the speaker is trying to establish on-line, are commonly found in spontaneous speech (and even written production) in many languages (Mauri & Sansò 2018).

In this paper, I will focus on a specific construction type, namely postnominal relative clauses (henceforth: RCs), and I argue that they may be used as a device for on-line category construction in Chinese. The canonical position for RCs in Mandarin Chinese, as well as in virtually all Sinitic languages, is actually *pre*nominal, and the combination of VO and RelN order is apparently attested almost only in Sinitic (see the data in Dryer 2013). However, postnominal RCs are also attested in Mandarin and in some other Chinese dialects (see Arcodia 2017; Wang & Wu 2020), where they are always a less frequent and more restricted construction (as opposed to canonical prenominal RCs). As hinted at above, I suggest that postnominal RCs in Mandarin Chinese may be analyzed as a syntactic strategy for on-line category construction, by contributing to identify the category referred to by the head noun. See the following example (from Fang 2004: 151):

---

1. URL: <http://ling.cuc.edu.cn/RawPub/> (last access: 4 March 2019).

(2) 你比如说你跟着那种水平不高的英语老师,

   nǐ   bǐrú-shuō   nǐ   gēn-zhe   nà   zhǒng shuǐpíng bù-gāo   de
   2SG for.instance-say 2SG follow-DUR that type   level   not-high SP

   Yīngyǔ lǎoshī
   English teacher

   他根本不知道那个纯正的英语发音,

   tā   gēnběn bù   zhīdào nà   ge chúnzhèng de Yīngyǔ fāyīn
   3SG at.all   not know   that CL pure   SP English pronunciation

   他英语语法也不怎么样, 你就全完了

   tā   Yīngyǔ yǔfǎ   yě   bù-zěnmeyàng, nǐ   jiù   quán   wán   le
   3SG English grammar too not-so.great   2SG then entirely finish PERF

   'If, say, you follow that kind of English teachers whose level is not so high, who
   don't know pure English pronunciation, and whose grammar is bad, then you're
   lost.'

In Example (2), the set of 'English teachers' is modified by the phrase 水平不高
的 *shuǐpíng bù-gāo de* 'whose level is not so high', and then followed by two rela-
tive(-like?) clauses, further specifying the set denoted by 'English teachers'. Here,
I argue, while the (ad hoc?) category label is explicitly named ('not-so-good English
teachers'), the function (or one of the functions) of the postnominal RCs is that of
narrowing down the scope of reference, i.e. "restricting the borders of the category"
(Mauri 2017: 323) by providing more "clues" (Mauri 2017; Barotto & Mauri 2018)
for the identification of the category represented by the head noun. My analysis is
based on data extracted with a manual search of postnominal RCs in excerpts of
transcribed spoken dialogue drawn from the National Broadcast Media Language
Corpus, maintained by the Communication University of China (中国传媒大学
Zhōngguó Chuánméi Dàxué).

   This paper is organized as follows. Firstly, I will introduce the theoretical back-
ground of my research, providing a working definition of RC in the context of the
Sinitic family, as well as a summary of the most relevant findings on this topic in
previous research, and I will argue that they fit within the domain of on-line cate-
gory building (§ 2). Then, I will briefly illustrate the methodological aspects of my
research (§ 3). I will then propose my generalizations on the data, focussing on the
interaction of postnominal RCs with other strategies for category-building, and on
the pragmatic and functional correlates of the use of postnominal RCs, as opposed
to canonical, prenominal RCs (§ 4). Lastly, I will summarize the main conclusions
of my survey and provide some hints for further research (§ 5).

2. Theoretical background

The research presented in this paper rests on the assumptions that in Mandarin Chinese there is a construction type which may be defined as RC, and that RCs may be placed after the head noun. Both ideas are, however, not uncontroversial, and require some preliminary discussion. Furthermore, their status as devices for on-line category building must be argued for in the framework of the published research on the subject. These will be the topics of the following sections.

2.1 The status of postnominal RCs in Chinese

RCs are a cross-linguistically common type of subordinate clause, and they are often used as an object of comparison in typological research (see e.g. Comrie & Kuteva 2013; Wiechmann 2015). They are generally defined as "clause-sized modifiers" of the noun (Givón 2001: 175), or as "subordinate clauses with a semantic pivot which they share with the matrix" (i.e. the head noun; Wu 2011: 570); from a semantic point of view, they may also be defined as clauses "narrowing the potential reference of a referring expression by restricting the reference to those referents of which a particular proposition is true" (Comrie & Kuteva 2013). It clearly appears that, from the semantic point of view, RCs are particularly suitable for the function of 'categorization clues', given that they basically restrict the reference of the head. I will get back to this in the next section.

If we focus on the structural aspects of the definition, the crucial difference between RCs and other types of modifiers lies in their clausal status (Arcodia 2017; see also Wang & Wu forthcoming).[2] However, in many (if not most) descriptive studies on Sinitic languages, clause-sized modifiers of the NP are often lumped together with adjectives, genitives, and other adnominal modifiers, in a macro-category termed 'modifiers' (Chin. 定语 *dìngyǔ* or 修饰语 *xiūshìyǔ*; see Liu 2005: 3). This is motivated by the formal identity of all constructions of adnominal modification in most Sinitic languages, including Mandarin. Compare:

(3) 张三的词典
 Zhāngsān de cídiǎn
 Zhangsan SP dictionary
 'Zhangsan's dictionary'

---

2. Here, by 'clausal status' we mean that RCs, by definition, must contain a verbal predicate (see Wang & Wu 2020). I would like to thank an anonymous reviewer for helping me to clarify this point.

(4)　说英语的同学
　　　shuō　Yīngyǔ　de　tóngxué
　　　speak　English　SP　classmate
　　　'the classmate(s) who speak/s English'

The possessive modifier in (3) and the clause-sized modifier (RC) in (4) both appear in the same prenominal position, and they are followed by the same marker (的 *de*), generally termed a 'structural particle' (结构助词 *jiégòu zhùcí*) in Chinese linguistics.

Moreover, RCs are sometimes conflated with nominalizations (see e.g. Li & Thompson 1981: 579–580); the fact that Sinitic languages generally use the same marker for relativization (i.e. adnominal modification) and nominalization further contributes to the blurring of the distinction, given that nominalizing constructions may be analyzed as headless relative clauses (Yap & Matthews 2008; exx. adapted from Li & Thompson 1981: 580):

(5)　a.　种水果的农民
　　　　　zhòng shuǐguǒ　de　nóngmín
　　　　　grow　fruit　　SP　farmer
　　　　　'farmers who grow fruit'
　　　b.　种水果的
　　　　　zhòng shuǐguǒ　de
　　　　　grow　fruit　　SP
　　　　　'fruit growers, those who grow fruit'

The clause in (5b) may be analyzed either as a nominalized construction, or as a headless version of the relative clause in (5a).[3]

In point of fact, the application of the notion of RC to Sinitic (as well as to many other languages of East Asia) has been called into question by Comrie (2003: 29), who prefers the label "general noun-modifying constructions" to indicate just any clause-sized modifier attached to the NP, without any syntactic operation as movement or gap-filling, and connected to the head noun by semantic-pragmatic relations. A case in point is the following Cantonese example (Chan, Matthews & Yip 2011: 219):

---

3. While this is the overwhelmingly dominant model throughout Sinitic, note that, as pointed out in Arcodia (2017: 38–39), the markers of adnominal modification (incl. RCs) and nominalization may not coincide e.g. in some Wu, Xiang, and also Yue dialects (Chao 1956[1928]; Yue 1995; Wu 2005); and, actually, in some cases, even different types of adnominal modification employ different strategies (e.g. in Tiantai Wu; Dai 2006: 99–100).

(6)   返學嗰對鞋[4]
    fāan   hohk   gó   deui   hàaih
    return school that pair shoe
    'the shoes for going to school'

In (6), the head noun 鞋 *hàaih* 'shoe(s)' has no argument relation with the verb 返 *fāan*; the association between the two appears to be based, again, on a semantic-pragmatic relation. According to Chan, Matthews and Yip (2011: 219), distinguishing this type of noun-modifying constructions with a clausal modifier from 'true' RCs (i.e. with a syntactic relation between the subordinate clause and the head NP) may prove "difficult, if not impossible" (on the inadequacy of the notion of RC for Sinitic, see also LaPolla 2017; Matthews & Yip 2017).

Despite the uncertainties as to the status of RCs in Chinese briefly discussed above, for the purposes of the present study I may propose a working definition of RC in Mandarin. In his typological survey, Arcodia (2017: 39) employs a very broad definition of RC as "a clause-sized (non-argumental) modifier built around a verb and headed", to the exclusion of adjectival predicates; the latter specification is relevant since verbs and adjectives are often conflated for Chinese (see Paul 2015). In Wang & Wu (2020), we find a more restrictive definition of RC. According to them, a RC, besides containing a verb, must also contain an unpronounced gap that is co-referential with its head noun, which in turn must have an argumental relation with the verb in the RC; this definition, thus, excludes adjunct RCs and noun complements. Also, Wang & Wu further argue that dependent status is a requisite for RC status; as dependent clauses, they should not be acceptable as standalone sentences. Since my focus here is on the use of RCs as devices for categorization, and hence the semantic-pragmatic aspects of the construction are arguably more relevant than the syntactic aspects, I will adopt Arcodia's (2017) definition, which is more inclusive for this construction type.

What about postnominal RCs? The amount of research on this construction type in Chinese is quite limited (see e.g. Dong 2003; Fang 2004; Liu 2008; see also Luke 1998 on Cantonese), although there is one very recent comprehensive study of the phenomenon (Wang & Wu 2020) in which an attempt has been made to identify the factors influencing the use of this non-canonical construction, including semantic, pragmatic and processing factors. There are actually somewhat different constructions that have been treated as postnominal RCs in the literature, and opinions differ as to their status. The first and least controversial construction type which has been analyzed as a postnominal RC is exemplified below (Dong 2003: 121):

---

4. Here, I follow the source in using traditional Chinese characters.

(7) 司法工作人员私放罪犯的 […]
*sīfǎ   gōngzuò-rényuán **sī-fàng   zuìfàn   de***
judicature working-personnel illicit-release criminal SP
'judiciary staff members **who illicitly release a criminal** […]'

The clause 私放罪犯的 *sī-fàng zuìfàn de* is structurally analogous to a prenominal RC; in point of fact, (7) would be perfectly fine (and, arguably, more natural) if the modifying clause were placed before the head NP.

This type of postnominal modifiers are said to be a characteristic feature of legal texts (Dong 2003), although they may also be found in colloquial Mandarin (Example from Zhao 2009: 87):

(8) 女作家撰写历史小说的, 孟瑶算是数一数二的了
*nǚ-zuòjiā   **zhuànxié lìshǐ   xiǎoshuō de** Mèng Yáo suànshì*
female-author write   history novel   SP Meng Yao consider.to.be.
shǔyīshǔ'èr   de le
one.of.the.best SP PERF
'Among female authors **who write historical novels**, Meng Yao is regarded as one of the best'

Zhao (2009) believes that these are not true RCs, but rather sentences with a subject-predicate construction as predicates (Chin. 主谓谓语句 *zhǔ-wèi wèiyǔ jù*). The arguments she provides in support of her position are that there can be a pause between the NP and the modifying clause, and a particle (as e.g. 啊 *a*, here marking a topic) can be added before the pause; also, conjunctions and adverbial modifiers can be added to the modifying clause. However, Arcodia (2017: 43) counterargues that a pause before or after a dependent clause is not unusual, cross-linguistically (as e.g. for English non-restrictive RCs); in point of fact, Liu (2008: 13) points out that, in Mandarin Chinese, a pause between head and modifier is odd only when the latter is placed *before* the head NP. Wang & Wu (2020) highlight that a pause between the relativizer 的 *de* and the postnominal RC is the norm in spoken Mandarin, and this is best explained, in their opinion, by the fact that postnominal RCs are mostly added as afterthoughts, motivated by the desire of the speaker to avoid referential ambiguity (a common function for RCs; Fretheim 1995; Podlesskaya 2011, qtd. in Wang & Wu 2020); this fits very well with my analysis of postnominal RCs as devices for on-line category building, and I will get back to this below. Furthermore, Arcodia (2017: 43–44) argues that the addition of adverbs to the modifying clause is acceptable in Mandarin also for prenominal RCs, and thus Zhao's (2009) second argument does not appear to be significant.

Another type of postnominal clause-sized modifier which has been described as an RC in the literature is shown in (9) (Fang 2004: 151):

(9)   你们班里你万一有谁吸毒的, 谁这个瞎搞的,
    *nǐ-men bān-li   nǐ   wànyī   yǒu   **shéi xī-dú   de shéi zhè-ge***
    2SG-PL class-loc 2SG supposing there.be s.o. do-drug SP s.o. this-CL
    xiāgǎo   de
    mess.around SP
    谁携抢的, 这谁受得了啊!
    shéi xié   qiāng de zhè shéi shòudeliǎo a
    s.o. carry gun SP this who stand   SFP
    'In your class, if there is anyone **who does drugs, messes around, or carries a gun,** who can stand this!!'

This construction type is particularly interesting as it cannot be regarded as equivalent to a 'canonical' prenominal RC, since 谁 *shéi* 'who, someone, anyone' in its indefinite use cannot be the head of a prenominal RC (*吸毒的谁都会被开除的 **xī-dú de shéi dōu huì bèi kāichú de* 'all those who do drugs will be expelled);[5] Fang (2004: 151–152) points out that, with an indefinite pronoun, virtually only postposed relatives are grammatical.

    Lastly, Fang (2004) proposes that constructions as (2), repeated here for the sake of convenience, may be analyzed as postnomimal RCs:

(2)   你比如说你跟着那种水平不高的英语老师,
    nǐ   bǐrú-shuō   nǐ   gēn-zhe   nà   zhǒng shuǐpíng bù-gāo   de
    2SG for.instance-say 2SG follow-DUR that type   level   not-high SP
    Yīngyǔ lǎoshī
    English teacher
    他根本不知道那个纯正的英语发音,
    tā gēnběn bù zhīdào nà   ge chúnzhèng de Yīngyǔ fāyīn
    3SG at.all   not know that CL pure   SP English pronunciation
    他英语语法也不怎么样, 你就全完了
    ***tā Yīngyǔ yǔfǎ   yě bù-zěnmeyàng,** nǐ jiù   quán   wán   le*
    3SG English grammar too not-so.great   2SG then entirely finish PERF
    'If, say, you follow that kind of English teacher **whose level is not so high, who doesn't know pure English pronunciation, and who has bad grammar,** then you're lost.'

This type of modifying clauses, lacking the relativizer (structural particle) 的 *de*, but having an 'empty' (and non-obligatory; Fang 2004: 152) third person pronoun 他 *tā* 'he', are non-restrictive (while the restrictive interpretation is normally available for canonical Chinese RCs; see Tang 2007: 140); and, again, there is a pause between the head NP and each RC, confirming the tendency for postnominal RCs

---

5.   Example courtesy of Hu Shenai.

to be less cohesive with respect to the NP they modify (if compared to prenominal RCs). Also, while postnominal 的 *de* RCs may always be turned into prenominal RCs just by moving them to their default position, a 他 *tā* RC cannot appear as such before the head NP: they would have to be recast as 的 *de* RCs. In other words, RCs based on the third person pronoun 他 *tā* can be only postnominal. Here, following Arcodia (2017), I treat all of the above as postnominal RCs, again because my focus is on their function, rather than on their structure.[6]

2.2 Postnominal RCs as category-building constructions

Categories have always been a key issue for philosophers of language, linguists and cognitive scientists. Since Barsalou's (1983) seminal paper, the existence of 'ad hoc' categories, i.e. novel categories, created on-line for purposes related to the actual discourse situation (as e.g. [TOURIST ACTIVITIES TO PERFORM IN BEIJING]), has been recognized alongside that of established categories residing in long-term memory (e.g. [CAT], Mauri 2017: 299). Actually, recent work on categorization (starting at least from Croft & Cruse 2004; see also Mauri 2017; Mauri & Sansò 2018) suggests that not only novel, ad hoc categories are created in interaction, but rather *all* categories are somehow 'negotiated' in discourse, and they are hence context-dependent and created on-line. Thus, words and phrases should not be understood as 'labels', but rather as 'clues' towards the identification of the referent, together with the context and with any other relevant information (Mauri & Sansò 2018). For instance, while *bird* is an established category label, if one says *we get lots of birds in our garden*, the category *bird* must be narrowed down to something like [MOST FAMILIAR TYPE OF SMALL GARDEN BIRD] (e.g. in a European urban setting), to exclude implausible exemplars as *eagle* or *ostrich* (Croft & Cruse 2004: 94).

Furthermore, besides top-down processes of contextual refinement of categories (i.e. starting from a category label, which is adapted to the context), the importance of bottom-up category building, starting from exemplars and moving 'upwards' toward the identification of the category, has been often stressed in the recent literature (Mauri & Sansò 2018). In actual language use, category labels, exemplars, and other categorization clues are not mutually exclusive: in point of fact, exemplars of the category often occur before or after a category label (in roughly 80% of the cases in Lo Baido's 2018 sample of Italian utterances), and help the hearer identify the borders of the category (see also Example (1) above). In this

---

6. Note that Wang and Wu (2020) take into consideration in their survey only those clauses marked by the structural particle 的 *de*, acting as a relativizer; thus, postnominal clausal modifiers as those exemplified in (2) are excluded under their definition.

context, RCs may be used to express categorizing properties. Consider the following Italian example (adapted from Lo Baido 2018: 88; glosses omitted for ease of presentation):

(10) *Posso immaginare una cosa del genere, che ne so, in un ostello universitario, in cui alla fine si fa una vita a parte, lontano dalla propria città e dalla propria famiglia, e si instaura di solito un grande cameratismo*
'I can imagine something like that, I don't know, in a student hostel, where in the end students are isolated from their environment, away from their hometown and family, and where often a strong sense of camaraderie develops'

In (10), *ostello universitario* 'student hostel' may be understood as an exemplar of a broader category; the following complex relative clause ('where in the end students are isolated from their environment […]' is analyzed by Lo Baido (2018: 88) as such:

the complex relative clause following the example 'student hostel' contributes to the understanding of the category comprising 'hostels, colleges, guesthouses, and so on…' […] After exemplifying, the speaker may keep on talking about the category or predicating something about the example, which may become the centre of attention.

Interstingly, Lo Baido (2018: 85) highlights that, in her sample, when a category label cooccurs with concrete exemplars and/or clues, the label most often (> 80% of instances) comes *before* the exemplars; obviously, this is also the case for postnominal RCs, in which the head noun normally acts as a (more or less specific; see the next example) category label. Compare the following Mandarin example (adapted from Wang & Wu 2020: 1526; my emphasis):

(11) 你说一个你曾经做过的你觉得最女性的
nǐ shuō yī gè nǐ céngjīng zuò-guo de nǐ juéde zuì nǔxìng de
2SG say one CL 2SG ever do-EXP SP 2SG think most feminine SP
就是对你来说是最难度大的一件事情,
jiùshì duì nǐ láishuō shì zuì nándù dà de yī jiàn shìqíng
COP to 2SG about COP most difficulty great SP one CL thing
跟生活有关的, 我是说就是
gēn shēnghuó yǒuguān de, *wǒ shì shuō jiùshì*
with life concern SP 1SG COP say COP
'Please, tell us about something you've done which you think is most 'feminine' and, like, more difficult, and **which is related to your life**, I mean.'

In (11), the generic label 事情 *shìqíng* 'thing (abstract)' acts as a 'placeholder', i.e. as a non-specific expression, the identification of which relies on exemplars and

modifiers, as e.g. RCs (see Lo Baido 2018: 86). The noun 'thing' is preceded by three canonical RCs, which frame it as 'most 'feminine'' and 'most difficult'; however, the speaker then resorts to a fourth, postnominal RCs, which acts as a further clue towards the identification of the properties identifying the set ([THING THAT IS 'FEMININE', DIFFICULT, AND RELATED TO LIFE]). See also the following example (adapted from Wang & Wu 2020: 1518–1519; my emphasis):

(12) 邀了几个蓝营的, 偏蓝的, 去过中国大陆的, […]
yāo-le   jǐ   gè lán-yíng   de piān-lán de **qù-guo Zhōngguó**
invite-PFV some CL blue-camp SP pro-blue SP **go-EXP China**
dàlù   de
continent SP
'(he) invited some people from the blue camp[7] **who have been to Mainland China**'

In Wang & Wu's analysis, 蓝营的 *lán-yíng de* 'from the blue camp' is the head, and 去过中国大陆的 *qù-guo Zhōngguó dàlù de* 'who have been to Mainland China' is a postnominal RC, delimiting the reference of the head. Here, I may add, the second modifier too serves the function of clarifying the reference of the head.

Since both prenominal and postnominal RCs, in principle, are modifiers, the function of which is to narrow down the reference of the head (see above, § 2.1), why would postnominal RCs be particularly suited as devices for on-line categorization? Wang & Wu (2020) highlight some distinctive features of postnominal RCs (as opposed to prenominal RCs) which, I believe, are relevant in this connection:

a. postnominal RCs are mostly independent intonational units, preceded and followed by a pause (see above, § 2.1) or by other prosodic markers
b. they can cooccur with an illocutionary marker (as 我是说 *wǒ shì shuō* 'I mean' in Example (11))
c. more than 76% of the RCs in their sample modify a sentential object

Properties a. and b. are regarded by Wang & Wu (2020) as indicative of the fact that, as mentioned earlier (§ 2.1), postnominal RCs in Mandarin may be understood as afterthoughts, mostly. Building on existing definitions of afterthoughts (see the references quoted in Wang & Wu 2020), they highlight that, besides occurring at the right periphery of a clause (and, hence, prenominal RCs are excluded by definition), they tend to be intonationally independent, may be accompanied by

---

7. In Taiwanese politics, 'blue' represents the Nationalist Party (a.k.a. 'Kuomintang' in English).

illocutionary markers, and serve the function of clarifying an ambiguous referent mentioned earlier.[8]

As to property c., namely the dominance of object-modifying postnominal RCs (*vs.* the subject-modifying dominance for canonical prenominal RCs), Wang & Wu propose that Chafe's (1987) 'One New Concept at a Time Constraint', predicting that an intonation unit expresses no more than one new concept, is at play here: as separate intonational units, postnominal RCs should be particularly suited as modifiers for objects, which are more likely to be new, if compared to subjects which are often given.[9] And, I dare say, the very idea of the co(n)textual construction of a category somehow implies that a new entity is being introduced; or, at least, that its reference is being negotiated.

Furthermore, following Hawkins's (1983) Heaviness Serialization Principle, RCs are the construction type most likely to be placed after the head noun, being the heaviest; thus, if multiple modifiers of the head are present (as in Example (2) above), the RC is expected to be the most likely candidate for postnominal placement (Wang & Wu 2020). Also, as I will show below (§ 4), on-line category building often involves the interaction of several elements; if, as in the case of RCs, these elements have the syntactic function of modifiers of the NP, a tendency to place them after the head is arguably consistent with the above-mentioned principle.

---

8. An anonymous reviewer pointed out that a lower degree of prosodic cohesion is typically associated with non-restrictive RCs. While 他 *tā* RCs (see Example (2)) are said to be always non-restrictive (see Fang 2004), postnominal RCs in legal texts are generally understood as restrictive (Dong 2003). However, the connection between prosodic cohesion and restrictiveness is not always straightforward in Mandarin. See the following example (from Liu 2008: 13):

> 吃螃蟹的毛利人
> *chī pángxiè de Máolìrén*
> eat crab ㏚ Māori
> 'the Māori who eat crab' (i.e. as opposed to those who do not)
> 'the Māori, who eat crab'

Thus, we do have cases in which there is no formal distinction between restrictive and non-restrictive RCs. It has been suggested that the distinction between the two types of RCs is actually based on semantics and pragmatics, rather than on structural differences (see Wang & Wu 2020 and the references cited therein). It appears that postnominal RCs may be both restrictive and non-restrictive, depending on the construction and on its semantic and pragmatic properties.

9. Note that, as pointed out by Wang and Wu (2020), in some Chinese dialects, as Cantonese (see Luke 1998) and Tunchang Min (Qian 2002), the head noun of a postnominal RC must be indefinite. This, however, is not true for Mandarin and for many other Sinitic varieties, as e.g. Kaiping Yue.

All of the features of postnominal RCs presented above are perfectly in line with my proposal, namely that they may be understood as devices for on-line category construction. As said earlier, the use of exemplars, reformulations, or just any categorization clue is quite often accompanied by an explicit mention of a category label (or placeholder); since the category is being introduced as something to be defined in discourse, it is likely to be new, rather than given, and is then retained as a topic and further elaborated upon (see Lo Baido 2018: 88), as hinted at above. In the literature on on-line categorization, we find a very broad range of linguistic constructions which have been analyzed as category 'triggers', or as 'clues': put very simply, 'triggers' are elements which signal the existence of other elements beyond the exemplars mentioned (as e.g. *etc.*, or the above mentioned non-exhaustive connectives 啊 *ā*… 啊 *ā*), while '(property) clues' are "linguistic elements in the co-text that provide explicit information useful to guide the inferential process" toward the identification of the property underlying the category, and, from the perspective of the speaker, "successive stages along the online construction of reference and meaning, whereby [the speaker] recurs to reformulation, elaboration, exemplification, and anaphoric encapsulation […]" (Barotto & Mauri 2018: 109). In my view, postnominal RCs may be seen as part of the latter class of elements; see the following Italian example (adapted from Barotto & Mauri 2018: 119; glosses omitted for ease of presentation; my emphasis):

(13) *questi aspetti **che poi il paziente ha molta resistenza a esprimere al medico** quindi tutto quello **che riguarda l'apparato genitale** [disfunzioni malformazioni eccetera] fanno sempre campo così di un campo su cui c'è molta eh reticenza a parlare*
'[there are] some things **that the patient is reluctant to talk about to the doctor**, that is everything **that concerns the genital apparatus** [malfunctioning malformations etc.], these are part of a field, so to say, about which there is much reticence'

In (13), we see a list of exemplars ('malfunctioning' and 'malformations') followed by a general extender ('etc.'), a typical trigger; the two RCs ('that the patient is reluctant to talk about to the doctor' and 'that concerns the genital apparatus') are analysed by Barotto & Mauri (2018: 119–120) as "abstract formulations" which, together with the explicit exemplars and the mention of 'reticence' ("which thus crucially characterizes the way in which the list items should be conceived in this specific situation"), provide "semantic clues" toward the correct identification of the relevant property of the category ([EMBARASSING CONDITIONS OF THE GENITAL APPARATUS]). However, Mauri and Sansò (2018) also point out that the distinction between exemplars, triggers and clues may become blurred, and one and the same construction may cover all of these functions.

To sum up, while any RC may serve the function of providing clues to the identification of a category, postnominal RCs, which are marked, less frequent constructions in Mandarin Chinese, appear to be particularly suited for this function, for discourse- and processing-related reasons. I will elaborate on this in § 4; let us now turn to the methodological aspects of my research.

3.   My survey

As mentioned in the Introduction (§ 1), the data I use for the present research are drawn from the National Broadcast Media Language Corpus of the Communication University of China (henceforth: NBMLC), a collection of transcribed texts from television and radio programmes, ranging from 2008 to 2013, totalling about 136 million words. The choice of the corpus was motivated by the need to obtain data of spoken Mandarin, and most freely available large corpora are based on written texts; another major spoken corpus of Chinese, the Corpus of the Centre for Chinese Linguistics at Peking University, has been mostly inaccessible recently, and thus I decided not to use it. Moreover, the NBMLC has already been used fruitfully in Wang & Wu's (2020) study of postnominal RCs, and hence we can be sure that it contains a significant number of instances of this construction type: using only a subset of the whole corpus, Wang & Wu could find as many as 247 postnominal RCs.

The NBMLC offers a wide range of parameters to restrict a query, including, for instance, the type of media (television *vs.* radio), the type of interaction (monologue, dialogue, discussion), the channel/station, etc. However, given that my analysis is qualitative in nature, and that I wanted to have the broadest coverage of different text types/genres, I did not add any filters. As to the extraction of postnominal RCs, following Wang & Wu (2020), I searched for the structural particle 的 *de*, followed by a comma, which generally indicates a pause in the transcriptions.[10] Also, since my definition of RC is less restrictive than that of Wang & Wu (see above, § 2.1), and includes also constructions based on an indefinite pronoun, as in (9), and constructions based on a third person pronoun, as in (2), I performed two more searches. As to the former construction type, I performed a combined search for 有谁 *yǒu shéi* '(lit.) there is someone' and the particle 的 *de* followed by a comma (within the same paragraph); as to the latter construction type, I searched for the pronoun 他 *tā* 'he' preceded by a comma, and followed by another 他 *tā*

---

10. Actually, Wang and Wu (2020) searched for 的 *de* followed by as many as seven punctuation markers. I chose to limit my query to a comma (arguably the most common graphic marker for a prosodic border) for the sake of simplicity.

'he' preceded by a comma (within a 15-character range). I chose to add the second 他 *tā* to narrow down my search to sequences as that exemplified in (2); a query with a single pronoun, in fact, returned way too many irrelevant results (basically, any sentence beginning with a third person pronoun preceded by a comma). Also, by using 他 *tā* as a keyword, I could extract also the instances in which the plural form 他们 *tāmen* 'they' was used, broadening the scope of my search.[11]

For the first search (i.e. 的 *de* + comma), I obtained as many as 211,930 results, the overwhelming majority of which were not postnominal RCs; I then manually extracted 24 RCs for my analysis. The second search, however, returned only 14 hits, most of which (again) were not postnominal RCs; I thus decided to exclude this construction type, given the limited significance of such a low number of occurrences. As to the third search, I had 16,623 hits, from which I manually extracted 26 RCs. Thus, I built a small corpus of 50 occurrences of RCs, in order to perform a more fine-grained qualitative analysis.

4. Presentation and analysis of the data

As to the first construction type, i.e. postnominal RCs marked by the structural particle 的 *de*, the features of the examples in my corpus are largely consistent with what has been observed in the previous literature. As many as 7 RCs in my corpus (14% of the total) appear to belong to legal discourse (Dong 2003; see above, § 2.1), as the following example:

(14) 劳动者因高温作业引起中暑的，
  *lǎodòngzhě **yīn**   gāo-wēn   zuòyè yǐnqǐ zhòngshǔ de*
  worker   because high-temperature work   cause heatstroke SP
  经诊断为职业病、认定为工伤的，
  jīng zhěnduàn wèi zhíyè-bìng   rèndìng wéi gōng-shāng de
  after diagnose as   professional-disease recognize as   work-injury SP
  享受工伤保险待遇。
  xiǎngshòu gōng-shāng bǎoxiǎn dàiyù
  enjoy   work-injury insurance treatment
  'Workers **who because of working at high temperatures suffer from a heat-stroke**, after a diagnosis as professional disease and recognition as work injury, will enjoy work insurance benefits'

---

11. Standard Mandarin Chinese employs a (purely) graphic distinction between the masculine (他), feminine (她) and neuter (它) forms of the character for the third person pronoun *tā*. For the sake of simplicity, here I chose to use only the masculine form for my query.

In (14), insurance provisions for work-related injuries are presented; the postnominal RC "whom because of working at a high temperature suffer from a heatstroke" is required to correctly identify the set of referents to which the provisions apply. This is fairly common in this text type, and sometimes involves very long RCs, arguably the best candidates for postnominal placement, as in the following example (from the same excerpt):

(15) 劳动者在工作时间和工作岗位上
lǎodòngzhě zài gōngzuò shíjiān hé   gōngzuò gǎngwèi shàng
worker   in   work   time   and   work   place   at
中暑死亡或中暑后48小时内
zhòngshǔ sǐwàng huò zhòngshǔ hòu 48 xiǎoshì nèi
heatstroke die   or   heatstroke after 48 hour   within
经抢救无效死亡的,
jīng qiǎng-jiù   wúxiào   sǐwàng de
after emergency-treatment ineffective die   SP
视为工伤, 享受工伤保险待遇
shì wéi gōng-shāng xiǎngshòu gōng-shāng bǎoxiǎn dàiyù
see as work-injury enjoy   work-injury insurance treatment
'Workers **who die after suffering a heatstroke during work hours and at their workplace or die within 48 hours of the event after ineffective emergency treatment**, will be considered as having suffered work injury, and will enjoy work insurance benefits'

In Example (15), the RC which defines further conditions under which death following a heatstroke is considered a work injury for insurance purposes is extremely long, and is (perhaps unsurprisingly) placed after the NP it modifies. Arguably, this is motivated by processing ease, as heavy modifiers placed before the head delay its recognition, increasing the "burden on short-term memory" (Hawkins 1983: 99; see above, § 2.2).

Thus, the use of postnominal RCs in legal discourse seems to be a specific stylistic feature of this genre (see Wang & Wu 2020), which might be explained by the fact that RCs in this type of texts are often quite long; whatever the reason behind this choice may be, here postnominal placement for RCs appears to be a planned discourse strategy. Moreover, it is safe to assume that, in this type of texts, the property clues and/or exemplars should identify a category/set in the least ambiguous terms possible. So, for instance, in an excerpt as (15), the conditions under which the predication (i.e. being entitled to insurance benefits) holds are spelled out exhaustively.

On the other hand, in more colloquial-oriented texts postnominal RCs do seem to possess the above-mentioned features of afterthoughts (§ 2.2), and the clues and exemplars are not necessarily exhaustive. They generally represent better examples of on-line category construction, with little planning and 'accumulation' of categorization devices. See the following example:

(16) 我带女儿过来, 江苏的, 请假过来的, […]
      *wǒ dài nǚ-ér   guòlai Jiāngsū de **qǐng-jià guòlai de***
      1sg take female-child come Jiangsu SP ask-leave come SP
      'I took my daughter, from Jiangsu, **who asked for a leave to come** […]'

The context of the sentence in (16), omitted due to space constraints, is that of visiting the Expo in Shanghai, and attending a concert which took place there. The (new) referent 'daughter', the object of the matrix clause, is introduced as a bare NP, but then two relevant clues are added: the fact that she lives in Jiangsu, and that she had to ask for a leave to join her father. The RC 'asked for a leave to come' does look like an afterthought, added on-line in spontaneous speech production. By adding these two modifiers, the speaker arguably frames the event of attending the show as something which was worth travelling for. See also the following example:

(17) 天天上班经过那条路, 没有树的, 我每天做防护措施
      *tiāntiān   shàng-bān jīngguò nà tiáo lù   **méi yǒu shù de** wǒ*
      every.day go-work   pass   that CL road NEG have tree SP 1sg
      měi-tiān   zuò fánghù cuòshī
      every-day do protect measure
      也没有用, 还是很晒的.
      yě méi yǒu yòng háishì hěn shài
      still NEG have use   still   very scorching
      'Every day, on my way to work, I walk through that road, **which has no trees,**
      I try to protect myself from the sun every day, and yet it's scorching anyway'

In (17), the generic label 路 *lù* 'road' is introduced, and the RC 没有树的 *méi yǒu shù de* 'which has no trees' adds the property which is most relevant in this context, namely that of being too scorching, in the speaker's perspective. Thus, differently from Examples ((14)–(15)), in (16) and (17) only contextually-relevant properties are spelled out, and those properties arguably help the hearer in correctly framing the events narrated.

As to the second construction type, namely RCs based on the third person pronoun 他 *tā*, their function as categorial clues emerges more clearly than for 的 *de* postnominal RCs. This type of RCs sometimes cooccur with illocutionary discourse markers; see the following example:

(18) 你比如说很多地方的官员, 他到北京来设一个
  nǐ bǐrú   shuō hěn duō dìfang de guānyuán tā dào   Běijīng
  2SG for.instance say   very many place SP official   **3SG arrive Beijing**
  shè yī gè
  set.up one CL
  驻京办, 为什么买那么多豪华车, 他哪怕是一个
  zhù-jīng-bàn   wèishénme mǎi nàme duō háohuá-chē tā nǎpà
  at-capital-office why   buy so   many luxury-car 3SG even.though
  shì yī gè
  COP one CL
  县处级的干部, 甚至是一个科局级, 就是县下面局级的,
  xiànchù-jí de gànbù shènzhì yī gè kējú-jí   jiùshì xiàn   xiàmiàn
  county-level SP cadre even   one CL section-level COP   county below
  jú-jí   de
  office-level SP
  这些干部, 他到北京来, 他都希望有一个豪华的名车
  zhè-xiē gànbù tā dào Běijīng lái   tā dōu xīwàng yǒu   yī gè
  this-PL cadre 3SG to   Beijing come 3SG all   hope   there.be one CL
  háohuá de míng-chē
  luxury SP famous-car
  来接他
  lái   jiē   tā
  come pick 3SG
  '**For instance**, officials from many places, **who come to Beijing and set up a
  representative office**, why do they buy so many expensive cars, **even though it's
  a county-level cadre, or even a section-level cadre'**, that is, below the county
  level, these cadres**, when they come to Beijing, they all want a nice car to pick
  them up'

The excerpt in (18) begins with the trigger 比如说 *bǐrú shuō* 'for instance', and
then the broad category 'officials (from all over the country)' is introduced. The
category is further specified by the postnominal RC 'who come to Beijing to set
up a representative office', and is then followed by the key predication, namely that
this type of officials spend a lot of money on luxury cars. This is in turn followed by
another clause, the status of which is ambiguous between another postnominal RC
and a concessive clause, providing further clues to the proper individuation of the
category: the officials at issue are low-level cadres; the illocutionary marker 就是
jiùshì 'that is' helps clarifying the relevance of 'section-level cadre' (i.e. the fact that
this is even lower than county-level in the administrative hierarchy). This in turn
helps the hearer in correctly framing the fact that said officials want nice, luxury
cars to drive them around in the capital Beijing: given their relatively low status,
this is presented as something which is excessive.

Interestingly, weight might be an even more important factor in the placement of this type of modifier clauses after the head, if compared to the more 'canonical' 的 *de* postnominal RCs discussed just above. This is because postnominal 他 *tā* RCs are not only long, but, also, often come in 'bundles' of two or more, as seen above for Example (2): while all of the 的 *de* RCs in my sample are single clauses, I found as many as nine sequences of more than one 他 *tā* RCs, and only three cases of isolated 他 *tā* RCs. The following excerpt is a case in point:

(19)   因为很多民进党支持者, 就我所知道, 我们有些基层,
       yīnwèi hěn duō Mínjìndǎng zhīchízhě jiù   wǒ suǒ zhīdao
       because very many DPP   supporter according 1sɢ that know
       wǒ-men yǒu xiē   jī-céng
       1-ᴘʟ   have some base-level
       他平常是洗碗工, 他甚至在市场卖菜,
       tā píngcháng shì xǐ-wǎn-gōng   tā shènzhì zài shìcháng
       3sɢ often   ᴄᴏᴘ wash-bowl-worker 3sɢ even   at market
       mài cài
       sell vegetables
       但是他长期都在资助民进党, 也就是他赚非常非常
       dànshì tā cháng-qī dōu zài   zīzhù   Mínjìndǎng yě jiùshì tā zhuàng
       but   3sɢ long-term all ᴘʀᴏɢ support DPP   or ᴄᴏᴘ   3sɢ earn
       fēicháng fēicháng
       very   very
       微薄的所得, 他还要拿一部分去支持民进党
       wēibó de suǒdě tā hái ná yī bùfen qù zhīchí Mínjìndǎng
       meager sᴘ income 3sg still take one part go support DPP
       'Because many Democratic Progressive Party supporters, to my knowledge, we have some grass root supporters, **who often are dishwashers, or even who sell vegetables at market stalls, but who are providing long-term support for the party, that is, who earn a very, very meagre income, but still spare some to support the DPP**'

This excerpt is part of a debate on the aftermath of a major political scandal in Taiwan, namely the conviction of Chen Shui-bian, former president and very prominent member of the Democratic Progressive Party, for corruption and embezzlement. The show anchor asks the utterer of (19) what are her thoughts concerning how DPP supporters feel after the case was made public. It thus appears that, in Example (19), the speaker wants to convey the point that there are grassroot supporters of the DPP who, even though they have very little disposable income, still provide constant financial support for the Party, and these are the ones what would feel more disappointed by the scandal. She introduces them as 'grassroot-level' supporters, and then adds two RCs to narrow down the category: people who work as

dishwashers, or even sell vegetables at market stalls. These are obvious exemplars of low-income people, at least in Taiwanese society. She then states this clearly, by making direct reference to their 'very, very meagre income' in the following RC; this is in fact the key property that defines the set of people she wants to talk about. Then, she gets to the point, stating that they would still spare something of this little money to support the Party. If we were to turn all of those RCs into prenominal modifiers, the result would be a rather awkward sequence of very long modifiers, with the head appearing only after as many as five clauses. While this is not outright ungrammatical (see Wang & Wu 2020), it does result in a very cumbersome sentence. Note, also, that just as in Example (18), the clause in which a key property is explicitly mentioned is introduced by the illocutionary marker 就是 *jiùshì* 'that is'. Interestingly, RCs based on the pronoun 他 *tā* are used to add exemplars and clues even when a very specific label is used, as in the following excerpt:

(20)   或是有一些安养中心, 安养中心就是一些老人, 甚至是
　　　huòshì yǒu　　　yīxiē　ānyǎng zhōngxīn ānyǎng zhōngxīn jiùshì yīxiē
　　　or　　　there.be some care　centre　care　centre　COP　some
　　　lǎo-rén　　　shènzhì shì
　　　old-person even　COP
　　　植物人, 他们已经没有什么知觉, 他们可能躺在病床
　　　zhíwù-ren　　tā-men yǐjīng　méi　yǒu　shénme zhījué　　　tā-men
　　　plant-person 3-PL　already NEG have any　consciousness 3-PL
　　　kěnéng zài　bìng-chuáng
　　　maybe at　sick-bed
　　　上没有什么感觉, […]
　　　shàng méi　yǒu　shénme gǎnjué
　　　on　　NEG have any　　feeling
　　　'Or, there are some elderly centres, at this elderly centres they have some old women and men, even people in a persistent vegetative state, **who already have no consciousness at all, who maybe lie on their hospital bed without feeling anything** […]'

The broader context for (20), again omitted due to space constraints, is that of a denunciation of unethical behaviour on the part of healthcare operators; in this excerpt, the speaker goes on telling that in some elderly centres, doctors would perform useless operations on patients in a persistent vegetative state to obtain refund from their health insurance. While the specific label 植物人 *zhíwùrén* '(lit.) vegetable person' is first introduced to identify the category of patients the speakers is talking about, s/he still uses two abstract formulations, expressed by RCs, to further illustrate what s/he means by this: people who have no consciousness, and lie in bed without feeling anything. These two properties of people in a persistent vegetative state are extremely relevant in the context, and stating them arguably

helps the speaker convey her/his point. This is, I believe, a clear case of the type of on-line construction of meaning at issue here, with the (seemingly) unplanned cooccurrence of category labels, abstract (re-)formulations, exemplars, and the like (see above, § 2.2).

Thus, postnominal RCs based on the pronoun 他 *tā* seem to be productively used to contribute to the on-line construction of categories, and they seem to be a very flexible tool for this purpose, given that, apparently, any number of postnominal 他 *tā* RCs may be added after a head NP, and they freely interact with illocutionary markers. As shown in the examples above (esp. Example (19)), they are used both to introduce exemplars and as categorial clues, often in the form of abstract formulations which spell out a defining property of the set of entities being introduced. Given that, as said above (§ 2.1), there is no direct prenominal equivalent for this construction type, and given that 他 *tā* RCs seem to be freely added as afterthoughts to just about any NP, I dare suggest that on-line category building is but one of the main functions of RCs based on the pronoun 他 *tā*; this appears to be the case for virtually all instances of this construction type in my sample, and it may well be that this pragmatic-discourse function has somehow become conventionalized for 他 *tā* RCs.

5. Summary and conclusions

In this paper, I proposed an analysis of Mandarin postnominal RCs in terms of their pragmatic-discourse function. Based on previous studies of postnominal RCs in Chinese and on the analysis of a small sample of spoken Mandarin, I argued that postnominal RCs are well suited as devices for on-line category building, by providing clues, as e.g. abstract formulations and/or exemplars, that narrow down the label introduced as head NP. I also argued that this is especially true for postnominal RCs based on the third-person pronoun 他 *tā*, which have no direct prenominal counterpart: they are particularly suitable for this function as they often appear in bundles, and freely interact with other devices for category construction, as well as with illocutionary markers. The more 'canonical' RCs marked by the structural particle 的 *de* are also often added as afterthoughts, but they seem to appear mostly one at a time; also, they are often found in legal texts, as part of a planned discourse strategy in which all of the relevant properties of the category are normally spelled out, with minimal inferential effort on the part of the hearer.

What also emerges from my survey is that the inferential process leading to the hearer's identification of the category being named may be more or less 'guided', depending on several factors. The fact that a category label (or, at least, a placeholder; see Example (11)) is necessarily present as the head of the RC does not entail that the process is always straightforward. I have already pointed out that postnominal

RCs as used in legal texts generally involve the least inferential effort, since the RCs (and other modifiers) normally provide all the clues and/or exemplars (exhaustively listed) for the identification of the category, for obvious reasons. In other cases, labels at different levels of specificity are used, and the clues and/or exemplars may allow for a greater or lower degree of 'referential ambiguity' (for lack of a better term). Still, even when a highly specific category label is introduced, as in Example (20), RCs may be anyway used as relevant exemplars or illustrations of the label itself; as pointed out earlier (§ 2.1), in on-line category building, exemplars are often used together with a category label, in order to help the hearer refine the set designed by the category in that context. Besides, postnominal RCs, particularly 他 *tā* RCs, are sometimes used to introduce exemplars or (re-)formulations as a discourse strategy which, I suggest, is employed for rhetorical purposes; namely, to provide the hearer with a more vivid picture (see Examples (19)–(20), among others).

Lastly, while postnominal 的 *de* RCs appear to be but a variant of the canonical prenominal RC, arguably motivated by processing reasons, 他 *tā* RCs seem to emerge as a separate construction type, a key function of which is to provide the means for on-line category construction; to the best of my knowledge, this has not been highlighted before in the literature. However, given the very limited size of my corpus, I believe that further research is needed to assess the appropriateness of this analysis.

Acknowledgements

References

Arcodia, Giorgio Francesco. 2017. Towards a typology of relative clauses in Sinitic: Headedness and relativization strategies. *Cahiers De Linguistique Asie Orientale* 46: 33–72. https://doi.org/10.1163/19606028-04601002

Barotto, Alessandra & Mauri, Caterina. 2018. Constructing lists to construct categories. *Italian Journal of Linguistics* 30(1): 95–134.

Barsalou, Lawrence W. 1983. Ad hoc categories. *Memory and Cognition* 11(3): 211–227. https://doi.org/10.3758/BF03196968

Chafe, Wallace. 1987. Cognitive constraints on information flow. In *Coherence and Grounding in Discourse* [Typological Studies in Language 11], Russell S. Tomlin (ed.), 21–52. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.11.03cha

Chan, Angel, Matthews, Stephen & Yip, Virginia. 2011. The acquisition of relative clauses in Cantonese and Mandarin. In *The Acquisition of Relative Clauses: Processing, Typology and Function* [Trends in Language Acquisition Research 8], Evan J. Kidd (ed.), 197–226. Amsterdam: John Benjamins. https://doi.org/10.1075/tilar.8.10cha

Chao, Yuen-Ren. 1956 [1928]. 现代吴语的研究. Bejiing: Kexue Chubanshe.

Comrie, Bernard & Kuteva, Tania. 2013. Relativization strategies. In *The World Atlas of Language Structures Online*, Matthew S. Dryer & Martin Haspelmath (eds). Leipzig: Max Planck Institute for Evolutionary Anthropology.

Comrie, Bernard. 2003. Typology and language acquisition: The case of relative clauses. In *Typology and Second Language Acquisition*, Anna Giacalone Ramat (ed.), 19–38. Berlin: Mouton de Gruyter

Croft, William & Cruse, Alan D. 2004. *Cognitive Linguistics*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511803864

Dai, Zhaoming. 2006. 天台方言研究. Beijing: Zhonghua Shuju.

Dong, Xiufang. 2003. "的" 字短语做后置关系从句的用法—兼评法律文献中的"的"字短语的用法. *Yuyan Wenzi Yinyong* 4: 120–126.

Dryer, Matthew S. 2013. Relationship between the order of object and verb and the order of relative clause and noun. In *The World Atlas of Language Structures Online*, Matthew S. Dryer & Martin Haspelmath (eds). Leipzig: Max Planck Institute for Evolutionary Anthropology.

Fang, Mei. 2004. 汉语口语后置关系从句研究. In 庆祝《中国语文》创刊*50*周年学术论文集, Editorial Board of Zhongguo Yuwen (eds). Beijing: Shangwu Yinshuguan.

Fretheim, Thorstein. 1995. Why Norwegian right-dislocated phrases are not afterthoughts. *Nordic Journal of Linguistics* 18(1): 31–54. https://doi.org/10.1017/S0332586500003097

Givón, Talmy. 2001. *Syntax: An Introduction*, 2 Vols. Amsterdam: John Benjamins.

Hawkins, John A. 1983. *Word Order Universals*. New York (NY): Academic Press.

LaPolla, Randy J. 2017. Noun-modifying clause constructions in Sino-Tibetan languages. In *Noun-modifying Clause Constructions in Languages of Eurasia: Rethinking Theoretical and Geographical Boundaries* [Typological Studies in Language 116], Yoshiko Matsumoto, Bernard Comrie & Peter Sells (eds), 91–103. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.116.05lap

Li, Charles N. & Thompson, Sandra A. 1981. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley CA: University of California Press.

Liu, Danqing. 2005. 汉语关系从句标记类型初探. *Zhongguo Yuwen* 1: 3–15.

Liu, Danqing. 2008. 汉语名词性短语的句法类型特征. *Zhongguo Yuwen* 1: 3–18.

Lo Baido, Maria Cristina. 2018. Categorization via exemplification: Evidence from Italian. In *Linguistic Strategies for the Construction of Ad Hoc Categories: Synchronic and Diachronic Perspectives*, Caterina Mauri & Andrea Sansò (eds). Special issue of *Folia Linguistica* 52(s39–1): 69–95. https://doi.org/10.1515/flih-2018-0007

Luke, Kang-Kwong. 1998. 粤语名词组中的后置修饰语. *Fangyan* 1: 48–52.

Matthews, Stephen & Yip, Virginia. 2017. Noun-modifying clause constructions in Sino-Tibetan languages. In *Noun-modifying Clause Constructions in Languages of Eurasia: Rethinking Theoretical and Geographical Boundaries* [Typological Studies in Language 116], Yoshiko Matsumoto, Bernard Comrie & Peter Sells (eds), 105–120. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.116.06mat

Mauri, Caterina & Sansò, Andrea. 2018. Linguistic strategies for ad hoc categorization: Theoretical assessment and cross-linguistic variation. In *Linguistic Strategies for the Construction of Ad Hoc Categories: Synchronic and Diachronic Perspectives*, Caterina Mauri & Andrea Sansò (eds). Special issue of *Folia Linguistica* 52(s39–1): 1–35. https://doi.org/10.1515/flih-2018-0001

Mauri, Caterina. 2017. Building and interpreting ad hoc categories: a linguistic analysis. In *Formal Models in the Study of Language*, Joanna Blochowiak, Cristina Grisot, Stephanie Durrleman-Tame & Christopher Laenzlinger (eds), 297–326. Cham: Springer. https://doi.org/10.1007/978-3-319-48832-5_16

Paul, Waltraud. 2015. *New Perspectives on Chinese Syntax*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110338775

Podlesskaya, Vera I. 2011. Relative clauses in spoken Russian and elsewhere: A corpus approach. In *Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue"*, 529–537. Moscow: Izdatel'stvo RGGU.

Qian, Dianxiang. 2002. 海南屯昌闽语语法研究. Kunming: Yunnan Daxue Chubanshe.

Tang, Zhengda. 2007. 关系化对象与关系从句的位置 – 基于真实语料和类型分析. *Dangdai Yuyanxue* 2: 139–150.

Wang, Fang & Wu, Fuyun. 2020. Postnominal relative clauses in Chinese. *Linguistics* 58(6): 1501–1542.

Wiechmann, Daniel. 2015. *Understanding Relative Clauses*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110339581

Wu, Tong. 2011. The syntax of prenominal relative clauses: A typological study. *Linguistic Typology* 15: 569–623. https://doi.org/10.1515/LITY.2011.036

Wu, Yunji. 2005. *A Synchronic and Diachronic Study of the Grammar of the Chinese Xiang Dialects*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110927481

Yap, Foong Ha & Matthews, Stephen. 2008. The development of nominalizers in East Asian and Tibeto-Burman languages. In *Rethinking Grammaticalization. New Perspectives* [Typological Studies in Language 76], María José López-Couso & Elena Seoane (eds), 309–341. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.76.15yap

Yue, Anne O.-K. 1995. 广东开平方言的 "的" 字结构. *Zhongguo Yuwen* 4: 289–297.

Zhang, Niina Ning. 2008. Encoding exhaustivity. *USTWPL* 4: 133–143.

Zhao, Zhengying. 2009. 后置关系从句的重新分析. *Yuwen Xuekan* 1: 87–88.

Et cetera, eccetera, etc. The development of a general extender from Latin to Italian

Ilaria Fiorentini and Elisabetta Magni
University of Pavia / University of Bologna

This paper discusses the range of functions of *et cetera* in Latin, as well as the multifunctionality and evolution of *eccetera* in Italian. Both forms pertain to the category of general extenders, i.e. markers whose main function is to 'extend' otherwise grammatically complete utterances. We will propose a qualitative analysis based on data from Latin, Old Italian, and contemporary written and spoken Italian. In particular, we will discuss the semantic-pragmatic expansion of *et cetera* and *eccetera*, highlighting how both expressions tend to develop intersubjective and procedural meanings in addition to their original functions. Moreover, since the original components of *eccetera* are by now blurred in contemporary Italian, we will show how recent developments allow its use in disjunctive contexts.

Keywords: general extenders, Latin, Italian, intersubjectivity

1. Introduction

In current research, the various epigones of the Latin locution *et cētĕra*, which literally means 'and (the) other things', have been analysed both as indicators of vagueness and as typical examples of general extenders. As is well known, this term designates a series of expressions such as Engl. *and stuff like that, and everything, or something*, that is, markers that tend to occur in phrase- or clause-final position and typically display a basic syntactic structure consisting of a conjunction followed by a generic nominal (or a proform), with an optional comparative or similative phrase. Such forms are attested in many languages and, as evidenced by the fortunate label coined by Overstreet (1999: 3), their basic characteristics are: to have nonspecific, 'general' reference and to 'extend' otherwise grammatically complete utterances.

Although many studies quote *et cetera* as a prototypical example of the category, no attempt has been made so far to investigate the evolution of this specific

expression.[1] To fill this gap, in this paper we will first illustrate the uses and functions of *et cetera* in Latin (Section 2), and then we will discuss the multifunctionality and recent developments of *eccetera* in Italian (Section 3). To this end, we will propose a qualitative analysis of the materials obtained from various databases:[2] for the Latin language the *LLT-Series A* (*Brepolis Library of Latin Texts*),[3] for Old Italian the OVI corpus (*Opera del Vocabolario Italiano*),[4] for contemporary written Italian the CoLFIS corpus (*Corpus e Lessico di Frequenza dell'Italiano Scritto*)[5] and ItTenTen 2016 (*Italian Web corpus*),[6] and finally, for spoken contemporary Italian, the LIP (*Lessico di frequenza dell'italiano parlato*),[7] LIT (*Lessico Italiano Televisivo*),[8] and KIParla[9] corpora. In the last part (Section 4), we will discuss the semantic and formal evolution of the general extender, which in spoken contemporary Italian displays intersubjective and procedural meanings in addition to its original objective and conceptual functions, but is also gradually developing new analytic variants.

2.   *Et cetera* in Latin

As in other languages, in Latin as well the category of general extenders has received little attention in traditional grammars. The data collected by Magni (2019) confirm the usual distinction between adjunctive and disjunctive general extenders (Overstreet 1999: 3): the former are introduced by *et, atque* or enclitic *-que* 'and', while the latter are introduced by *aut* and *uel* 'or'. In addition, the two types display both short forms (e.g. *ceteraque* 'etcetera', *et reliqua* 'and the remaining things', *aut similia* 'or the like') and long variants (e.g. *et sic cetera* 'and similar other things', *et alia de hoc genere* 'and other things of this kind', *aut aliquid eiusmodi* 'or something like that').

---

1.   For an overview of the forms and functions of general extenders in Latin see Magni (2019). For a discussion of Italian *eccetera* see Fiorentini (2018) and Fiorentini (2019).

2.   The source of the data will be specified within brackets after each example.

3.   Available at: <http://apps.brepolis.net/BrepolisPortal/default.aspx>. Cf. Tombeur (2016).

4.   Available at: <http://gattoweb.ovi.cnr.it>

5.   Available at: <http://linguistica.sns.it/CoLFIS/Home.htm>

6.   Available at: <https://www.sketchengine.co.uk/ittenten-corpus/ >(2016 version).

7.   Available at: <http://badip.uni-graz.at/en/> Cf. also De Mauro et al. (1993); Voghera et al. (2014).

8.   Available at: <http://www.italianotelevisivo.org/contenuti/36/banche_dati>

9.   Available at: <https://kiparla.it/>. Cf. Goria & Mauri (2018) (KIP module).

The overall inventory of Latin adjunctive general extenders found on the Brepolis database, within the period *Antiquitas* (works of so-called Classical Antiquity, from ca. 200 BC to ca. 200 AD), is summarized in Table (1), adapted from Magni (2019: 701):[10]

Table 1. Adjunctive general extenders in Latin

	Short forms	Long forms	Total
(et) cetera / ceteraque 'and the other things'	40	17	57
(et) alia 'and the other things'	7	38	45
et similia 'and similar things, and the like'	20	19	39
et multa 'and many things'	1	14	15
(et) reliqua 'and the remaning things'	9	0	9
et talia 'and such things'	5	0	5
et deinceps 'and so on'	2	1	3
et porro 'and so forth'	0	1	1
	84	**90**	**174**

According to these data, *et cetera*, which contains the neuter plural of the adjective *ceterus, -a, -um* 'the other, that which exists besides, the other part', is evidently the most frequent and, as we will see, the most versatile form.

Considering its consistency with the basic structural features and constraints of the category, it has been observed that *boni scriptores* may at times omit connectors (Hand 1832: 471), as illustrated in Example (1):

(1) *si est nihil nisi corpus, summa erunt illa: ualetudo, uacuitas doloris, pulchritudo, cetera.* (Cic. *fin.* 4, 35)
'If there is nothing except the body, these will be most relevant: health, freedom from pain, beauty, and so on.'

In addition, this expression can be found not only in clause-final position, which is typical of most general extenders, but also in clause-internal position, as shown in (2):

---

10. Disjunctive general extenders, which will not be discussed in this paper, display lower frequency and variability with respect to adjunctive ones, as also in other languages. More specifically, the locution *aut aliquid* 'or something' is attested 4 and 6 times in short and long forms, respectively; *aut / uel similia* 'or similar things' occurs 1 and 2 times, respectively; *aut quiduis* 'or anything, or whatever' is found only once, in the long variant *aut quiduis generis eiusdem* 'or anything of that sort'.

(2)   *Nam mentem, fidem, spem, uirtutem, honorem, uictoriam, salutem, concordiam*
      ***ceteraque huius modi** rerum uim habere uidemus non deorum.*

                                                              (Cic. *nat. deor.* 3, 61)

      'For mind, faith, hope, virtue, honor, victory, health, concord and the like, we
      see them to have the force of things but not of gods.'

Despite these particularities, throughout the history of Latin *et cetera* behaves as a
normal general extender, and the last example confirms that it can also have long
variants with modifiers denoting similarity (e.g. *et cetera huius modi, et cetera de
hoc genere, et sic cetera*). Let us therefore analyse its functions by discussing some
examples taken from literary texts.

2.1   Functions of *et cetera*

2.1.1   *Completing lists and marking sets*

Adjunctive general extenders, which have the basic meaning 'there is more'
(Overstreet 1999: 126), display the typical values of addition when ending a series
of three or more elements. This usage is in fact quite frequent for *et cetera*, as shown
in Examples ((1)–(2)), and also in (3):

(3)   *Nauigia atque agri culturas, moenia, leges,*
      *arma, uias, uestes <et> **cetera de genere horum**,* […]   (Lucr. 5, 1448–1449)
      'Navigation, the cultivation of fields, walls, laws,
      arms, roads, clothes and things like that …'

In these cases, *et cetera* has an enumerative function (Cortés Rodríguez 2006) and
works as a *list completer* (Jefferson 1990). Of course, since in non-exhaustive lists
items are not random but are linked to an idea, the ending general extender usually
suggests an expansion based on objective connections in a particular context, such
as the testimonies of human and civil progress itemised in Example (3).

The role of context and associations is even more prominent in (4), where *et
cetera* does not occur in a list but follows a single exemplar, broadening its reference
and marking it as belonging to a homogeneous group of entities: in this case 'small
insects and parasites'.

(4)   *sentimus nec priua pedum uestigia quaeque,*
      *corpore quae in nostro culices **et cetera** ponunt.*   (Lucr. 3, 389–390)
      'We do not feel each of the footsteps that
      mosquitoes and suchlike place on our body.'

Here, the expression has mainly an illustrative function and works as a *set marker*
(Dines 1980). For Dines, the main function of general extenders, which in her
view are mostly considered as carrying referential meanings, is in fact "to cue the

listener to interpret the preceding element as an illustrative example of some more general case" (Dines 1980: 22). Interestingly, it has also been observed that these expressions are particularly useful in spoken and informal language, because they "provide a way of talking about groups of entities or actions that spontaneously need to be referenced together when no established referring expression for the group is known (or even exists)" (Overstreet 1999: 43). This function is illustrated in (5), where Seneca discusses curious analogies between the earth and the human body, which both have veins and canals in which liquids and air circulate:

(5) *in quaedam uero terra umorque putrescunt, sicut bitumen **et cetera huic similia***.

(Sen. *nat*. 3, 15, 2)

'Some [substances] originate from the decay of the earth and its fluids, such as bitumen and others like it.'

Clearly, when the philosopher was writing the *Naturales Quaestiones*, the category implicated by *bitumen et cetera huic similia* was a non-lexicalized one, but even today the proper superordinate noun, that is, 'natural hydrocarbons', may not easily come to mind and a general extender could thus serve to overcome the lack of adequate scientific knowledge.

2.1.2   *Sharing knowledge and building categories*

Considering the usages discussed above, general extenders can be viewed as *vague category identifiers* (Channell 1994). In some cases, especially when used after a single exemplar, they can also serve to access fluid and temporary conceptual associations, i.e. categories that are "inherently variable, and created on-line as and when needed" (Croft & Cruse 2004: 92). In this role, the expressions at issue do not suggest a well-defined and stable set, but rather imply abstraction over the given exemplar(s) through a context-driven associative reasoning, which leads to infer an *ad hoc* category (Barsalou 1983). In other words, a heterogeneous group of entities whose individuation is crucially conditioned by the exemplar provided, the linguistic context, the purpose of the text and the interlocutors' pragmatic knowledge. This peculiar function of *et cetera* can be found in Example (6):

(6) *Ego, dum panes **et cetera** in nauem parantur, excurro in Pompeianum,*

(Cic. *Att*. 10, 15, 4)

'While the bread, etc. is being made for the ship, I am running to my place at Pompei.'

In Cicero's letter, the locution *panes et cetera* does not suggest a precise set of items, but rather invites Atticus to both try an associative reasoning starting from the mentioned exemplar and to recall well-known information about the necessary (and disparate) 'things for a journey by ship'.

Previous researches on the forms like *et cetera* have centred on the list complet-
ing and set marking functions. Nonetheless, more recently the relevance of their
context-driven interpretation has led scholars to recognize that general extenders
can also be exploited for a number of pragmatic and interpersonal purposes.

2.1.3   *Pragmatic and interpersonal functions*
As observed by Overstreet (2014: 120–121), general extenders often reflect the
Gricean maxim about the quantity of information when they are used to suggest
that enough is said "for the current purposes of the exchange" (Grice 1975: 45),
thus displaying hedging functions. In Latin, *et cetera* is in fact frequently used to
shorten well-known quotations or formulae, like the proverb in (7), or to abbreviate
a series of examples, as in (8):[11]

(7)   *'agas asellum' et cetera*.   (Cic. *de orat.* 2, 258)
       '« Drive the ass », etc.' (sc. *cursum non docebitur* 'he will not be taught the way')

(8)   *dicimus lauo manus, sic pedes et cetera*.   (Varro *ling.* 9, 107)
       'We say: I wash my hands, my feet, etc.'

In similar cases, the general extender seems to develop the implied meaning 'this
is enough', which is probably also suggested by its frequent ending position, and it
is therefore employed to replace something that the writer considers superfluous,
since it is supposed to be already known to the reader.

As has been observed, an interesting correlation of this kind of usage in natu-
rally occurring conversations is that general extenders can often "help to establish
and maintain a sense of rapport among the interlocutors" (Overstreet 1999: 18).
If we apply this statement to the textual discourse, we may say that, similarly, *et
cetera* can help to establish a connection between the writer and the reader, thus
performing interpersonal functions.

As we have seen from the preceding examples, the completion of a list or the
recovery of a specific category can be secondary aspects in the use of general ex-
tenders: these functions become in fact irrelevant when these expressions are used
to imply shared experience or evaluation, and familiarity or solidarity in interac-
tion. This observation seems confirmed by Example (9):

(9)   *additurum principem defunctae templum et aras **et cetera** ostentandae pietati*.
                                                                                (Tac. 14, 3, 3)
       'The emperor would add a temple and shrines and the like for the deceased
       lady, to display filial affection.'

---

11. Considering other Latin general extenders, this usage is peculiar of *et cetera*, which in this
function can be compared to analogous expressions in Ancient Greek such as καὶ τὰ ἕτερα, καὶ
τὰ λοιπά, or καὶ τὰ ἑξῆς.

In this case, *et cetera* ends the reported speech whereby Tacitus imagines that the freedman Anicetus was not only abetting Nero to murder his mother but, at the same time, also inviting complicity and confirming solidarity through the implicit message: 'I don't have to tell you everything because we share this scheme'.

We can thus observe that, in the uses of *et cetera*, the implication of associative reasoning and inference triggers the interplay between objective and subjective meanings, while the suggestion of shared knowledge and common ground complements the emergence of intersubjective meanings "centred on the addressee" (Traugott 2010: 30). As we will see, it is also by virtue of this multifunctionality that the expression becomes more frequent over time.

2.2 The evolution of *et cetera*

Expanding the research on *et cetera* to Late Latin, we find that in the period *Aetas Patrum I* on the Brepolis database (works of Late Antiquity, from ca. 200 to ca. 500) the form tends to gain in frequency. To give an idea of this phenomenon, it is sufficient to say that from 38 occurrences (plus 2 *ceteraque*) in the entire *Antiquitas* period, we reach a total of 314 attestations only in the works of Augustine (Magni 2019: 709). Furthermore, in the long run the overall increase in use reflects in the spread of abbreviations, which are found in medieval manuscripts since the VII century, and in the later univerbation *etcetera*, which is attested since the XI century.

Before these formal changes, in the early imperial age, phenomena of decategorialization (Hopper & Traugott 2003: 106–109) announce semantic bleaching, whereby *et cetera* starts to lose certain morpho-syntactic features by having categories other than noun as referents. For instance, it is interesting to notice that the expression can also be found after verb phrases, as exemplified in (10):

(10) *Eius enim esse inuenire, disponere, eloqui **et cetera**.* (Quint. *inst*. 3, 3, 11)
'For it is his business to invent, arrange, express, etcetera.'

The increasing mismatch between the properties of the 'host' element and those of the proform in the extender (a neuter plural) is even more evident when the preceding item refers to animate entities, like *pedisequos* 'servants' in Example (11):[12]

(11) *Vestem, uniones, pedisequos **et cetera***
Illi adsignate, uitam quae luxu trahit. (Phaedr. 4, 5, 36)
'The clothes, the pearls, the servants and the like,
give them to the one who spends her life in luxury.'

---

12. An anonymous reviewer remarked that the list provided by Phaedrus contains possessions, and that slaves as well were deemed as objects that could be possessed.

Unsurprisingly, these phenomena, partly favoured by the increasing (and sometimes opaque) use after quotations, become more frequent in later authors, as illustrated in (12):

(12) *Det nobis perseuerare in mandatis suis, ambulare in uia recta eruditionis suae, placere illi in omni opere bono, **et cetera talia***.   (Aug. *c. Pel.* 13, 60)
  'May He allow us to persevere in his commandments, to walk the straight path of his instruction, to please him in every good work, and other such things.'

In the example above, Augustine is explaining that Pelagian heretics deny the value of the liturgical prayer of blessing, whose contents are shortened by using *et cetera talia* after a series of clauses. Similar cases signal the referential ambiguity that results from loss of a precise grammatical connection between *cetera* and its antecedents, which leads to the extension to new contexts.

To sum up, we have observed that a Latin expression encoding the basic meaning 'there is more' frequently triggers heuristic procedures such as associative reasoning and similarity inference, whereby speakers construct lists, sets and categories basing on shared knowledge. In some cases, however, we have noticed that the typical ending position of *et cetera* intensifies the additional meaning 'this is enough', whereby speakers conclude, abbreviate, and approximate utterances drawing on Gricean conversational heuristics. Therefore, according to the contextual emergence of different semantic nuances, the form *et cetera* can be used not only to complete lists, mark sets and build ad hoc categories, but also – and increasingly – to perform hedging functions at the textual and interpersonal level.

The scheme in Figure (1), summarizes these coexisting functions, which range along a continuum that reflects the shift from objective to subjective and intersubjective meanings:

LIST COMPLETING - SET MARKING - AD HOC CATEGORIZING - HEDGING (TEXTUAL/INTERPERSONAL)

'there is more'   -   'this is enough'

objective/conceptual > subjective/pragmatic > intersubjective/procedural

----
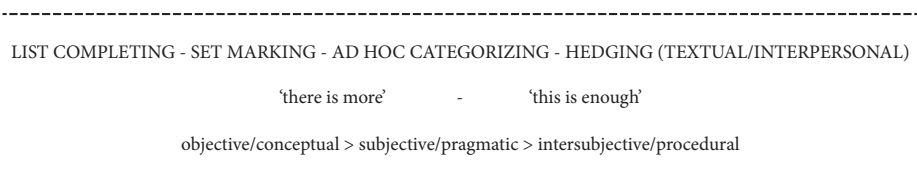
Figure 1. Functions and meanings of Latin *et cetera*

On the whole, the inference of additional meanings that we have posited for *et cetera* is consistent with recent findings concerning the paths of pragmaticalization already observed in other languages (Diewald 2011). In particular, it is in keeping with those phenomena of semantic-pragmatic expansion by which, in the long run, "the set-marking meaning of GEs (*sc.* general extenders) gradually recedes while

their intersubjective and other pragmatic/procedural meanings increasingly come to the fore" (Pichler & Levey 2011: 18). Of course, we have to admit that the interpersonal uses of Latin general extenders appear less developed than textual ones, but this also depends on the nature of our data, which prevents us from grasping the extent of pragmatic functions in spoken language. These aspects will become clearer in the next section, where we will discuss the uses and the recent developments of *eccetera* in contemporary Italian.

3. From Latin to Italian

In Old Italian, *et cetera* underwent a phonological reduction which led to the univerbation in *etcetera* (XI century, see Section 2.2), and subsequently in *eccetera* (XIV century). These forms were often abbreviated (*etc., ec., ecc.*) and are usually found after a list, a description or a quotation, to concisely replace something considered redundant or already known to the interlocutor. The use in a fixed position (i.e. at the end of sentences) was especially common after quotations (Fiorentini 2019), as in (13), and common sayings or formulaic expressions, as in (14):

(13)   *quando disse a Cristo:* "*facciamo qui tre tabernacoli in sul monte Taborre, eccetera*"   (OVI, Bosone da Gubbio, *Fortunatus siculus (l'Avventuroso Ciciliano)*, 1333, L.2, osservazioni, 312, 15)
'When he said to Christ: "Let us make three tabernacles on mount Taborre, etcetera".'

(14)   "(…) *passasti in cielo laddove tu se' risplendente nel mezzo della divina schiera ne lo splendore de la incorruttibile gloria, in allegrezza de la insaziabile gioconditade rallegrandoti, eccetera*". *Queste cose disse Teodoro.*   (OVI, *Leggenda Aurea*, XIV c., ch. 118, S. Bartolomeo – volume 3, page 1040, line 22)
'"You passed into heaven where you are shining in the midst of the divine legion, rejoicing the splendor of the incorruptible glory, the joy of the insatiable playfulness, etcetera". These are the things that Theodore said.'

These uses, as we have seen, were already possible for Latin *et cetera*. The form *eccetera*, however, continued to evolve over time: after a brief overview on previous studies, we will focus on its functions in spoken Italian, in order to account for its multifunctionality and its more recent developments.

3.1 Previous studies on Italian *eccetera*

Although there is a shortage of studies dealing with Italian general extenders as a category (cf. Cucchi 2010; Fiorentini 2018), some works focus on *eccetera* in contemporary written and spoken Italian.[13] Galli de' Paratesi (1969: 43) proposes a comparison with semantically vague expressions like *cosa* 'thing', which can also be used to talk euphemistically about uncomfortable topics and to avoid negative themes (cf. O'Keeffe 2004). For instance, in (15) *ecc.* replaces something that the author of the letter preferred to omit (Galli de' Paratesi 1969: 121, cf. also Prandi 1990; Fiorentini 2019):

(15) *un giorno mi misi con una mia coetanea e andammo in campagna con i nostri fidanzati. Arrivati lì, si sa come si fa, baci, carezze* **ecc.**
      'One day I got together with a girl my age, and we went to the countryside with our boyfriends. When we got there, you know how it goes, kisses, caresses etc.'

Similarly, *eccetera* can be used to shorten well-known texts or formulae, in order to "downgrade the 'more' that is considered routine and predictable" (Overstreet 1999: 131), as in (16):

(16) *L'uomo potrebbe essere, nel migliore dei casi, un molestatore, nel peggiore un vio-lentatore. La donna, per dire pane al pane* **eccetera,** *una ninfomane.* (COLFIS)
      'The man could be, in the best-case scenario, a molester, in the worst, a rapist. The woman, to call a spade a spade etcetera, a nymphomaniac.'

In the example, the second segment of the Italian popular saying *dire pane al pane e vino al vino* (literally 'to call bread bread, and wine wine') is replaced by *eccetera*. In order to be able to complete the saying, the reader must already know it; hence, s/he does not have to infer other possible elements, but rather to import something from her/his memory.

     This function is typical of written Italian, whereas it is rarer in oral speech (Fiorentini 2019: 258). In particular, it can be found when the speaker is quoting part of something previously uttered by one of the participants, or when s/he is reading something out loud to the hearer, as in (17):

(17) *venga si scriva questa lettera signora # allora mette la dottoressa Melania Angotta* **eccetera eccetera** *no prende gli estremi da $*   (LIP_Florence_A_12_A)
      'Please come here and write this letter, Mrs... So, write "Doctor Melania Angotta" etcetera etcetera, no?, you can take the information from [xxx]'

---

13. *Eccetera* can also act as a noun: *ci sono troppi eccetera* 'there are too many ecceteras'; *con tutti questi eccetera* 'with all these ecceteras' (cf. Treccani online, <http://www.treccani.it/vocabolario/eccetera/>).

Here, the speaker is dictating something to the interlocutor while the text is physically available for both participants; therefore, there is no need to read it through completely. At the same time, the hearer does not need to abstract anything.

According to Guil (1999), *eccetera* is one of those "proformas alusivas" (literally 'hinting proforms'), which are typical of oral speech. These proforms are usually placed after a list of heterogeneous (yet similar) elements and refer to the existence of other elements similar to the expressed one(s). Such elements are considered analogous, but nonetheless informatively new, as in (18) (Guil 1999: 95):

(18) *c'è tutto un sistema adesso offre la Macintosh di boh non so quanti computer schermi cose in rete **eccetera** per cui la stampante viene gratis*
'There is a whole system now, Macintosh offers I do not know how many computers, screens, network things, etcetera, this is why the printer is free.'

Furthermore, like other general extenders, *eccetera* has been described as a vagueness marker. For instance, Cucchi, who labels these forms as "vague expressions" or "vague items", states that *eccetera* "may give an impression of vagueness and non-conclusiveness" (Cucchi 2010: 101). On the other hand, Ghezzi (2013: 65) underlines that it may be used to imply a vague categorization, as well as to signal an approximation of the propositional content of the utterance, as shown in (19) (Ghezzi 2013: 142):

(19) *è comodo il cellulare perché magari sei in giro per strada chiami un amico chiami un'amica **eccetera***
'The mobile phone is useful, because if you are outside, on the street, you call a boyfriend, you call a girlfriend, etcetera.'

Here, the speaker uses two examples (*chiami un amico, chiami un'amica* 'you call a boyfriend, you call a girlfriend') to approximate "the choice of a lexical item in relation to a particular concept"; nevertheless, the presence of *eccetera* signals that such approximation "is somehow unsatisfactory" (Ghezzi 2013: 142).

To sum up, previous studies (which did not take into account the diachronic dimension) have focussed mainly on two values of *eccetera*: its referential function (Guil 1999), considering it only "as a form that indicates additional members of a list, set, or category" (Overstreet 1999: 11), and, subsequently, its role in marking vagueness (Cucchi 2010; Voghera 2012; Ghezzi 2013) and also reticence (Galli de' Paratesi 1969). Nevertheless, these two values do not account for the whole range of functions of the form, which will be discussed in the following section.

3.2   Functions of *eccetera* in contemporary spoken Italian

3.2.1   *'There is more': Completing lists and building categories*

According to Guil (1999), *eccetera* is the most frequent general extender in contemporary spoken Italian: as a matter of fact, its 191 occurrences represent 59.3% out of a total of 322 general extenders in the LIP corpus (cf. Fiorentini 2018: 27). Similar to Latin *et cetera* (Section 2.1.1), one of its main functions concerns list completion. In these cases, *eccetera* is generally placed after two or more elements, signalling that 'there is more':

> (20)   *i figli di immigrati hanno tante altre possibilità di rinforzare il loro italiano tramite queste associazioni giornali la radio **eccetera** e quindi anche qualche accenno all'importanza delle comunità italiane in Australia*   (LIP_Naples_A_12_C)
> 'Immigrants' children have many other possibilities to improve their Italian through these organizations, newspapers, the radio etcetera, and therefore [I want] to mention also the importance of Italian communities in Australia.'

In addition, *eccetera* can be used to extend the reference of a given exemplar so as to include other (non-explicit) referents. Therefore, it signals that there are other elements which share a property with the expressed one(s), consisting in the fact that they can co-occur in a frame (*ad hoc* categorization, Barsalou 1983), and it can combine "with a named exemplar (or exemplars), whose characteristics make it possible for the hearer to infer a category the speaker has in mind" (Overstreet 1999: 11). In other words, "[b]y etcetera we mean that there are others, but not any others" (Sacks 1992: 246, quoted by Overstreet 1999: 47):

> (21)   *questa è l'ultima lezione all'ultima lezione possono partecipare anche i parenti amici **eccetera** che vogliono venire*   (LIP_Milan_A_23_D)
> 'This is the last class, the last class can also be attended by relatives, friends, etcetera who want to come.'

The speaker in (21) mentions two examples of people who can attend to a swim class: *parenti* 'relatives' and *amici* 'friends' are therefore members of a larger set characterized by the property 'people who are near and dear to the hearer'. The identification of the property is facilitated by the comparison among the two examples (cf. Barotto 2018), which, together with *eccetera*, triggers an associative inference towards it. The function of *eccetera* is therefore to signal that there are other members belonging to the same category, which will remain unexpressed.

Furthermore, in spoken interactions, the participants can cooperate in the construction of the category, as in (22):

(22) A. *i vari records, e:h che sono: (.) e::h caratterizzati da vari campi, vari fields.*
          quindi autore, (.) titolo dell'opera::=m:h (.) anno di [pubblicazione, soggetto
          ***eccetera,**]*

      B. [*soggetto:, (.) editore,*] (.) [*luogo di pubblicazione*],

      A. [*luogo di pubblicaz*]*ione,(.) e eh che m::h=nsomma l'insieme di questi records,*
          danno vita alla banca dati.   (KIParla)

      'A: The different records, which are characterised by various fields, various
           fields, that is author, title of the book, year of publication, topic etcetera'

      B: Topic, publisher, place of publication

      A: Place of publication, and, well, the combination of these records brings the
           database to life.'

Speaker A first lists four items belonging to a vague category (*vari campi* 'various
fields'), followed by *eccetera*. After that, he overlaps (as indicated by square brackets)
with speaker B, who repeats the last exemplar mentioned by A and adds two more
items. In turn, speakers A repeats the last exemplar cited by B and then concludes
his turn. In this case, the general extender triggers the co-construction (cf. Du Bois
2014) of the category, and speaker B demonstrates her understanding and cooper-
ation by adding other relevant members.

3.2.2 *'You know what i mean': Indexing shared knowledge*
In the examples considered so far, *eccetera* (like *et cetera* in 4–6) entails the existence
of unexpressed members of a category, which share a property with the expressed
ones. Nonetheless, in other cases the form does not point to other elements, but
only to a property, and refers to some knowledge that only the participants in the
interaction share:

(23) B: *e poi è a Ostia 'sta scuola*
      A: *ah*
      B: *morta'*
      A: *ma tu avevi fatto domanda*
      B: *ma l'ho fatto mica quest'anno perché col fatto che stavo ancora al Camilli*
          ***eccetera** l'avevo fatte l'anno scorso le domande*
      A: *ah dall'anno scorso*   (LIP_Rome_B_2_B)
      'B: And, moreover, this school is in Ostia.
      A: Oh!
      B: Damn!
      A: But did you apply…
      B: Well, I didn't apply this year because, since I was still at Camilli etcetera,
           I applied last year.
      A: Oh, last year.'

The speakers in (23) are talking about the school in Ostia[14] where speaker B is going to teach the following school year. To understand what she is referring to with *eccetera*, the hearer must know that the "Camilli" mentioned by B is another school in Ostia, which is quite far from where she currently lives. In similar cases, "the knowledge shared between participants seems completely inaccessible to non-participants" (Overstreet 1999: 70). To some extent, this is similar to the use exemplified in (16): the main difference is that the first function refers to a fixed text, whereas this one usually refers to the common ground of the interlocutors.

Furthermore, it is worth noting that in this example *eccetera* can easily be deleted without changing the propositional content of the segment (Fraser 1999: 944), whereas in the previous ones it cannot. For instance, in (21) the deletion of *eccetera* would indicate that only the four listed exemplars (i.e. *autore, titolo dell'opera, anno di pubblicazione, soggetto* 'author, title of the book, year of publication, topic') were possible.[15] On the contrary, in (23) the general extender functions like a discourse marker, in the sense that it does not contribute "to the propositional meaning of either segment" (Fraser 1999: 944; cf. also Hansen 1998b).

Another interesting example is the following:

(24) *vedete nel primo rigo. gli accordi della A di* (.) "*But not for me*", *sono* (.) *molto semplici*
[plays piano for 16 seconds]
eccetera. okay?   (KIParla)
'You can see it in the first line. The chords of [line] A of 'But not for me' are very simple. Etcetera. okay?'

Here, *eccetera* follows a non-verbal turn, during which the piano teacher plays some chords which he claims to be very simple. After that, he stops and begins to speak again, uttering *eccetera* followed by an interactional discourse marker (i.e. *okay?*) which requests the agreement of the interlocutors. In this case, the form is not simply replacing something that the participants already know, but rather could be paraphrased as "you know what I mean by saying that they are very simple".

To sum up, in Examples ((22)–(23)) *eccetera* marks some degree of shared knowledge (Overstreet 1999: 18), it seems "to underline the shared experience" (Cheshire 2007: 182) of the participants and can be paraphrased as "you know what I mean". Therefore, it has an interactional function, similar to other discourse markers (cf. Pons Borderìa 2006; Hansen 1998a; Hansen 2006). This is consistent with the view of general extenders as primarily "markers of intersubjectivity

---

14. A large district of Rome.

15. However, the non-exhaustivity and the existence of alternatives could also be indicated by prosody.

through which speakers indicate solidarity, self-connection or an assumption of shared experience" (Cheshire 2007: 158).

Contrarily to the other functions of *eccetera*, by which "a speaker implicates a category, so that a hearer can infer additional or alternate members of the category the speaker has in mind" (Overstreet 1999: 66), in this case "the meanings apparently recognized and shared by the participants seem to be unfathomable to anyone else" (Overstreet 1999: 65). Similar interpersonal uses were already present in Latin (Section 2.1.3); nonetheless, while in written texts the suggestion of shared knowledge and common ground is less evident, it emerges quite clearly in spoken data, where it is particularly frequent.

3.2.3 *'There are alternatives': Disjunctive uses*

Occasionally, *eccetera* can signal that 'there are alternatives', i.e. other elements that share a property with the expressed elements, which makes them alternatives to each other. In these cases, it functions like a disjunctive general extender (similar to *o cose del genere* 'or stuff like that'), indicating that the exemplars represent an approximation which "may not be exactly right" (Overstreet 1999: 112):

(25) *vediamo che cosa proponete a chi vi viene a chiedere eh un viaggio per la Norvegia o per la Svezia* **eccetera** *# benissimo giriamo la carta e vediamo un attimo quali itinerari proporreste* (LIP_Florence_C_5_A)
'Let's see what you would propose to somebody who asks you for a trip to Norway, or to Sweden, eccetera… Very well, let's turn the map and see which itinerary you would propose.'

(26) B: *se tu diciamo fai un intervento*
A: *ah*
B: *o per ricoveri* **eccetera** *ti decorre ti decorre dopo dopo sei mesi*
(LIP_Rome_B_7)
'B: If you, say, have surgery…
A: Ah
B: …or for hospital days etcetera, it shall apply after six months.'

In these examples, the speakers offer two alternatives to exemplify the category they are referring to; such alternatives are connected by the disjunction *o* 'or' (Ariel and Mauri 2018). In (25), the speaker, a geography teacher who is examining her students by means of role playing, mentions two possible members of the category she is talking about (i.e. Nordic countries), so that the students can infer other elements. In (26), the participants are discussing the benefits to take out insurance; the examples mentioned by speaker B are two possible cases to which the insurance would apply. In both cases, *eccetera* signals that there are alternatives which will be not explicitly mentioned.

The alternatives can also be mutually exclusive, as in the case of antonyms, exemplified in (27):

(27)   *come avviene la surgelazione altro problema tecnico velocemente lentamente* **eccetera eccetera** *poi il problema sono il trasporto di questi di questi surgelati*
(LIP_Milan_C_11_B)
'How does the freezing happen? [This is] another technical problem, rapidly, slowly, etcetera etcetera, then there is the problem of the transport of these frozen products.'

In this case, the two exemplars (*velocemente* and *lentamente*) represent two opposite freezing paces, i.e. rapidly or slowly, and are incompatible with each other. Nevertheless, since they stand for the two extremes on a scale, they are gradable: therefore, the other possible alternatives indicated by *eccetera* are the intermediate points of deep-freezing pace.

These uses clearly show that the original adjunctive meaning of *eccetera* is becoming opaque. This point will be developed in the next Section.

3.3   Recent developments

The use of *eccetera* in disjunctive contexts is due to the semantic bleaching of its original constituents and represents a further step in its process of grammaticalization. More specifically, a first reanalysis concerns the conjunction *e* 'and' (Latin *et*), which is no longer transparent nor recognizable in the univerbated form. Therefore, the original boundary (Langacker 1977: 119) is recreated by means of adding the conjunction back (*e eccetera, ed eccetera*):

(28)   *nessuno che sappia dove stia 'sta:: porta dell'ade, e abbia visto questo cane a più:: mh=eh questo cane, apparentemente a più teste, con il collo da serpente::: (.)* **ed** **eccetera eccetera**   (KIParla)
'Nobody knows where this Hades door is, and [nobody] has seen this dog with multiple, uhm, this dog, allegedly multi-headed, with a snake neck…and etcetera etcetera.'

The reanalysis is probably caused, or at least facilitated, by the fact that *eccetera* differs from the prototypical structure of Italian general extenders, which are for the most part analytic (Fiorentini 2018: 27) and transparent with respect to "the operations underlying their function" (Mauri 2014: 13), since they have the following structure: [conjunction + proform (+ similative)]. In cases like (28), *eccetera* is reanalysed as a proform (similar to *cose* 'stuff' in *e cose del genere* 'and stuff like that'), with the subsequent addition of an adjunctive conjunction.

The forms [conjunction + *eccetera*] are still quite rare in the corpora of spoken Italian consulted for this research. Nonetheless, some instances can be found in web-based corpora, which represent well colloquial and informal usage of Italian, however written (Fiorentini & Sansò 2019: 108). A simple search of concordances on the ItTenTen corpus on SketchEngine (2016 version, 4.9 billion words)[16] returned 46 occurrences of *e(d) eccetera*:

(29) *Nasce a Torino (Italia) nel 1937. Vive, lavora **ed eccetera** nel capoluogo piemontese.*
'Born in Turin (Italy) in 1937. He lives, works and etcetera in the Piedmontese capital.'

(30) *Mentre gli scienziati fanno complicati calcoli su velocità, traiettoria, massa, accelerazione, resistenza eolica, impatto **ed eccetera simili**, e mentre i politologi riscrivono Machiavelli e discutono prezzi con i moderni principi, lo zapatista si avvicina alla mela, la guarda, l'annusa, la tocca, l'ascolta.*
'While scientists make complicated calculations on speed, trajectory, mass, acceleration, wind resistance, impact and etcetera like that, and while political scientists rewrite Machiavelli and discuss prices with modern princes, the zapatista [militant] approaches the apple, he looks at it, smells it, touches it, listens to it.'

In Example (30), the reanalysis of *eccetera* as a proform is particularly evident. In addition to the conjunction, the writer adds *simili* 'like that' (literally 'similar'), consistently with the above-mentioned structure [connective + proform (+ similative)]. As a result, the general extender becomes more semantically explicit and with a higher degree of internal complexity.

A further step of the reanalysis concerns the loss of the adjunctive meaning ('there is more'). As we have already seen (Examples (25)–(27)), *eccetera* can be used in disjunctive contexts, and, in some cases, the adjunctive value has blurred to the point that *eccetera* is preceded by a disjunction, i.e. *o* 'or'. Although there are no instances of *o eccetera* in LIP and KIParla corpora, we found 13 occurrences in ItTenTen 2016:

(31) *Mi piace anche se spesso non sono d'accordo con quello che scrive, o non mi piace il suo tono **o eccetera**.*
'I like her, although I often disagree with what she writes, or I do not like her tone or etcetera'.

---

16. The Italian Web corpus (itTenTen), which consists of written texts collected on the Internet (websites, forums, blogs, and so on), is part of the TenTen corpus family, i.e. a set of web corpora built up using the same method, with a target size of more than 10 billion words <https://www.sketchengine.eu/ittenten-italian-corpus/>.

(32) *Quindi – riassumendo – ľomosessualità (o bi- **o eccetera**-) è una cosa perfetta-mente in natura, a mio parere.*
'So – to sum up – homosexuality (or bi- or etcetera-) is a perfectly natural thing, in my opinion.'

In (32), the opaqueness of the form is further proved by the fact that it is followed by a hyphen. This means that the writer is treating it like a prefix, exactly as *bi-* in the previous exemplar. Nonetheless, since it is a generic and semantically empty word, it indicates that its slot could be alternatively filled by other prefixes.

4.   Discussion and conclusive remarks

The analysis presented in Section 2 has shown how Latin *et cetera* displays both objective and (inter)subjective functions. Its basic meaning 'there is more' can trigger associative reasoning and similarity inference, which lead to the construction of lists, sets and categories basing on shared knowledge, while the additional meaning 'this is enough' can favour hedging functions at the textual and interpersonal level.

We observe the same multifunctionality in Italian *eccetera* (Section 3), along with further developments. As regards multifunctionality, it is possible to identify three basic meanings, which correlate with three main functions:

LIST COMPLETING – BUILDING CATEGORIES – SHARED KNOWLEDGE (AND RETICENCE)

'there is more' – 'you know what I mean' – 'this is enough'

objective/conceptual > subjective/pragmatic > intersubjective/procedural

----

Figure 2. Functions and meanings of Italian *eccetera*

Such functions and meanings are for the most part consistent with those of Latin *et cetera* (see Figure 1), with some differences. In particular, in more recent data, and especially in oral speech, the interpersonal functions of *eccetera* are clearly deployed: (i) it can be used to (co)-construct categories in interaction (Example (22)); (ii) it can index shared knowledge between the interlocutors. In the latter case, it does not invite the hearer to infer other elements in addition of the one expressed, rather pointing to something which is supposed to be present in the common ground of the interlocutors. Similar intersubjective meanings were already present in Latin, although they appear to be less prominent than the other functions, due to the nature of written texts, which prevents us from grasping the extent of the

pragmatic functions of *et cetera*. Furthermore, spoken interaction generally depends more highly on context, and is more deeply rooted in the common ground shared by speakers.

On the other hand, we have observed that the more recent developments of *eccetera* are caused by the opaqueness of the form (probably also due to its frequency in speech; cf. Fiorentini 2018). Although these instances are still quite rare, Examples ((28)–(32)) clearly show that *eccetera* is undergoing a further process of semantic bleaching, since it can be used as a proform (like *cose* 'stuff') and, as such, it can form analytic general extenders (*ed eccetera simili* 'and eccetera like that', Example (30)).[17] Therefore, we can schematise the evolution of the form as in Figure (3):

(et) cetera > *etcetera* > *eccetera* > *e(d) eccetera*
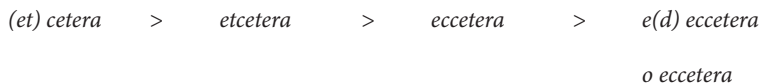
o eccetera

Figure 3. The development of the form

The figure shows how, after the univerbation and the subsequent assimilation (*etcetera > eccetera*), in Old Italian, the form has been further reanalysed in contemporary Italian. More specifically, the conjunction *e* 'and' has blurred to the point that *eccetera* can be preceded by another conjunction (*e(d) eccetera*) or even a disjunction (*o eccetera*).[18] In the latter case, the form signals that there are alternatives, thus losing its adjunctive meaning and functioning as a disjunctive general extender (similar to *o cose del genere* 'or stuff like that').

In conclusion, we have observed a semantic-pragmatic expansion of both Latin *et cetera* and Italian *eccetera*, which have developed intersubjective and procedural meanings besides the original (objective and conceptual) functions. Finally, in contemporary Italian *eccetera* is undergoing a further development, due to the opaqueness of its components, which is leading to the creation of new analytic general extenders: *e(d) eccetera, o eccetera*.

---

17. Since they are semantically empty, such proforms can be preceded either by a conjunction or a disjunction; besides *e/o cose così* 'and/or things like that' (Fiorentini 2018), see for instance also French (*et/ou) machin* '(and/or) stuff' (Béguelin & Corminboeuf 2017).

18. Interestingly, this is also the case of English *etcetera*, which sometimes can be preceded by *and* or *or*, as in the following examples (extracted from EnTenTen 2013 corpus, <https://www.sketchengine.eu/ententen-english-corpus/>):

1. But, depending on your background you will add other services to stay competitive and also to utilize what you know: website design, Internet research, article writing and submission, event planning **and etcetera**.
2. And since so many of us like to fit in workouts during our lunch breaks **or etcetera**, we often don't think about what we should eat.

Acknowledgements

References

Ariel, Mira & Mauri, Caterina. 2018. Why use or? *Linguistics* 56(5): 939–993. https://doi.org/10.1515/ling-2018-0020

Barotto, Alessandra. 2018. The role of exemplification in the construction of categories: The case of Japanese. In *Linguistic Strategies for the Construction of Ad Hoc Categories: Synchronic and Diachronic Perspectives*, Caterina Mauri & Andrea Sansò (eds). Special issue of *Folia Linguistica* 52(s39–1): 37–68. https://doi.org/10.1515/flih-2018-0002

Barsalou, Lawrence W. 1983. Ad hoc categories. *Memory and Cognition* 11(3): 211–227. https://doi.org/10.3758/BF03196968

Béguelin, Marie-José & Corminboeuf Gilles. 2017. *Ou comme ça, machin* et autres marqueurs d'indétermination dans les listes. *Discours* 20. https://doi.org/10.4000/discours.9275

Channell, Joanna. 1994. *Vague Language*. Oxford: OUP.

Cheshire, Jenny. 2007. Discourse variation, grammaticalization, and stuff like that. *Journal of Sociolinguistics* 11(2): 155–193. https://doi.org/10.1111/j.1467-9841.2007.00317.x

Cortés Rodríguez, Luis. 2006. Los elementos de final de serie enumerativa en el discurso oral. *Investigaçoes* 19(2): 9–36.

Croft, William & Cruse, Alan D. 2004. *Cognitive Linguistics*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511803864

Cucchi, Costanza. 2010. Vague expressions in the European Parliament: A marker of cultural identity? In *Discourse, Identities and Roles in Specialized Communication*, Giuliana Garzone & James Archibald (eds), 85–107. Bern: Peter Lang.

De Mauro, Tullio, Mancini, Federico, Vedovelli, Massimo & Voghera, Miriam. 1993. *Lessico di frequenza dell'italiano parlato*. Milano: Etaslibri.

Dines, Elizabeth R. (1980). Variation in discourse: 'And stuff like that'. *Language in Society* 9(1): 13–31. https://doi.org/10.1017/S0047404500007764

Diewald, Gabriele. 2011. Pragmaticalization (defined) as grammaticalization of discourse functions. *Linguistics* 49(2): 365–390. https://doi.org/10.1515/ling.2011.011

Du Bois, John W. 2014. Towards a dialogic syntax. *Cognitive Linguistics* 25(3): 359–410. https://doi.org/10.1515/cog-2014-0024

Fiorentini, Ilaria. 2018. *Eccetera eccetera e così via di seguito*. I *general extenders* dell'italiano contemporaneo. *CLUB Working Papers in Linguistics* 2: 20–39.

Fiorentini, Ilaria. 2019. E le rimanenti cose. *Eccetera* tra reticenza e inferenza. In *Spazi bianchi. Le espressioni letterarie, linguistiche e visive dell'assenza*, Alfonsina Buoniconto, Gerardo Salvati & Raffaele Cesaro (eds), 249–260. Soveria Mannelli: Rubbettino.

Fiorentini, Ilaria & Sansò, Andrea. 2019. Reformulation markers and their functions. Two case studies from Italian. *Journal of Pragmatics* 120: 54–72. https://doi.org/10.1016/j.pragma.2017.08.010

Fraser, Bruce. 1999. What are discourse markers? *Journal of Pragmatics* 31: 931–952. https://doi.org/10.1016/S0378-2166(98)00101-5

Galli de' Paratesi, Nora. 1969. *Le brutte parole*. Milano: Mondadori.

Ghezzi, Chiara. 2013. Vagueness Markers in Contemporary Italian: Intergenerational Variation and Pragmatic Change. PhD Dissertation, Università di Pavia.

Goria, Eugenio & Mauri, Caterina. 2018. Il corpus KIParla: Una nuova risorsa per lo studio dell'italiano parlato. In *CLUB Working Papers in Linguistics*, Vol. 2, Francesca Masini & Fabio Tamburini (eds), 96–116. Bologna: Circolo Linguistico dell'Università di Bologna (CLUB).

Grice, Paul. 1975. Logic and conversation. In *Syntax and Semantics, Vol. 3: Speech Acts*, Peter Cole & Jerry L. Morgan (eds), 41–58. New York NY: Academic Press.

Guil, Pura. 1999. Proformas alusivas en lengua oral. In *Linguistica testuale comparativa. In Memoriam Maria-Elisabeth Conte*, Gunver Skytte & Francesco Sabatini (eds), 89–98. Copenhagen: Museum Tusculanum Press.

Hand, Ferdinand G. 1832. *Tursellinus seu de particulis Latinis commentarii*, Vol. II. Leipzig: Weidemann.

Hansen, Maj-Britt Mosegaard. 1998a. *The Function of Discourse Markers. A Study with Special Reference to Spoken Standard French* [Pragmatics & Beyond, New Series 53]. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.53

Hansen, Maj-Britt Mosegaard. 1998b. *The semantic status of discourse markers. Lingua* 104: 235–260.

Hansen, Maj-Britt Mosegaard. 2006. A dynamic polysemy approach to the lexical semantics of discourse markers (with an exemplary analysis of French *toujours*). In *Approaches to Discourse Particles*, Kerstin Fischer (ed.), 21–46. Amsterdam: Elsevier.

Hopper, Paul & Traugott, Elizabeth C. 2003. *Grammaticalization* (2nd edn). Cambridge: CUP. https://doi.org/10.1017/CBO9781139165525

Jefferson, Gail. 1990. List-construction as a task and resource. In *Interaction Competence*, George Psathas (ed.), 63–92. Washington DC: University Press of America.

Langacker, Ronald W. 1977. Syntactic reanalysis. In *Mechanisms of Syntactic Change*, Charles N. Li (ed.), 57–139. Austin TX: University of Texas Press.

Magni, Elisabetta. 2019. General extenders in Latin. In *Lemmata linguistica Latina, Vol. 2: Clause and Discourse. Proceedings of the 19th International Colloquium on Latin Linguistics (Munich April 24–28, 2017)*, Lidewij van Gils, Caroline Kroon & Rodie Risselada (eds), 697–714. Berlin: De Gruyter. https://doi.org/10.1515/9783110678222-013

Mauri, Caterina. 2014. What do connectives and plurals have in common? The linguistic expression of ad hoc categories. In *Linguistic Papers Dedicated to Jacques Moeschler*, Joanna Blochowiak, Stéphanie Durrlemann-Tame, Cristina Grisot & Christopher C. Laenzlinger (eds), 1–21. Genève: Université de Genève.

O'Keeffe, Anne. 2004. 'Like the wise virgins and all that jazz': Using a corpus to examine vague categorization and shared knowledge. *Language and Computers* 52(1): 1–26.

Overstreet, Maryann. 1999. *Whales, Candlelight and Stuff Like That: General Extenders in English Discourse*. Oxford: OUP.

Overstreet, Maryann. 2014. The role of pragmatic function in the grammaticalization of English general extenders. *Pragmatics* 24: 105–129. https://doi.org/10.1075/prag.24.1.05ove

Pichler, Heike & Levey, Stephen. 2011. In search of grammaticalization in synchronic dialect data: General extenders in northeast England. *English Language and Linguistics* 15: 441–471. https://doi.org/10.1017/S1360674311000128

Pons Bordería, Salvador. 2006. A functional approach to discourse markers. In *Approaches to Discourse Particles*, Kerstin Fischer (ed.), 77–99. Amsterdam: Elsevier.

Prandi, Michele. 1990. Una figura testuale del silenzio: La reticenza. In *Dimensioni della linguistica*, Maria-Elisabeth Conte, Anna Giacalone Ramat & Paolo Ramat (eds.), 217–239. Milano: Franco Angeli.

Sacks, Harvey. 1972. On the analyzability of stories by children. In *Directions in Sociolinguistics: The Ethnography of Communication*, John Gumperz & Dell Hymes (eds.), 325–345. New York NY: Holt, Rinehart and Winston.

Tombeur, Paul (ed.). 2013. *LLT-Series A. Library of Latin Texts*. Turnhout: Brepols.

Traugott, Elizabeth C. 2010. Revisiting subjectification and intersubjectification. In *Subjectification, Intersubjectification and Grammaticalization*, Kristin Davidse, Lieven Vandelanotte & Hubert Cuyckens (eds), 29–70. Berlin: De Gruyter.

Voghera, Miriam. 2012. Chitarre, violino, banjo e cose del genere. In *Per Tullio De Mauro. Studi offerti dalle allieve in occasione del suo 80°compleanno*, Anna M. Thornton & Miriam Voghera (eds), 341–364. Roma: Aracne.

Voghera, Miriam, Iacobini, Claudio, Savy, Renata, Cutugno, Francesco, De Rosa, Aurelio & Alfano, Iolanda. 2014. VoLIP: A searchable Italian spoken corpus. In *Complex visibles out there. Proceedings of the Olomouc Linguistics Colloquium: Language Use and Linguistic Structure*, Ludmila Veselovská & Markéta Janebová (eds), 628–640. Olomouc: Palacký University.

Morphopragmatics of rhyming and imitative co-compounds in Russian

Anna Alexandrova and Valentina Benigni

Sapienza University of Roma / University of Roma Tre

Coordinative compounding is a phenomenon at the crossroads between word-formation, syntax and discourse, adopted for encoding non-lexicalized concepts in a wide range of languages. We propose a morphopragmatic account of two specific coordinative compounding strategies in Russian, namely, rhyming and imitative co-compounds. They exhibit non-compositional semantics, inasmuch as they perform a generalizing/categorizing function, which, in turn, is tightly intertwined with a strong pragmatic evaluative component, involving irony and downgrading of the referent. Furthermore, different means of categorization tend to cluster in discourse, thus resulting in pervasive overcoding of the respective *ad hoc* concepts. Since categorization is a result of an online process of meaning negotiation between speakers, we have to do with low-frequency single instances of a frequent phenomenon.

Keywords: coordinative compounding, rhyming co-compounds, imitative co-compounds

1. Introduction

1.1 Co-compounds and their morphosyntactic properties in Russian

Russian relies on several types of compounds as productive processes of word-formation. In this paper, we focus on one of the most frequent processes, which is coordinative compounding.

Coordinative compounds, or co-compounds in Wälchli's (2005) terms, are formed by mere juxtaposition of at least two conceptually related lexical units, belonging to the same word class. In Russian, these formations are predominantly noun-based (*chleb-sol'* [bread.SG.M-salt.SG.F] 'hospitality'), although verbal compounds are also rather common (*est'-pit'* [eat.IPFV.INF-drink.IPFV.INF]

'eat and drink; perform activities involving the consumption of food and drinks'; *poit'-kormít'* [water.IPFV.INF-feed.IPFV.INF] 'wine and dine', *odevát'-obuvát'* [dress. IPFV.INF-put.on.shoes.IPFV.INF] 'buy clothes, shoes and other necessary things for someone; provide for someone'). The meanings of these compounds are not a mere sum of the lexical meanings of their bases, inasmuch as these formations trigger a metaphorical or metonymical interpretation, with a varying degree of bleaching, and serve a categorizing function. In sum, they are not straightforwardly compositional. From a pragmatic and discourse perspective, they often have evaluative meanings, associated with metarepresentational/figurative uses of language, such as irony, sarcasm, and humour.

The aim of this paper is to provide a morphopragmatic account of two specific types of co-compounds in Russian, namely, rhyming and imitative co-compounds. The focus of the study is on how these morphological patterns convey the categorizing function, as well as on the contribution of iconicity to the construction of complex concepts and to the expression of evaluative meanings. Our starting point is the assumption that (morpho)syntactic constructions can have "regular pragmatic meanings" (Dressler & Barbaresi 1994: 58). However, pragmatic meaning is not fully predictable either, in that it strongly depends on the speech situation and on the speaker's intention. Hence, discourse-level aspects should not be overlooked in the analysis of the phenomenon. We view rhyming and imitative co-compounds as word-formation strategies that contribute to the construction of the sentence-level pragmatic meaning and constitute patterns in discourse organization.

Furthermore, we analyse rhyming and imitative co-compounds as morphological means of *ad hoc* categorization. *Ad hoc* categories are "goal-driven abstractions, which respond to the need to categorize the world under particular contextual circumstances" (Mauri & Sansò 2018: 2).[1] We discuss the respective underlying pragmatic processes from the perspective of Relevance Theory (Wilson & Carston 2007; Carston 2010). We depart from its central claim that human cognition tends to maximize relevance, and comprehension of utterances is guided by a selective process. Thus, in processing any new input, whether it is derived from external reality or from internal representations, human minds pick out only those subsets of information that are most relevant to a specific context. We view the co-compound components as exemplars of a target category. In order to process the category correctly, the receiver resorts to contextual and background knowledge and selects only those properties of the components that are relevant.

Wälchli (2005: 1) defines co-compounds as "word-like units consisting of two or more parts which express natural coordination". If we take a glance at the respective Russian linguistic terminology, co-compounds are mainly found in the

---

1. For an in-depth discussion of *ad hoc* categories, see Barsalou (1983).

literature under the labels *parnye sočetanija* 'pair <word> combinations' (Plotnikova 2012) and *parnye slova* 'pair words' (Minlos 2004; Pacsai 2013). Furthermore, Weiss (1993, 2013a, 2013b) adopts the term *dvojnye glagoly* 'double verbs' for verbal co-compounds, such as *sidit rabotaet* [sit.IPFV.PRS.3SG work.IPFV.PRS.3SG] '(s/he) is sitting (and) working', which passed somewhat unobserved in the older accounts of Russian compounding strategies. Such compound verbs formally resemble serial verb constructions and encode complex or situationally related events.

From a structural point of view, coordinative compounds are similar to appositional compounds (e.g. *diván-krovát'* [sofa.SG.M-bed.SG.F] 'sofa bed', *chudóžnik-restavrátor* [painter.SG.M-restorer.SG.M] 'painter-restorer', etc.). However, their semantic structures differ in several ways that are described below. In well-formed appositional N+N compounds, the nominal constituents do not refer to a superordinate concept, even though they belong to the same semantic frame. For instance, both *diván* 'sofa' and *krovát'* 'bed' are pieces of furniture for sleeping and/or lying, but the compound *diván-krovát'* does not give rise to a superordinate concept of the type *[PIECES OF FURNITURE FOR SLEEPING, INCLUDING BEDS, SOFAS AND THE LIKE]. On the contrary, the lexical meaning of an appositive co-compound, as a whole unit, is more specific than the meaning of each of its constituents. As Wälchli (2005: 76) put it, they are rather "referentially intersective, as both […] denote a single referent".

Co-compounding used to be a notoriously widespread formulaic language strategy in Russian folklore, where it was used both as a mnemonic technique and as an ornamental device.[2] Most of these co-compounds belong to the synonymic sub-type (*naród-ljúdi* [people.SG.M-people.PL.M] 'people', *put'-doróžka* [way.SG.M-path.DIM.SG.F] 'way', *pečál'-toská* [sadness.SG.F-anguish.SG.F] 'sadness and anguish', *porá-vrémja* [time.SG.F-time.SG.N] 'time', *otéc-bátjuška* [father.SG.M-father.SG.M] 'father', *muravá-travá* [young.grass.SG.F-grass.SG.F] 'grass' etc.).[3] In this subtype, the second constituent of the co-compound does not contribute new information to the meaning of the first constituent. It rather serves the purpose of generalization, thus emphatically repeating the propositional meaning of the first noun.

Besides synonymic co-compounds, another pattern, common in folklore, is presented by what Wälchli (2005: 141–143) terms the 'collective' co-compound. Its

---

2. Numerous instances thereof are attested in (but are not limited to) *byliny*, traditional East Slavic oral epic poems.

3. When the oral tradition was still alive and improvisation was an essential part of the genre, the use of the pattern was not limited to the standard repertoire of fossilized forms (Lipec 1939). New co-compounds, such as, for instance, *sila-armija* [army.SG.F-army.SG.F] 'army' in the recordings of M. S. Krjukova's performances (Borodina & Lipec 1939), were coined by the singers under the influence of other genres of oral poetry, such as soldiers' songs.

components are co-hyponyms, such as, for instance, *zláto-sérebro* [gold.SG.N-silver. SG.N] 'precious metals', *knjázi-bojára* [knyaz.PL.M-boyar.PL.M] 'noblemen, nobility', *medvédi-vólki* [bear.PL.M-wolf.PL.M] 'dangerous wild animals', *chleb-sol'* [bread. SG.M-salt.SG.F] 'hospitality'.[4] Both constituents equally contribute to the overall meaning of the compound, which, ultimately, refers to a superordinate concept.

From a syntactic-semantic perspective, then, the relation between the constituents of a co-compound is regarded by Wälchli (2005) as an instance of natural coordination. Unlike accidental coordination, natural coordination implies that "there is a close and salient relation between the entities denoted by the conjuncts" (Dalrymple & Nikolaeva 2006: 830). At first glance, the contrast between accidental and natural coordination seems to reside in the lexical/semantic and pragmatic domains. However, earlier studies (Corbett 1983; Dalrymple & Nikolaeva 2006) have shown that the contrast in question has reflexes in morphosyntactic patterns, such as agreement, in a range of languages, among which Russian, Kurdish, and Tundra Nenets.[5] More specifically, in Russian, natural coordination requires a plural-marked adjective, as shown in (1)–(2), while accidental coordination constructions exhibit closest-conjunct agreement (3):

(1) *Do čego že   glupy   byli   **vaši**   **amerikanskie***
    just.so EMPH stupid.PL be.PST.PL POSS2PL.NOM.PL American.NOM.PL
    babuška   *i   **deduška**,*   *esli priechali*
    grandmother.NOM.SG.F and grandfather.NOM.SG.M if   arrive.PFV.PST.PL
    sjuda.
    here
    'It was just so stupid of your American grandmother and grandfather to come here.'   [RNC. E. Changa. *Pro vsë*, 2000]

(2) *Neobchodimo vynesti   na svalku   **starye***
    is.necessary   take.away.PFV.INF on landfill.ACC.SG.F old.ACC.PL
    divan   *i   **krovat'**.*
    sofa.ACC.SG.M and bed.ACC.SG.F
    'It is necessary to take an old sofa and (an old) bed to a landfill.'
                                       [Web, <https://molodechno.kabanchik.by/tag/
                                       vyvoz-staroj-mebeli-na-svalku>]

---

4. For further examples, see Pacsai 2013.

5. Le Bruyn and de Swart (2012) also account for bare coordination as a manifestation of natural coordination. However, this is not a directly applicable criterion for Russian, in that it does not overtly express definiteness of a noun phrase with a grammaticalized device, such as articles, and apparent bare nominals may receive both generic and definite interpretations, depending on the syntactic structure. The parameter of definiteness is also intertwined with the verbal aspect in a complex way, but this topic is too broad to be discussed here.

(3) *Čajf*   *vypustili*   **novyj**   **klip**
    <band name>.NOM release.PFV.PST.PL new.ACC.SG.M music.video.ACC.SG.M
    i   *pesnju.*
    and song.ACC.SG.F
    'The Čajf band has released **a new music video and a (new) song**.'
        [Web, <https://piter.tv/event/Video_Chajf_vipustili_novij_klip_/>]

However, Dalrymple and Nikolaeva (2006) also observe that the same entities can be alternatively conceptualized as instances of natural or accidental coordination in different contexts and hence exhibit both patterns of adjectival agreement, which somewhat weakens the authors' position. For instance, compare (4)–(7):

(4) *Možete*   *priglasit'*   *znakomych*   **so**
    can.IPFV.PRS.2PL invite.PFV.INF acquaintance.ACC.PL with
    svoej   **koškoj**   *i*   *sobakoj*   *k vam*   *v*
    REFLPOSS/own.INS.SG.F cat.INS.SG.F and dog.INS.SG.F to 2PL.DAT in
    gosti…
    guest.ACC.PL.M
    'You can invite acquaintances with **their own cat and dog** to your place.'
        [Web, <https://www.toybytoy.com/stuff/If-you-decide-to-have-a-pet>]

(5) *Ja*   **so**   **svoimi**   **koškoj**   *i*
    1SG.NOM with REFLPOSS/own.INS.own.INS.PL cat.INS.SG.F and
    sobakoj   *tak*   *i*   *ezžu.*   *Tak*   *oni*
    dog.INS.SG.F this.way EMPH drive.IPFV.PRS.1SG this.way 3PL.NOM
    vedut   *sebja*   *gorazdo spokojnee.*
    behave.IPFV.PRS.3PL REFL.ACC much calm.COMP
    'I keep traveling with **my cat and dog**. It makes them feel safer/more relaxed.'
        [Web, <http://ru-cats.livejournal.com/21409262.html>]

(6) *on*   […] *i*   *napoil*   *ostal'nuju*
    3SG.NOM.M   and water.PFV.PST.SG.M remaining.ACC.SG.F
    našu   *kompaniju*   *obaldenno vkusnym*   *čaem*   *s*
    POSS1PL.ACC.SG.F gang.ACC.SG.F incredibly tasty.INS.SG.M tea.INS.SG.M with
    pečenjuškami,   *i*   *pokormil*   **svoich**
    cookies.INS.PL.F and feed.PFV.PST.SG.M REFLPOSS/own.ACC.PL
    košku-sobaku   […]
    cat.ACC.SG.F-dog.ACC.SG.
    'He treated the whole gang to an incredibly tasty tea with cookies and fed **his cat and dog**.'
        [Web, <https://www.tury.ru/otzyv/id/158003-indiya-kalangut-dona-
          terezinha-3-severnyy-goa-indiya-otel-dona-terezinha-2–3>]

(7)   *Kogda-to   v   molodosti,   iz-za   moego*
      some.time in youth.PREP.SG.F because.of my.GEN.SG.M
      vsechžaljuščego   charaktera,   u   nas   v
      sorry.for.everyone.GEN.SG.M personality.GEN.SG.M at/by 1PL.GEN in
      kvartire   často pojavljalis'   vsjakie
      flat.PREP.SG.F often appear.IPFV.REFL.PST.PL various.NOM.PL
      koški-sobaki-ptički   ⟨…⟩
      cat.NOM.PL.F-dog.NOM.PL.F-bird.DIM.NOM.PL.F
      'When I was young, because of my feeling-sorry-for-everyone personality, there
      were often **various cats, dogs, birds (and the like)** in our flat ⟨…⟩'
                                      [Web, ⟨http://pikabu.ru/story/
                                      koshka_sobaka_rebenok_vpridachu_5847082⟩]

In (4) above, the adjectival agreement is controlled by the adjacent noun, which apparently suggests that it is a case of accidental coordination, whereas in (5), the same pair of coordinate nouns is assumed to behave as an instance of natural coordination, with plural agreement on the possessive adjective. However, the context does not provide any hints that the relation of coordination in (4) is really conceptualized as accidental. Finally, in (6) and (7), we have two co-compounds with different morphosyntactic and semantic properties. The co-compound in (6), triggering plural agreement on the adjective, is a representation of two specific animals, a cat and a dog, as a single entity with plural reference. In (7), the co-compound is built upon a non-exhaustive list of three plural-marked nominal constituents, constructing the *ad hoc* category [ANIMALS IN NEED OF HELP]. The above examples show that neither coordinate noun phrases nor co-compounds behave, syntactically and semantically, in a uniform way. Apparently, the distribution of agreement patterns is more of a tendency rather than a rule. There is a gradient between syntactic coordination and word formation. Moreover, there is variation with respect to levels of semantic cohesion between the constituents. The co-compound *kóška-sobáka* [cat.SG.F-dog.SG.F] 'cat (and) dog' in (6) refers to the domestic animals as a whole entity. Its reference is characterized by specificity. Furthermore, in (7), the process of categorization gets on a new level of abstraction, as the co-compound *kóški-sobáki-ptíčki* [cat.NOM.PL.F-dog.NOM.PL.F-bird.DIM.NOM.PL.F] 'cats, dogs, birds (and the like)' contextually refers to a potentially unlimited set of [+animate] entities, which is pragmatically restricted by two major properties, (a) the fact of belonging to the class of animals, and (b) being in need of help.

There may be more than one functional explanation of the fact that co-compounds, derived from singular-marked noun bases, can trigger both singular and plural agreement. Firstly, singular agreement is logically coherent with the conceptual unity of the co-compound, which is expected to be stronger in the case of a lexicalized/conventionalized entity, as compared to syntactically coordinated

conjuncts. Secondly, besides semantic/pragmatic criteria, a purely morphological factor may be responsible for the variation. Co-compounds that present combinations of different-gender nouns may be more prone to plural agreement as a strategy of avoidance of gender assignment by the speaker,[6] inasmuch as plural adjectives are not marked for gender in Russian. For instance, compare (8) and (9) below, where the co-compounds are respectively modified by a plural and a singular adjective:

(8) *I*   ***imja-familija***   *kakie-to*
     and first.name.NOM.SG.N-last.name.NOM.SG.F somewhat.NOM.PL
     čudnye,   *i*   *nacija*   *strannaja.*
     weird.NOM.PL and nationality.NOM.SG.F strange.NOM.SG.F
     'The name and surname are **somewhat weird**, and the nationality is strange.'
     [RNC. N. Bogoslovskij. *Zametki na poljach šljapy*, 1997]

(9) *A* *svoj-to*   *dom,*   *svoja*
     but own.NOM.SG.M-EMPH house.NOM.SG.M own.NOM.SG.F
     čaška-ložka   *uže*   *i*   *nenužnye*   *tebe?*
     cup.NOM.SG.F-plate.NOM.SG.F already and.EMPH useless.NOM.PL 2SG.DAT
     'But do you (really) no longer need you own house, **your own possessions** (*lit.* cup-plate)?'   [RNC. S. Zalygin. *Komissija*, 1976]

In (8), the bases of the co-compound *ímja-famílija* [first.name.NOM.SG.N-last. name.NOM.SG.F] 'first name (and) last name' pertain to different genders (they are neuter and feminine, respectively) and trigger plural agreement on the adjective. Conversely, in (9), the co-compound *čáška-lóžka* [cup.NOM.SG.F-plate.NOM.SG.F] is derived from same-gender bases, and the adjective exhibits singular agreement.[7]

    When the 'close and salient' relation, typical of natural coordination, is realized at the word-formational level, different paradigmatic relations between the constituents of compounds, going beyond synonymy and co-hyponymy, are possible. For instance, the conjuncts can instantiate binary opposites, particularly, those related to social roles, as in the names of common role-play games (*kóški-mýški* [cat. PL.F-mouse.PL.F] 'cat-and-mouse game'; *dóčki-máteri* [daughter.DIM.PL.F-mother. PL.F] 'house game', *s"edóbnoe-nes"edóbnoe* [edible.SG.N-indedible.SG.N] 'edible or inedible', *kazakí-razbójniki* [Cossack.PL.M-robber.PL.M] 'Cossacks and Robbers' etc.). Noun bases denoting time spans can encode a chronological relationship (*subbóta-voskresén'e* [Saturday.SG.F-Sunday.SG.N] 'week-end'), in turn, intrinsically

---

6. This topic needs to be addressed in a dedicated experimental study, which is beyond the scope of our paper.

7. An analogous situation is observed in (4), where two same-gender coordinate nouns trigger singular agreement on the adjective: *so svoej koškoj i sobakoj* [with REFLPOSS/own.INS.SG.F cat. INS.SG.F and dog.INS.SG.F] 'with their own cat and dog'.

associated with temporal approximation (*gde-to v apréle-máe* [somewhere in April. PREP.SG.M-May.PREP.SG.M] 'in April or May; some time between April and May'). Both verbal and nominal (deverbal) compounds can be based on a reversive relationship (*ssórit'sja-mirít'sja* [argue.IPFV.INF.REFL-make.peace.IPFV.INF.REFL] 'have a stormy relationship', *vdoch-výdoch* [inhalation.SG.M-exhalation.SG.M] 'inhalation and exhalation act', *otkrýtie-zakrýtie* [opening.SG.N-closing.SG.N] 'opening and closing, open-and-shut' etc.). The bases of a reversive co-compound are usually ordered iconically and, hence, reflect the temporal causal relationship between the events. Finally, co-compounds often present, in Barsalou's (1985) terms, goal-derived categories. Such categories are composed of entities that serve the same goal or plan, while the exemplars of the category *per se* may bear no surface resemblance to each other and lack common features. For instance, money and an identity card may be conceptualized in a certain context as exemplars of the goal-derived category [OBJECTS NEEDED FOR TRAVEL], as in (10), where we have the co-compound *dén'gi-dokuménty* [money.NOM.PL.F-document.NOM.PL.M] 'money and documents':

(10) *A   vsjakie   **den'gi-dokumenty**   pri tebe*?
     and various.NOM.PL money.NOM.PL.F-document.NOM.PL.M by 2SG.PREP
     'Do you have **money, documents and the like** on you?'
                        [Web, <http://restonweb.livejournal.com/338484.html>]

Although co-compounding is not a frequent phenomenon in 'standard' Russian (Wälchli 2005: 204), it exhibits a relatively high level of productivity in colloquial speech. Specifically, it is widely exploited as a discourse strategy for naming non-lexicalized conceptual entities, which can be activated by the context.

The categorizing function is also a typical semantic/pragmatic feature of rhyming and imitative co-compounds, predominantly occurring in colloquial speech. One salient difference between this type of co-compounds and the other aforementioned types is concerned with their formal structure. Specifically, rhyming and imitative co-compounds follow a relatively rigid template. Rhyming co-compounds, besides the word-final rhyme, often exhibit repetition of segments (cf. *chólod-gólod* [hunger.SG.M-cold.SG.M] 'poverty and starvation', *véliki-róliki* [bike.PL.M-skate.PL.M] 'amateur wheel sports' etc.). As far as imitative co-compounds are concerned, they are built upon a content word, followed by an echo form, which is phonologically based on the first constituent and is devoid of independent lexical meaning. The (morpho)phonological operations applied to the echo form are language specific. In Russian, they consist in copying the phonological material of the meaningful component and substituting the onset of the first syllable with *m-*, *shm-* (e.g., *dokazatel'stva-šmokazatel'stva* [evidence.PL.N-SHM.evidence.PL.N] 'evidence and all that stuff') or an obscene root morpheme.

Co-compounds can be considered an example of gradience between syntax and morphology as structures of uncertain status, one foot in morphological compounds and one in asyndetic binomial constructions. It is a case of intersective gradience (Aarts 2007: 94), which obtains when two categories converge due to a shared set of properties. Many co-compounds are non-exhaustive list constructions (in the Construction Morphology sense) with a categorizing, non-compositional meaning, operating at the lexical/morphological level (Masini, Mauri & Pietrandrea 2018). In written discourse, co-compounds are subject to orthographic variation, which reflects their categorial indeterminacy, as well as their informal, uncodified status in the 'standard' language system. These formations may be hyphenated or written separately, as juxtaposed words.[8] Consider the structures containing three or more coordinate members, as in the following example:

(11) *Tože s   udovol'stviem   by   tuda s"ezdila*   [...] *v*
      too   with pleasure.INS.SG.N COND there go.PFV.PST.SG.F   in
      italiju   [...] *v gory   pervym   delom,*
      Italy.ACC.SG.F   in mountain.ACC.PL.F first.INS.SG.N thing.INS.SG.N
      nu   i   voobšče proechat'sja   po
      EMPH and in.general take.a.ride.PFV.INF.REFL across/about
      rimam-milanam-venecijam-florencijam)))
      Rome.INS.PL.M-Milan.INS.PL.M-Venice.INS.PL.F-Florence.INS.PL.F
      'I would love to go there [...], to Italy [...], first thing in the mountains, and in general [I would] take a ride throughout **Rome, Milan, Venice, Florence and the like**.'   [Web]

In the above context, the co-compound *rímy-milány-venécii-floréncii* [Rome.NOM.PL.M-Milan.NOM.PL.M-Venice.NOM.PL.F-Florence.NOM.PL.F] is built upon the names of four Italian cities, commonly viewed as prototypical tourist destinations. The categorizing nature of this formation, referring to the higher-level category [FAMOUS ITALIAN ART HERITAGE CITIES], is suggested by the plural form of the proper names. However, theoretically speaking, this form could be alternatively analysed as an asyndetic open-list syntactic construction with the same semantic/pragmatic functions.

Naturally, the problem of the syntax-morphology gradience and criteria for wordhood is also closely related to the productivity issue. On the one hand, these formations are rather frequent, as a class, in discourse, which means that they present a productive pattern. However, highly lexicalized items co-exist with a wide

---

8.   Be that as it may, orthographical conventions are of little help in establishing their categorial status.

range of scarcely lexicalized, context-related formations that do not belong to the speakers' common lexical stock. A situation, in which a frequent morphological pattern involves a large number of *hapax legomena* and low-frequency items, is potentially a sign of a high expanding productivity (Baayen 1993).

Another salient parameter for deciding the morphosyntactic status of co-compounds is their inflectional pattern. In subordinative compounds (e.g., *bíznes-škóla* 'business school'), whereby a complement relation holds between the components, the syntactic dependency is marked on the head (e.g., *s bíznes-škóloj* [with business. NOM.SG.M-school.INS.SG.F] 'with a business school'), which is inflected for case. By contrast, in co-compounds, both constituents exhibit the same case marking (*bez divána-krováti* [without sofa.GEN.SG.M-bed.GEN.SG.F] 'without a sofa-bed'). This feature makes co-compounds similar to a syntactic structure, namely, a sequence of coordinated lexemes.

Finally, as far as the semantic criteria for wordhood are concerned, one could say that rhyming and imitative co-compounds are semantically non-compositional, and non-compositionality is generally perceived as a feature that is more prototypical for compound words rather than for phrases. However, there are numerous syntagmatic phenomena exhibiting semantic non-compositionality, and hence, the semantic criterion itself is not reliable enough (Haspelmath 2011).

1.2   Data, theoretical framework, and scope of the study

Our data were obtained from several sources. Rhyming and imitative co-compounds were extracted from the Russian National Corpus (RNC),[9] as well as from Araneum Russicum Maius (ARM) (Benko 2014). Moreover, a large set of examples was collected from the web by searching specific items via Google. Since the corpora are not annotated for this class of multi-words, co-compounds are not easy to retrieve automatically. We are adopting a combination of both corpus-driven and corpus-based approaches as the most effective way to analyse occasional items with hardly any automatically identifiable (morpho)phonological templates, such as synonymic and collective co-compounds. Given the heterogeneous nature of our dataset, we could not conduct a quantitative analysis.

One of the salient features of Russian rhyming and imitative co-compounds at the sentence level is their tendency to cluster together. We claim that this property is intrinsically related to their categorizing function. A characteristic example of co-compounds occurring in series is shown below. In (12), two asyndetically connected rhyming co-compounds, each of which is based on same-root deverbal

---

9.   The Russian National Corpus can be accessed at <http://www.ruscorpora.ru/new/>.

nouns with different prefixes, occur in adjacent position, giving rise to a higher-level category that can be labeled as [STRAIN OF TRAVELLING]:

(12) *Nenavižu   vse   èti   **s.bory-raz.bory,***
      hate.IPFV.1SG all.ACC.PL this.ACC.PL packing.ACC.PL.M-unpacking.ACC.PL.M
      ot.lëty-pri.lëty.
      departure.ACC.PL.M-arrival.ACC.PL.M
      'I hate all this **packing-unpacking, departures-arrivals**.'
                                    [Web, <https://neferjournal.livejournal.com>]

In (13), the co-compound *pívo-vódka* (*lit.* 'beer-vodka') encodes the *ad hoc* category [CHEAP ALCOHOL TO GET DRUNK ON]. The selection process, involved in the construction of the *ad hoc* category, can be traced throughout the chunk of discourse:

(13) - […] *Mužčina ne v sostojanii adekvatno sebja vesti daže pri postroenii otnošenij*
      s odnoj ženščinoj, kakaja tut otvetsvennost' za rebenka budet?
      - Nikakoj - budet   s   devkami   v igrovych
      none.GEN.SG.F be.FUT.3SG with chick.INS.PL.F in slot.PREP.PL
      *avtomatach   sidet'   **pivo-vodku***
      machine.PREP.PL.M sit.IPFV.INF beer.ACC.SG.N-vodka.ACC.SG.F
      pit'   […]
      drink.IPFV.INF
      '[…] [That] man is unable to behave properly even in a relationship with just one woman, how could he take responsibility for the child?'
      'No way! He will spend his time with chicks at slot machines drinking **beer, vodka (and the like)** […]'
                              [RNC, *Ženščina + mužčina: Psichologija ljubvi* (forum), 2004]

The components of the co-compound *pívo-vódka* [beer.SG.N-vodka.SG.F] 'beer, vodka and the like' constitute a non-exhaustive list, whose items are picked out from the class of inexpensive and, so to speak, 'inelegant' alcoholic beverages. In common world knowledge, their consumption has the aim of getting drunk, hence, neither wine nor champagne may qualify as members of this category. The concept is thus created *ad hoc* by the speaker by selecting names of drinks with properties that are relevant for the intended meaning. The encoding and decoding processes are, hence, simultaneously guided by several factors:

– world knowledge regarding alcoholic drinks and the related gender stereotypes, shared by the speaker and the hearer (for instance, the cultural attitude towards men as more prone to binge drinking than women);
– context, i.e. the interlocutors' shared knowledge regarding a specific topic;

– co-text, i.e. textual environment, containing references to other supposedly 'deplorable' habits, such as spending time with females, labelled with the degoratory term *devki* 'trollops, strumpets' (*búdet s dévkami…*), and playing slot machines (*v igrových avtomátach*), as well as an explicit statement regarding the man's inability to behave properly (*mužčína ne v sostojánii adekvátno sebjá vestí…*). The choice of the term *devki* to refer to females reflects the negative attitude of the speaker to the whole situation and, specifically, to the male referent.

All the aforementioned elements allow for a correct interpretation of the phrase *pit' pívo-vódku* [drink.IPFV.INF beer.ACC.SG.N-vodka.ACC.SG.F] 'drink beer, vodka, and the like', used informally to refer to excessive alcohol consumption. In combination with other relevant information in the co-text (spending time with girls, playing slot machines), it contributes to identify the *ad hoc* category [SHOWING IRRESPONSIBLE BEHAVIOUR TOWARDS CHILDREN]. The concept is thus overcoded by the speaker to guide the receiver's selective processing of the available information.

Co-compounds that are characterized by the categorizing function and consist of two different meaningful nominal bases, as in (14)–(15), are, supposedly, a distinguishing feature of present-day Russian, as compared to the rest of Slavic:[10]

(14) ***Ruki-nogi*** *cely* *– uže* *chorošo.*
arm/hand.NOM.PL.F-leg/foot.NOM.PL.F undamaged.PL already good
'The limbs and other body parts are undamaged, which is already good (in itself).' [RNC. Polina Vološina, Evgenij Kulkov. Marusja (2009)]

(15) *Ničego vam ne budet! Taščite iz sadov*
nothing 2PL.DAT NEG be.FUT.3SG steal.IPFV.IMP.2PL from garden.GEN.PL.M
čaški-ložki*, vorujte*
cup.ACC.PL.F-spoon.ACC.PL.F steal.IPFV.IMP.2PL
luk-morkov'*…*
onion(s).ACC.SG.M-carrot(s).ACC.SG.F
'You can just get away with it! You can steal tableware (from country houses), steal vegetables from gardens…'
                    [RNC. "Timurovcy" v zakone // Sel'skaja nov', 2003.09.16]

---

10. Co-compounding, as mentioned above, was a common morphological pattern in Russian folk literature. Some, for instance, Tkačenko (1979), have attributed its development to contact of northern Russian dialects with Finno-Ugric. The scholar observes that Finno-Ugric and East Slavic languages used a similar verbal co-compound as an opening formula in fairy tales (in Russian *žili-býli* [live.PST.PL-be.PST.PL] 'once upon a time…'). This form used to be spread over the whole East Slavic area and is still known in contemporary Russian. However, its relative frequency is much higher in Russian than in Belarusian and Ukrainian. Hence, probably, it would be reasonable to hypothesize areal influence from Russian towards the rest of East Slavic, outside of which it does not seem to have counterparts.

It can be deduced from parallel corpora data, as in (16–17), that some of the most lexicalized Russian co-compounds of the type under discussion have no correspondence in Polish and Bulgarian. In these languages, syntactic coordination is found in the contexts where Russian resorts to morphological co-compounding. Compare:

(16) Russian
  *Tol'ko by **ruki-nogi**   byli*
  only/just COND arm/hand.NOM.PL.F-leg/foot.NOM.PL.F be.PST.PL
  cely…
  undamaged.PL
  Bulgarian
  *Samo **răcete** i krakata mu da sa   zdravi…*
  only   arm.PL.F.DEF and leg.PL.M.DEF 3SG.DAT.M SBJV be.PRS.3PL healthy.PL
  'If only his arms, legs and other body parts were undamaged…'
              [RNC > parallel corpus, Č. Ajtmatov. *Placha*, 1987]

(17) Polish
  *pókim żyw i **rękoma**   a kolanami ruchać*
  until   alive and hand/arm.INS.PL.F and knee.INS.PL.N move.IPFV.INF
  mogąc
  be.able.CVB
  Russian
  *dokole živ i **rukami-nogami**   dvigat'*
  until   alive and arm/hand.INS.PL.F-leg/foot.INS.PL.F move.IPFV.INF
  mogu
  be.able.PRS.1SG
  '…until I am alive and able to move my body…'
            [RNC > parallel corpus, Henryk Sienkiewicz.
                  Pan Wołodyjowski, 1887–1888]

As far as imitative co-compounds (Belikov 1990; Wälchli 2005; Voinov 2012; Rozhanskiy 2015) are concerned, they consist of a content noun, followed by a semantically empty component. The latter is obtained by phonetic modification of the first syllable of the semantic head according to one of the three formal patterns, available in the present-day language. *M-* and *shm-*reduplications, such as *šúšera-múšera* 'riff-raff', *dokazátel'stva-šmokazátel'stva* 'evidence and all that stuff', and so forth, are among the most famous instances of constructional borrowing in the domain of word-formation (from Turkic and Yiddish, respectively).[11] From

---

11. Minlos (2004: 16) points out that the Russian folklore presents a variety of patterns of imitative reduplication with different phonetic modifications, for instance, *solóma-volóma* < solóma 'straw', *sáchar-báchar* < *sáchar* 'sugar', *solómina-jalómina* < *solómina* '(single) straw' etc. These patterns of reduplication are an internal East Slavic development.

the perspective of language contact, both patterns are notoriously widespread across large areas, including the languages of Central and Eastern Europe. Thus, for instance, *shm*-reduplication is rather common in Polish, whereas in Balkan Slavic (specifically, in Bulgarian and Macedonian) *m*-reduplication is particularly diffused. The Russian inventory of imitative co-compounds is further enriched with reduplications based on taboo words, which follow a similar morphonological pattern and represent a salient informal discourse strategy.

The remaining part of the paper is organized as follows. Sections 2 and 3 are focused, respectively, on rhyming and imitative co-compounds, whereas Section 4 contains conclusive remarks.

2.   Rhyming co-compounds

Rhyming co-compounds consist of two rhyming content nouns and convey a generalizing/categorizing meaning. The rhyming effect can be obtained by a mere repetition of the final part of the base, containing derivational and inflectional suffixes, as well as of the stress pattern. Consequently, the two bases often share the same derivational and inflectional morphemes, resulting in the same word-formational pattern. For instance, there is a common pattern, involving the suffix *-lka* (preceded by a stem vowel). It is predominantly exploited in the formation of deverbal instrument nouns, denoting tools (e.g., *kop-í-lka* 'piggy bank' < *kop-í-t'* 'save [money]'; *meš-á-lka* 'agitator, mixer' < *meš-á-t'* 'mix'). Since the suffix requires the insertion of the thematic vowel of the infinitive/past stem of the verb base, the rhyming noun bases tend to be derived from verbs sharing the same thematic vowel, although, as is shown below, it is not a rigid rule. Consider the co-compound *mazílki-krasílki* [ointment.PL.F]-[greasepaint.PL.F] 'makeup-shmakeup, makeup and the like' in (18):

(18) *ja*   *pol'zovalas'*   *dešëvoj*   *kosmetikoj*   *do 20*
      1SG.NOM use.IPFV.PST.SG.F.REFL cheap.INS.SG.F cosmetics.INS.SG.F till 20
      let,   *poka*   […] *ne   stala*   *sama*
      year.GEN.PL till/before   NEG begin.PFV.PST.SG.F oneself.NOM.SG.F
      zarabatyvat', čtoby   pozvolit'   sebe   firmennye
      earn.IPFV.INF so.that allow.PFV.INF REFL.DAT trendy.ACC.PL
      mazilki-krasilki.
      ointment.ACC.PL.F-greasepaint.ACC.PL.F
      'I used cheap cosmetics when I was under 20, before I started earning my own money so that I could afford trendy **makeup-shmakeup**.'
                          [RNC, *Krasota, zdorov'e, otdych: Krasota* (forum), 2005]

There is a tight semantic connection between the two components of the compound in (18), referring to the higher-level category [SUBSTANCES ONE'S FACE CAN BE SMEARED WITH]. From a perceptual perspective, reference to the more specific non-lexicalized category [SKINCARE AND MAKE-UP PRODUCTS] is inferred from world knowledge. This interpretation is also guided by the co-text, as in the left context an explicit reference to the use of cheap cosmetics by the speaker is present. The nominal bases of *mazílki-krasílki* are derived, respectively, from the verbs *mázat'* 'grease, smear' and *krásit'* 'paint' by means of the suffix *-(i)lka*. Theoretically speaking, both *mázalka* and *mazílka* (< *máz-*) exist and can be used with the meaning 'ointment, a substance that can be spread on a surface', as is illustrated in (19)–(20). However, the variant with the vowel *i* was selected in (18), apparently, to fit the rhyming pattern:

(19) *vot,   deljus'   vkusnym   i èlementarnym*
here.it.is share.IPFV.PRS.1SG tasty.INS.SG.M and elementary.INS.SG.M
receptom   "mazalki"   na buterbrody.
recipe.INS.SG.M spreadable.stuff.GEN.SG.F on sandwich.ACC.PL.M
'Hey there, I am going to share (with you) a tasty and simple recipe of **spreadable stuff** for sandwiches.'   [Web, http://bulka.livejournal.com/103698.html]

(20) *U staršej   vetrjanka.   <…> Govorjat   tam   u*
at elder.GEN.SG.F chickenpox.NOM.SG.F   say.IPFV.PRS.3PL there at
vas   est'   čudo-mazalka,   kotoraja
2PL.GEN be.PRS miracle.NOM.SG.N-ointment.NOM.SG.F which.NOM.SG.F
snimaet   zud.
relieve.IPFV.PRS.3SG itching.ACC.SG.M
'My oldest daughter has chickenpox. <…> They say you have a miraculous **ointment**, relieving itching.'
[Web, <https://eva.ru/travel/messages-2159785.htm>]

Another example of *(V)lka*-suffixed rhyming co-compounds is shown below. The conjuncts of the co-compound *sop-é-lka-pycht-é-lka* [wheeze-V-NMLZ.SG.F-puff-V-NMLZ.SG.F] 'wheezing (and) puffing, nonsensical, primitive text' are, respectively, derived from the almost synonymous verbs *sopét'* 'wheeze' and *pychtét'* 'puff', sharing the thematic vowel *-e-* in the infinitive/past stem. *Sopélka* and *pychtélka* are the names of certain types of hums composed by Winnie-the-Pooh, as they were rendered by the Russian translator Boris Zachoder. The co-compound was born as a creative formation with a specific referent. However, subsequently, as it became part of common cultural knowledge (also due to a popular cartoon), it underwent full lexicalization. Its lexical meaning broadened, and it developed the categorizing properties, typical of structurally similar co-compounds. In (21) below, where the co-compound refers to the *ad hoc* category [UNSOPHISTICATED

LYRICS], it is preceded by the indefinite adjective *kakoj-nibud'* 'any, some; no more than', performing an additional operation of downgrading:

(21) *Kak-to otec   Andrej Kuraev   privël*
      once   father.NOM.SG.M <male.name>.NOM.M provide.PFV.PST.SG.M
      udačnyj   primer   iz   "Vinni-Pucha".
      felicitous.ACC.SG.M example.ACC.SG.M from Winnie-the-Pooh.GEN
      Nel'zja   že   vzjat'   i   zaprosto napisat'
      it.is.impossible PTCL take.PFV.INF and simply   write.PFV.INF
      kakuju-nibud' sopelku-pychtelku.   *Poèzija*
      some.ACC.SG.F wheezing.ACC.SG.F-puffing.ACC.SG.F poetry.NOM.SG.F
      sama   nachodit   tebja.   Vot   tak že   i   ljubov'.
      oneself.NOM.SG.F find.IPFV.3SG 2SG.ACC here.is so   PTCL and love.NOM.SG.F
      'Once Father Andrej Kuraev used one good example from "Winnie-the-Pooh".
      One cannot just compose **some hum** out of nowhere. Poetry finds its way to
      you by itself. Just as love.'   [Web, <http://sp-g.ru/gazeta/1/1n.htm>]

Now, let us have a look at partial rhymes in co-compounds. In (22), the two rhyming bases *vél-ik-i-ról-ik-i* [bike-DIM-PL.M]-[roller.blade-DIM-PL.M], both of which are also used as independent lexemes in Russian, share the same number of syllables, stress pattern, as well as a rhyme. The rhyme is partial, as the vowels of the stressed syllables do not overlap. It is worth noting that the superficial parallelism in the morphemic structure of these nouns is a result of different underlying word-formation operations. The noun *vélik* 'bike' is a truncated form of the noun *velosipéd* 'bicycle', with the addition of the diminutive suffix *-ik-*. By contrast, *róliki* 'rollers, rolls; roller blades' is a mere truncation of the Adj+N phrasal lexeme *rólik-ovye kon'kí* 'roller blades', containing the adjective *ról-ik-ov-ye* [roll-DIM-ADJZ-PL], in which the suffix *-ik-* is already present.[12] The syllabic rhyme between the bases of the co-compound is thus obtained due to the overlap of the final unstressed chunks, containing identical derivational and inflectional suffixes:

(22) *Vmesto p'janych   kompanij   sportivnaja   molodëž'*
      instead drunken.GEN.PL group.F.GEN.PL athletic.NOM.SG.F youth.NOM.SG.F
      na velikach-rolikach.
      on bike.PREP.PL.M-roller.blade.PREP.PL.M
      'Instead of groups of drunken guys, athletic youths with **bikes, roller blades
      etc**.'   [RNC, A. Starobinec. *Padenie čudo-grada* //
                                              Russkij reporter, 45 (223), 2011]

---

12. The noun *rólik* 'small wheel, roller, roll' exists in Russian, however, in the case of the plural form *róliki* 'roller skates' it is reasonable to consider it a truncation of the respective full phrasal lexeme, *rólikovye kon'kí*.

The morphological rhyme between the bases iconically suggests a semantic connection between the two concepts and guides the exemplar-driven categorization process. In the context shown in (22) above, bikes and rollers represent the *ad hoc* category [amateur wheel sports as leisure activities for young people]. These items are selected as prototypical members of the category. The latter potentially includes other items with the same functional properties, such as skateboards or scooters.

One more example of synonymous rhyming bases with diachronically different morphological structures is given in (23) below. In this example, the co-compound *bardák-kavardák* [brothel/mess.sg.m-mess.sg.m] is based on two nouns in -*ak*. While *bardák*, in fact, is parsed as containing the colloquial multifunctional suffix -*ak* (< *bordel* 'brothel'), *kavardák* is a morphologically unsegmentable Turkic borrowing. The nouns are synonyms and have the meaning 'mess, disorder'. When they occur as bases of the co-compound, they yield an intensifying and categorizing effect, resulting in the *ad hoc* category [extremely messy situation]. Intensification is thus iconically encoded by the co-occurrence of two phonetically similar synonymic nouns:

(23) *Zabavno, čto otnošenija   načalis'   so   2 na 3*
       it.is.funny that relationship.nom.pl.n start.pfv.pst.pl.refl from 2 on 3
       janvarja   na anime-pati,   potom byl
       January.gen.sg.m at anime-party[prep.sg] then be.pst.sg.m
       bardak-kavardak..
       brothel/mess.nom.sg.m-mess.nom.sg.m
       'It is funny that the relationship started between the 2nd and 3rd of January during an anime-party, and then there was a **terrible mess**…'

                                    [Web, https://svetazara.livejournal.com]

One might wonder why speakers should use occasional co-compounds, when the concepts involved are already lexicalized. A possible answer is that exemplar-driven categorization and existing category labels (e.g., *mazílki-krasílki* [ointment.pl.f-greasepaint.pl.f] vs. *kosmetika* 'cosmetic products') are not fully equivalent, either semantically or pragmatically. *Mazílki-krasílki* is not a mere sum of skincare products and make-up. An evaluative component, involving irony and intimacy with the topic of the message, as well as with the interlocutor, is added. Moreover, the reduplicative structure of the co-compound, along with plural marking, iconically suggests an indefinitely large quantity of matter. In turn, the co-compound *véliki-róliki* [bike.pl.m-roller.blade.pl.m] does not simply encode the meaning 'wheel sports'. It more generally refers to outdoor leisure activities for young people. To sum it up, the co-compound components are exemplars of a target category. However, in order to correctly process the category, the hearer must select only

those properties of the compound constituents that are relevant for the speaker. In doing this, the hearer relies on contextual and background knowledge.

Further discussion deals with two other subtypes of rhyming co-compounds, the appositive and the prefixed subtype.

2.1   The appositive subtype

In the literature on compounds in Russian, 'appositional'/'appositive' compounds appear to represent a heterogeneous group. For instance, Ward (1973) puts into this category *górod-prízrak* [town.SG.M-ghost.SG.M] 'ghost town', *žénščina-vrač* [woman.SG.F-doctor.SG.M] 'female doctor', as well as *vagón-restorán* [carriage.SG.M-restaurant.SG.M] 'dining car'. Wälchli (2005: 161–162) follows the same approach. However, it is obvious that the semantic structure and, respectively, the underlying syntactic relations differ within these formations. According to the classification by Bisetto and Scalise (2005) and Scalise and Bisetto (2011), compounds, such as *žénščina-vrač* [woman.SG.F-doctor.SG.M] 'female doctor', *vagón-restorán* [carriage.SG.M-restaurant.SG.M] 'dining car', and *ávtor-ispolnítel'* [author.SG.M-performer.SG.M] 'singer/songwriter, singer/performer' must be classified as belonging to the coordinate additive type. Supposedly, there is an implicit 'and'-relation between their constituents, in that a female doctor is, at the same time, a female and a doctor, a singer/performer is both a singer and a performer, and so forth. It has been observed by Arcodia et al. (2009) that such compounds predominantly pertain to two semantic groups, respectively, denoting professions and instruments.

Then, compounds of the type of *górod-prízrak* [town.SG.M-ghost.SG.M] 'ghost town' and *čelovék-zagádka* [man.SG.M-riddle/mystery.SG.F] 'man of mystery' should be classified as appositive-attributive (ATAP). The semantic non-head in ATAP compounds is an attribute of the head and is often used metaphorically. In fact, a ghost town is not a town and a ghost at the same time. The noun *prízrak* 'ghost' only projects a property onto the noun *górod* 'town', which is the semantic head of the compound. Thus, in this paper, we only understand by appositive co-compounds a subtype of rhyming co-compounds, in which semantic/pragmatic features of the second constituent are projected onto the first constituent.

The appositive subtype of co-compounds can be distinguished on syntactic, as well as semantic and pragmatic, grounds. Both bases of an appositive co-compound are content nouns. However, only the first noun functions as the semantic head, whereas the second rhyming noun adds a connotative value to the denotative meaning, expressed by the semantic head. Consider, for instance, (24) below. In this example, the first base of the co-compound *družóček-pirožóček* [friend.DIM.SG.M-pie.DIM.SG.M] 'sweetie pie' (lit. 'friend-pie'), referring to a kitten, is the semantic head,

whereas the second base is an appositional attribute, expressing affection and endearment. Moreover, this pragmatic meaning is overcoded, inasmuch as it is also expressed by the diminutive suffixes, applied to both bases, as well as by the lexical meaning itself of the first nominal base (*družoček* 'sweet little friend' < *drug* 'friend'):

(24) *Sročno! Otdam*   *ètogo*
       urgent give.away.PFV.FUT.1SG this.ACC.SG.M
       družoček-a-pirožoček-a   *v dobrye*   *ruki,*
       friend.DIM-ACC.SG.M-pie.DIM-ACC.SG.M in good.ACC.PL hand.ACC.PL.F
       očen' neposedlivyj,   *laskovyj*   *kotënok,*
       very energetic.NOM.SG.M affectionate.NOM.SG.M kitten.NOM.SG.M
       kušaet   *VSË!*   *Kotënok*   *ot*
       eat.IPFV.PRS.3SG all.ACC.SG.N kitten.NOM.SG.M from
       koški-krysolovki!
       cat.GEN.SG.F-rat.catcher.GEN.SG.F
       'Urgent! I am looking for a good home for this **sweetie pie**. (It is) a very energetic and affectionate kitten, (he) eats EVERYTHING! A rat catcher's son!'
                [Web, <https://vk.com/wall-57249769?q=%23pets&offset=20>]

The lexicalized co-compound *ljubóv'-morkóv'* [love.SG.F-carrot.SG.F] 'affection-confection; together-forever' ironically refers to romantic feelings. The ironic effect is generated by the semantic clash between an abstract and a concrete noun, evoking different frames: poetic fine feelings (*ljubóv'* 'love') vs. non-poetic peasant life (*morkóv'* 'carrot'). The two nouns are morphonologically similar. They belong to the same non-productive, highly restricted and hence marked nominal declension paradigm, including feminine nouns with the rare stem marker *-ov'*. However, they pertain to different lexical semantic domains and, consequently, produce a comic effect when used together. It can be hypothesized that such semantically unrelated but phonologically similar words, pertaining to a single closed morphological class, are easily extracted together from the mental lexicon. Furthermore, from a cultural and pragmatic perspective, the co-compound is an ironical imitation of the highly predictable rhyme *krov'*/*ljubóv'* ('blood'/'love'). The noun *ljubóv'* 'love' is widely assumed to be difficult to rhyme due to its phonological pattern. This is an illustration of how the decoding process can be rooted in the background knowledge, shared by the speakers and hearers.

In (25) below, the plural-marked co-compound *ljubóvi-morkóvi* [love.PL.F-carrot.PL.F] 'together-forever (and the like)', placed at the end of a list,[13] performs a reformulating function:

---

13. For a more detailed account of the categorizing function of lists in Russian, see Benigni 2016 and 2018.

(25) *V Rossii   […] rimejki   i   kon"junkturnye*
      in Russia.PREP.SG.F   remake.NOM.PL.M and time.serving.NOM.PL
      "festival'nye"   podelki   bratkov   Michalkovych,
      festival.NOM.PL makeshift.piece.NOM.PL.F bro.GEN.PL.M Mikhalkov.GEN.PL
      telemuviki,   fil'my   vychodnogo   dnja…
      telemovie.DIM.NOM.PL.M film.NOM.PL weekend.GEN.SG.M day.GEN.SG.M
      ofisnye   strasti…   vse   èti
      office.NOM.PL drama.NOM.PL.F all.NOM.PL this.NOM.PL
      ljubovi-morkovi…
      love.NOM.PL.F-carrot.NOM.PL.F
      'In Russia […] remakes and time-serving "festival" makeshift pieces by the
      Mikhalkov bros, telemovies, weekend films … office dramas … all these
      together-forever…'
                      [RNC, *Obsuždenie fil'ma "Korol' govorit!"* (forum), 2010–2011]

With this co-compound, the speaker ironically targets the *ad hoc* category [FILM OR
PIECE OF LITERATURE WITH MELODRAMATIC SUBJECT MATTER]. The first compo-
nent thereof (*ljubóvi* [love.PL.F] 'loves') is used metonymically (as the content of the
object referred to). As for the second component (*morkóvi* [carrot.PL.F] 'carrots'),
it was originally based on the defeated expectancy effect, but now it is perceived
as semantically bleached and merely ironical. Furthermore, the use of the plural
number, along with the indefinite and demonstrative adjectives *vse èti* 'all these',
reinforce the reformulating nature of the co-compound and allude to the consist-
ency of the *ad hoc* category.

Another appositive co-compound, whose level of lexicalization is rather ad-
vanced, is *lečénie-mučénie* (lit. 'treatment-torment'). Its use is illustrated in (26).
Its components are linked by an underlying cause-effect relationship, revealing
negative attitudes towards doctors and medicine. The side effects from medical
treatment are thought to be worse than the disease itself. At the syntactic level, the
compound appears to be derived from an equative copular clause (*treatment is tor-
ment*). In fact, on the web, *lečénie-mučénie* is sometimes typed with the equal sign
(=), substituting the hyphen. Thus, the first component is the semantic head of the
co-compound, whereas the second one adds an evaluative (in this case, negative)
meaning thereto. Consider the example below, where Russian-speaking immigrants
discuss the public healthcare system in their country of residence. The semantic
head of the co-compound is modified by the adjective *besplátnoe* 'free of charge',
whereas the second nominal base occurs as a purely attributive element, projecting
a property on the head:

(26) *Koroče besplatnoe*   **lečenie-mučenie,**
in.sum free.of.charge.NOM.SG.N treatment.NOM.SG.N-torment.NOM.SG.N
chotja esli učest' skol'ko nalogov s tebja sderut za takoj udovol'stvie stanovitsja
obidno.   [Web]
'In sum, (it is) free **treatment-torment**, although, if you consider how many
taxes you have to pay for this pleasure, it makes you frustrated.'

[Web, <https://www.kharkovforum.com>]

The collective interpretation of the same co-compound is also possible. In (27), it
refers to the non-lexicalized, higher-level concept [SUFFERING DUE TO DISEASE AND
EFFORTS TO FIND A CURE]:

(27) *Žizn'*   *rebenka*   *prevraščaetsja*   *v*
life.NOM.SG. child.GEN.SG.M transform.IPFV.PRS.3SG.REFL in
splošnoe   **lečenie-mučenie,**
continuous/total.ACC.SG.N treatment.ACC.SG.N-torment.ACC.SG.N
nos   *ne*   *dyšit,*   *po nočam*   *rebenok*
nose.NOM.SG.M NEG breathe.IPFV.PRS.3SG at night.DAT.PL.F child.NOM.SG.M
chrapit   *ili zadychaetsja*
snore.IPFV.PRS.3SG or choke.IPFV.PRS.3SG.REFL
'The child's life is transformed into a continuous [medical] **treatment-torment**,
s/he cannot breathe out of her/his nose, the child snores or chokes at night.'

[Web, <http://otzovik.com/review_211151.html>]

In the example below, the co-compound *vrač-portáč* [doctor.SG.M-good.at.nothing.
SG.M] 'quack, incompetent doctor' consists of an agent noun, denoting a profession,
and a rhyming noun with the meaning 'a person that screws the things up':[14]

(28) **Vrač-portač**   *chočet*
doctor.NOM.SG.M-good.at.nothing.NOM.SG.M want.IPFV.PRS.3SG
priznat'   *obvinjaemogo*   *psichom*   *i*   *zakatat'*
recognize.PFV.INF defendant.ACC.SG.M psycho.INS.SG.M and lock.PFV.INF
v durku.
in nut.house.ACC.SG.F
'The **incompetent doctor** wants to declare the defendant a psycho and lock
him in a nut house.'   [Web, <http://www.yaplakal.com>]

---

14. Both nouns share the final segment -*ač*. *Vrač* 'doctor' etymologically contains the agent suffix
-*ač* (however, the speakers are no more aware of it and perceive the noun as a simplex word). As
for *portáč*, due to reanalysis it is synchronically perceived as an agent formation with the suffix
-*ač* from the verb *pórtit'* 'ruin, screw up', although the actual etymology of the word is different
(most probably, from Polish *partacz* < Latin *a parte paternitatis* 'a craftsman that is not a member
of a corporation').

2.2   The prefixed subtype

In the previous section it was shown how co-compounds with a morphological rhyme, given by derivational and inflectional suffixes, behave in discourse. Now, we turn to another subtype of rhyming co-compounds, in which the rhyming element is (a) the root morpheme, or (b) the root morpheme plus a suffix, whereas the initial segments, represented by prefixes, are different.

Russian verbal prefixes, most of which still retain their original spatial meanings, notoriously present a case of grammaticalization from spatial paths, goals and sources into telicity markers (see the overview in Josephson 2015). Deverbal nouns inherit the prefixes from the respective verbs. As is shown in (29), they are often created by back formation, so that the final segment of the noun is the root morpheme without any word-formation suffixes on the right:

(29) *pri-chod-í-t'* [PRX:at/by-go-V-INF] 'come' → *pri-chód* [PRX:at/by-go.NMLZ] 'arrival'
pod-chod-í-t' [PRX:under-go-V-INF] 'come up to, approach' → *pod-chód* [PRX-:under-go.NMLZ] 'approach'

The alternation of spatial and non-spatial meanings of some of the Russian verbal prefixes, frequently occurring in the co-compounds under discussion, is briefly illustrated below:

(30) *pod-* 'under'
bežat'[IPFV] 'run' → *pod-bežat'*[PFV] 'run up close to (the target)' [spatial]
est'[IPFV] 'eat' → *pod-"est'*[PFV] 'eat up every bit of something' [non-spatial]
do- 'up to, till'
bežat'[IPFV] 'run' → *do-bežat'*[PFV] 'run up to, reach (the target)' [spatial]
pet'[IPFV] 'run' → *do-pet'*[PFV] 'finish singing/sing to the end' [non-spatial]
raz(o)-/ras- 'apart'
rvat'[IPFV] 'tear' → *razo-rvat'*[PFV] 'tear apart' [spatial]
ljubit'[IPFV] 'love' → *raz-ljubit'*[PFV] 'cease to love' [non-spatial]

As was argued by Janda et al. (2013: 179), verbal prefixes function as verbal classifiers in Russian, inasmuch as they "sort the verbs according to the typical type of event expressed, similar to the way that numeral classifiers sort nouns according to the typical type of object referred to". The classifying function of the prefixes is also exploited in deverbal nominal co-compounds, where a highly frequent verb stem with a generic meaning is combined with different prefixes, which restrict and/or modify the meaning of the verb stem. The different prefixes, applied to the same verb stem, contribute to the creation of *ad hoc* categories, associated with complex situations and/or concepts. Prefixed same-root co-compounds typically have generalizing and categorizing functions, as in (31):

(31) *vse   analizirujut*
all.PL analyze.IPFV.PRS.3PL
rasklady-doklady   *načaľnika*
PRX:apart.putting.ACC.PL.M-PRX:up.to.putting.ACC.PL.M boss.GEN.SG.M
našego   GUUR Trubnikova   za prošlyj   god.
our.GEN.SG.M GUUR Trubnikov.GEN.SG.M for last.ACC.SG.M year.ACC.SG.M
'All [of them] are still analyzing the last year **reports and all that stuff** by
Trubnikov, the boss of our GUUR.'   [RNC, V. Čerkasov. *Černyj jaščik*, 2000]

Both components of the co-compound *ras.klády-do.klády* [PRX:apart.putting.
PL.M-PRX:up.to.putting.PL.M] 'reports and the like' (lit. 'accounts-reports') are
derived from prefixed verbs, sharing the root morpheme *-klad-* 'put, place hori-
zontally'. The noun *rasklád* 'situation; state of affairs; account of a state of affairs',
containing the distributive verbal prefix *raz-/ras-* (< *ras.kládyvat'* 'put in separate
locations, lay out; explain, relate'), has several related meanings, but only the rel-
evant one (i.e., 'account of the state of affairs') is picked out, as it is combined
with the noun *doklád* 'report' (< *do.kládyvat'* 'report, announce'). The respective
co-compound encodes the higher-level category [REPORTS, EXPLANATIONS AND
THE LIKE].

In the example below, the co-compound *zakósy-perekósy* [behind.oblique.
PL.M-over.oblique.PL.M] (lit. 'deviations-distortions'), refers to [DEVIANT SEXUAL
BEHAVIOR]. Moreover, it is inserted into a rhyming list pattern. Although the
basic colloquial meaning of the noun *zakos* is 'imitation, impersonation', it re-
ceives the contextual interpretation 'deviation (from the normality)'. Both compo-
nents rhyme with the preceding independent plural noun *zasósy* 'love bites'. This
three-component rhyming listing structure, including both a free noun form and
a co-compound, suggests conceptual unity:

(32) *I   psichoterapevt,   člen   mnogich*
and phychotherapist.NOM.SG.M member.NOM.SG.M many.GEN.PL
Akademij,   daët   razbor
academy.GEN.PL.F give.IPFV.PRS.3SG critical.analysis.ACC.SG.M
polëtov   – vsjakie   zasosy,
flight.GEN.PL.M various.NOM.PL love.bite. NOM.PL.M
zakosy-perekosy   *v sekse.*
deviation.NOM.PL.M-distortion.NOM.PL.M in sex.PREP.SG.M
'And a psychotherapist, member of many Academies, provides a detailed critical
analysis of the matter – various **love bites, deviations and distortions** in sex.'
                        [Web, <https://meggirita.livejournal.com/647348.html>]

In the example below, two near-synonymous cognate nouns with the meaning
'rotation', forming a co-compound, give rise to the *ad hoc* category [ROTATIONS
AND THE LIKE]:

(33) *Vsjakie*   ***povoroty-razvoroty***   *tablic*
      various.ACC.PL rotation.ACC.PL.M-rotation.ACC.PL.M table.GEN.PL.F
      lučše   *delat'*   *v Excel.*
      good.COMP do/make.IPFV.INF in Excel[PREP.SG]
      'It is better to perform various **rotations** of tables **(and other similar operations)**
      in Excel.'   [Web, <http://wordexpert.ru/forum/viewtopic.php?id=3718>]

Prefixed deverbal nominal co-compounds are often based on reversive pairs. The
semantic properties of these reversive co-compounds are determined by the double
semantic nature of Russian verbal prefixes, which, on the one hand, retain their
basic spatial meanings, and, at the same time, encode actional semantics, associated
with telicity and result states. If a spatial schema is adopted, co-compounds can
potentially describe complex events, involving motion towards a goal and return-
ing along the same path to the point of departure. However, in most cases, they
rather refer to an *ad hoc* category, including both directions of motion, which does
not imply that these events are part of a single complex situation. This is the case
in the sample sentences below, where the co-compounds *posádka-výsadka* [entry.
SG.F-alighting.SG.F] 'drop-off and pick-up process' (< *posadít'* 'sit (causative)', *výsa-
dit'* 'drop off') and *v"ézdy-výezdy* [in.going.PL.M-out.going.PL.M] 'relocations' have
a purely categorizing function:

(34) ***Posadka-vysadka***   *čut' lučše,*   *čem u*
      entry.NOM.SG.F-alighting.NOM.SG.F a.bit good.COMP than at/by
      nekotorych   *odnoklassnikov.*
      some.GEN.PL same.category.car.GEN.PL.M
      '**Entry and alighting** are somewhat better than in some cars of the same cate-
      gory.'   [RNC, A. Budkin. *Plašč toreadora* // Za rulëm, 2003]

(35) *Na svoëm*   *ètaže*   *Katja*   *nikakich*
      on REFLPOSS.M.PREP.SG floor.PREP.SG.M Katja.NOM.SG.F not.any.GEN.PL
      v"ezdov-vyezdov   *v poslednee*   *vremja*   *ne*
      entry.GEN.PL.M-exit.GEN.PL.M in recent.ACC.SG.N time.ACC.SG.N NEG
      nabljudala,   *značit, madam*   *s*   *četvërtogo.*
      notice.IPFV.PST.SG.F so   madam.NOM.SG.F from fourth.GEN.SG.M
      'Katja didn't notice any **movements in and out** lately, so it means that the lady
      is from the fourth floor.'
                            [RNC, M. Zosimkina. *Ty prosněš'sja. Kniga pervaja*, 2015]

In the example below, the co-compound *vchódy-východy* [in.going.PL.M-out.going.
PL.M] with deverbal nominal bases (< *vchodít'* 'enter', *vychodít'* 'exit'; they are not
action nouns, however, they have to do with directionality) refers to the *ad hoc*
category [ENTRANCES AND EXITS OF ANY TYPES]:

(36) ***Vchody-vychody***   *po-prežnemu byli*
entrance.NOM.PL.M-exit.NOM.PL.M still   be.PST.PL
otkryty   nastež',   na zimu   ne   bylo
open.PFV.PST.PASS.PTCP.PL wide.open on winter.ACC.SG.F NEG be.PST.SG.N
sdelano   daže krovli.
make.PFV.PST.PASS.PTCP.SG.N even roofing.GEN.SG.F
'**Entrances and exits** were still wide open, even the roofing wasn't done.'
[RNC. K. Mikhajlov. *Styd i požar* // Ogonëk, 2014]

As can be evinced from the above examples, when spatial paths are involved, the ordering of the prefixed bases of the co-compound is driven by iconicity. The natural sequence of events here involves the attainment and subsequent cancellation of the result state of a telic event, encoded by the deverbal noun.

Obviously, reversive co-compounds are not always based on spatial paths. Non-spatial reversives often include a second constituent with the prefix *raz-/ras-*, encoding the spatial meaning of separation ('apart'), as well as undoing/reversal of the result state of an event, which is a secondary non-spatial meaning, directly derived from the primary spatial meaning (Janda & Nesset 2010). For instance, consider (37). In this example, the bases of the co-compound *dochódy-raschódy* [income.PL.M-expenditure.PL.M], sharing the same verbal root (*-chod-* 'go'), have the prefixes *do-* 'up to' and *raz-/ras-*, respectively. This co-compound is a lexicalization of the higher-level category [INCOMES AND EXPENDITURES], involving two opposite processes:

(37) *Zastavit   medikov   po-novomu   sčitat'*
force.PFV.FUT.3SG doctor.ACC.PL.M in.a.new.way count.IPFV.INF
*svoi   **dochody-raschody**   i*
REFLPOSS.ACC.PL income.ACC.PL.M-expenditure.ACC.PL.M and/also
smena   ich   pravovogo   statusa.
change.NOM.SG.F POSS3PL legal.GEN.SG.M status.GEN.SG.M
'Another thing that will force doctors to count their **revenues and expenditures** in a new way is the change of their legal status.'
[RNC. E. Kostjuk. *Nas ždët reanimacija* // Vremja MN, 2003]

Furthermore, co-compounds with nominal bases, derived from verbs of motion with near-synonymous prefixes, can encode the approximation of a spatial path. Russian verbal prefixes often encode highly specific spatial trajectories. For instance, there is no generic prefix of directed motion in Russian. However, the prefix *za-* ('behind, beyond') may indicate motion with a final point located inside the goal, whereas *pod-* ('under') encodes a motion path with a final point, located close to the

external boundary (outside) of the goal.[15] In the example below, the co-compound *zaézdy-pod''ezdy* [behind.driveway.PL.M-under.driveway.PL.M] (< *ézdit'* 'ride, drive') encodes the *ad hoc* category [DRIVEWAYS OF DIFFERENT TYPES AND THE LIKE]. Moreover, due to the combination of the two prefixes in question, it encodes [MO-TION TOWARDS A GOAL WITH AN UNDERSPECIFIED FINAL POINT], which, in turn, also constitutes an *ad hoc* category:

(38) *Otel'*   *dejstvitel'no polnost'ju specializirovan*   *na*
hotel.NOM.SG.M really   completely specialize.PASS.PTCP.PST.SG.M on
semejnyj   otdych.   Est'   i   krovatki   v
family.ACC.SG.M vacation.ACC.SG.M COP.PRS and infant.bed.NOM.PL.F in
nomere   i   stul'čiki   v stolovoj.
room.PREP.SG.M and high.chair.NOM.PL.M in dining.room.PREP.SG.F
*Pandusy,   **zaezdy-pod''ezdy**.*
ramp.NOM.PL.M driveway.NOM.PL.M-driveway.NOM.PL.M
'The hotel is really specifically designed for family vacations. There are baby beds in the rooms and highchairs in the dining room. Ramps, (various) **driveways and the like**.'   [Web, https://otpravkin.ru/strany?hotel= amelia-beach-resort-hotel-spa&showreviews=1]

In the same vein, the *ad hoc* category [ACCESS POINTS OF DIFFERENT TYPES AND THE LIKE] is constructed in (39) below. The co-compound *podchódy-zachódy* [under.way.PL.M-behind.way.PL.M] (< *chodít'* 'go, walk')] :

(39) *A   ved'   ja   tam   očen' často byvala   i*
but however 1SG.NOM there very   often be.ITER.IPFV.PST.SG.F and
*vse   **podchody-zachody**   znaju.*
all.ACC.PL access.point.ACC.SG.M-access.point.ACC.SG.M know.IPFV.PRS.1SG
'However, I used to come around that place very often and I know all the access points (and the like).'
[Web, <https://neglinka-msk.livejournal.com/145482.html>]

3.   Imitative co-compounds

As was mentioned in the introduction, Russian exhibits three main patterns of imitative co-compounds: (a) *m*-reduplication, (b) *shm*-reduplication, and (c) ta-boo root reduplication. Like other types of co-compounds, discussed in this paper, imitative co-compounds, also known as echo-reduplications, express 'blurred

---

15. Needless to say, all Russian verbal prefixes are polysemous. However, a discussion of all their spatial and non-spatial interpretations with verbs of motion falls beyond the scope of this paper.

similarity' (Rozhanskiy 2015: 1010), i.e. meanings of the type 'N and the like'. As imitative co-compounds may refer to a class of similar items, they are frequently exploited for widening the domain of meaning of a noun. In terms of Construction Grammar, they can be hence described as follows:

FORM: [N-IMI$_N$]
MEANING: <N and the like; N and so on>

N is a content noun, whereas IMI$_N$ is a non-referential form obtained by phonetic modifications on the first syllable of N. The pejorative connotation, often associated with this class of co-compounds, is iconically reflected in the imperfect imitation of the original item, provided by the second component of the co-compound.

M-reduplication is an expressive discursive strategy, characteristic of colloquial Turkish. Armoskaite and Kutlu (2014) define it as "a case of lexical number, namely similative plural". The categorizing function of this type of reduplication can be illustrated by an expressive example, cited by Wiese and Polat (2016: 247), namely, an excerpt from Recep Erdoğan's election speech. An *m*-reduplication of the noun *Twitter* is a means used to express the speaker's general attitude towards the category [TWITTER AND OTHER SOCIAL MEDIA]:

(40)   Turkish
        Twitter, mwitter hepsinin kökünü kazıyacağız.
        'We will eradicate Twitter and so on.'

According to Wiese and Polat (2016), *m*-reduplication can be also observed in urban Germany, in settings of German-Turkish language contact, where it acquires a pejorative, dismissive, and distancing function.

As far as Russian is concerned, *m*-reduplication is a well-known case of morphological borrowing, a pattern borrowing. The reduplication template itself, as a word-formational operation requiring a total reduplication and substitution of the initial consonant cluster with a labial element in the second component, is borrowed. Furthermore, the respective semantic and pragmatic (pragmalinguistic, as well as sociopragmatic) functions also have passed to the recipient language. In colloquial Russian, *m*-reduplication is related to situations of bilingualism and other rather complex contact situations, involving Turkic and Caucasian languages. *M*-reduplication is known to be an areal feature, which is spread over a vast territory. The area comprises the languages that have (or used to have in the past) some connection with the Turkish zone of influence, ranging from Armenian to the Balkan Sprachbund (Chirikba 2008; Stolz 2008). The folk joke cited in (41) highlights some sociopragmatic aspects of this phenomenon, such as stigmatization of bilingual speakers of regional varieties of Russian on behalf of speakers of 'standard' Russian. In this case, it is Baku Russian, a variety in which extensive use of *m*-reduplications should be attributed to Azeri influence:

(41) *Priezžajut žurnalisty iz Rossii v Baku i obraščajut vnimanie, čto mestnye žiteli*
      kak-to stranno iz"jasnjajutsja:
      Frukty-mrukty   *prodaem,*
      fruit.ACC.PL.M-MRED.ACC.PL.M sell.IPFV.PRS.1PL
      šašlyk-mašlyk   *žarim,*
      barbeque.ACC.SG.M-MRED.ACC.SG.M fry.IPFV.PRS.1PL
      neft'-meft'   *dobyvaem…*
      oil.ACC.SG.F-MRED.oil.ACC.SG.F extract.IPFV.PRS.1PL
      Prichodjat za raz"jasnenijami k ministru kuľtury Azerbajdžana. – Skažite
      požalujsta, počemu u vas tak stranno govorjat: proiznosjat slovo, a zatem
      povtorjajut ego, zameniv pervuju bukvu na "m" ? – Nu kak vam ob"jasniť…
      *Navernoe, prosto **kuľtur-muľtur**   ne*
      apparently simply culture.GEN.PL.F-MRED.GEN.PL.F NEG
      chvataet.
      be.enough.IPFV.PRS.3SG
      'Journalists come from Russia to Baku and notice that local people express
      themselves in a somewhat weird manner: "we sell **fruit-mruit**, we cook
      barbeque-marbeque, we extract **oil-moil**…" So, they come to the minister of
      culture of Azerbaijan and ask him for explanations, "Tell us, please, why do
      your people speak in such a weird way? They utter a word but then they repeat
      it, substituting the first letter with an *m*." – "Well, how can this be explained…
      Apparently, there is a lack of **culture-multure**."

                                          [ARM, <https://rusenergy.com>]

As can be evinced from the above example, the construction can be used with the
ironic metalinguistic intent to downgrade the Baku variety of Russian and its speak-
ers, "putting a spotlight on links between pejoration and other cognitive domains"
(Wiese & Polat 2016: 243).

     Both functions typically associated with *m*-reduplication, namely, blurred sim-
ilarity and pejorative/distancing connotation, can co-exist within the same context:

(42) *To esť proverjaď / podchodiď ostorožno ko vsemu /*
      potomu čto edut   ne toľko xorošie   ljudi
      because   go.IPFV.PRS.3PL NEG only good.NOM.PL people.NOM.PL
      ottuda /   no i   edut   vsjakie
      from.there but and/also go.IPFV.PRS.3PL various.NOM.PL
      šušera-mušera   *i   kriminal.*
      scum.NOM.SG.F-MRED.NOM.SG.F and criminals.NOM.SG.M
      'That is to say [you have to] control, be cautious with everything / because not
      only nice people come from there / but also all that **scum** and criminals.'

In (42), the categorization process is directly triggered by the presence of the definite quantifier *vsjákie* 'every; all'. Furthermore, the co-compound *šúšera-múšera* [scum.SG.F-MRED.SG.F] 'scum and the like' is followed by another collective noun, *kriminál* 'criminals; criminal world', which guides the hearer towards the intended category, [SOCIALLY DANGEROUS PEOPLE].

Yiddish-style *shm*-reduplication, which is also rather common in Russian, is functionally similar to *m*-reduplication. However, it is mostly associated with downgrading the referent. In the following example, extracted from a Russian translation of Romain Gary's novel *La Danse de Gengis Cohn*, a negative, sarcastic attitude towards the concept of happiness (*sčást'e*) is expressed by the imperfect reduplication *sčást'e-šmást'e* [happiness.SG.N-SHMRED.SG.N]:

(43) - […] *Nakonec-to ona budet udovletvorena! Amerikancy dadut ej sčast'e!*
      - ***Sčast'e-šmast'e.***   *Ničego   u   amerikancev*
      happiness.NOM.SG.N-SHMRED.NOM.SG.N nothing.GEN at American.GEN.PL.M
      ne   polučitsja.
      NEG manage.PFV.FUT.3SG
      Sliškom oni pylkie, sliškom toroplivye, sliškom neterpelivye, vse oni pomešany na skorosti, tak čto končitsja u nich pšikom; […]
      '[…] She finally will be satisfied! Americans will bring her happiness!
      Happiness-shmappiness. Americans won't make it. They are too ardent, too hasty, too impatient, they are all crazy about speed, so it's all fluff.'
                                     [ARM, <http://lib.ru/INPROZ/GARI/la_danse.txt>]

In the example cited below, the categorizing function of the co-compound *krízisy-šmízisy* [crisis.PL.M-SHMRED.PL.M], is intertwined with the use of plural morphology, the quantifier (*vse* 'all') and the demonstrative determiner (*èti* 'these'), followed by the attributive adjective *obščemirovýe* 'global'. The latter items widen the semantic domain of the reduplicated item. The extension of the word *krízis* 'crisis' is widened, while its intension is modified. The large scale of the phenomenon and its supposedly scarce impact on Russian economy are hence implied at the same time. The reduplicative form thus contributes to designate a new concept, [GLOBAL CRISES WHOSE IMPACT IS LITTLE FELT IN RUSSIA]:

(44) *A   vse   èti   obščemirovye*
      and all.NOM.PL this.NOM.PL global.NOM.PL
      krizisy-šmizisy   *do   nas   dokatyvajutsja*
      crisis.NOM.PL.M-SHMRED.NOM.PL.M up.to 1PL.GEN reach.IPFV.PRS.3PL.REFL
      oposredovano, sil'no   oslabev.
      indirectly   strongly weaken.PFV.GER
      'All these global **crises-shmises** only indirectly affect us, having lost much of their strength.'   [ARM, <https://nstarikov.ru/blog/18351>]

We have seen that both *m-* and *shm-*reduplications share the same ludic and down-grading function and the same meaning of 'blurred similarity', which contribute to widen the reference of existing concepts and to design new ones in the communicative process.

Numerous occurrences for *m-*reduplication come from the web sources that are related to the geographical areas where *m-*reduplication is known to be highly diffused (in particular, Caucasian region). In turn, *shm-*reduplication is often found on web pages related to Jewish culture. These facts confirm the sociolinguistic relevance of the phenomena in question as identity features.

Both *m-* and *shm-*patterns display some highly lexicalized forms, i.e. *šašlýk-mašlýk* [shashlik.SG.M-MRED.SG.M] 'kebab; shashlik', and *táncy-šmáncy* [dance.PL.M-SHMRED.PL.M] 'dances'. Their use is primarily ludic with proper names, such as, for instance, *Mál'boro-Šmál'boro* [Marlboro-SHMRED] or *Pútin-mútin* [Putin.SG.M-MRED.SG.M], where only a pejorative connotation, but not a generalization of the meaning of the noun, can be observed.

The third type of imitative co-compounds involves reduplicative phenomena based on obscene/taboo lexical morphemes. Russian obscene lexical morphemes easily fit into different word-formation patterns as nonreferential elements, whereby the morphological construction itself is meaningful. Krongauz (1998: 28) termed them as word-formation jokers. When they are involved in the formation of co-compounds, we have a reduplication, in which the pretonic part of the base is replaced with a taboo word (typically, *chuj* 'penis').[16] Just like the *m-* and *shm-*reduplication, the *chuj-*reduplication also encodes the categorizing and evaluative (specifically, pejorative) function. Euphemisms can also occur in the same position, as in the notorious example from Solženicyn's *One Day in the Life of Ivan Denisovich*, cited below (via Belikov 1990 and Rozhanskiy 2015). The phonetic form of the taboo word is euphemistically reshaped (/χ/ is substituted with the labiodental fricative /f/). Moreover, in this specific case, the emphatic coordinate conjunction *da* is inserted between the conjuncts, thus transforming the co-compound into a coordinate phrase:

(45) *Na stolike   u nich   **maslice***   ***da***
on table.PREP.SG.M at 3PL.GEN butter.DIM.NOM.SG.N and
fujaslice…
TABOORED.DIM.NOM.SG.N
'On the table they have **f***ing butter and what not**.'
      [A. Solženicyn, *Odin den' Ivana Denisoviča*, 1961]

---

16. For an account of the phonological and prosodic properties of the *chuj-*pattern, see Kukhto & Piperski (2017).

The reduplicative syntactic construction in (45) above refers to butter and other unspecified food on the table. The modified phonetic form of the taboo root in the reduplicant adds a derogatory meaning. In fact, butter and the like are associated, for the character of the novel, with a different life he has nothing to do with, while he is heading towards death.

The use of the 'phallic' euphemism *chren* (literally 'horseradish') instead of the original taboo word is highly common. In (46), it substitutes the entire group of pretonic syllables, involving not only the root morpheme but also the suffix *-iz-* of the deverbal noun *modernizácija* 'modernization', as part of the base verb ([[*modern-iz*]-*ácija*] < *modern-ír-ir-ova-t'* 'modernize'):

(46)   *Da tut nikto ne spravitsja kogda v strane idet politika uničtoženija naselenija, a v uši dujut, čto*
     *vse   prekrasno, **modernizacija-chrenacija**…*
     all.NOM.SG.N fine.SG.N modernization.NOM.SG.F-CHRENRED.NOM.SG.F
     'Nobody will ever solve anything here. When in a country there is a policy of destruction of own people, and in the meanwhile they put it in your ears that everything is fine, **modernization and all that stuff**…'

                                    [Web, <http://biwork.ru>]

The reduplicant echoing the noun *modernizacija* suggests that the policy of modernization, proclaimed by authorities, is a deliberately created illusion. Hence, reduplication here does not have a generalizing meaning. It only affects the connotative level, but not the denotative one.

It should be noted that occurrences of the form *chrenizácija*, where the suffix *-iz-* is retained and hence only the root morpheme of the base noun is substituted, are also found on the Web. However, only some of them really occur as the second constituent of a co-compound. In most cases, it is inserted in a listing construction of nouns with a morphological rhyme pertaining to the same lexical semantic domain as a closing element. The *chren*-element contributes to the construction of an *ad hoc* category and simultaneously downgrades the referent:

(47)   *PASE-MASE, **organizacija-chrenizacija**…*
     PACE-MRED organization.NOM.SG.F-CHRENRED.NOM.SG.F
     'F***ing PACE (=Parliamentary Assembly of the Council of Europe), **f***ing organization and all that stuff**…'

                       [Web, <https://fishki.net/anti/1412008-chto-my-poterjali-na-uhod-rossii-iz-pase.html>]

(48) *Sejčas že* **optimizacija,** **modernizacija** *i*
now EMPH optimization.NOM.SG.F modernization.NOM.SG.F and
pročaja **chrenizacija** *mediciny,* *poètomu*
other.NOM.SG.F CHRENRED.NOM.SG.F health.service.GEN.SG.F that's.why
vrači *naznačajut* *samyj* *minimum.*
doctor.NOM.PL.M prescribe.IPFV.PRS.3PL very.ACC.SG.M minimum.ACC.SG.M
'Nowadays there is an ongoing **optimization, modernization and what not** of
health service, that's why doctors prescribe only the bare minimum (of tests).'
[Web, <http://www.woman.ru/health/woman-health/thread/4966211/>]

(49) *Načaľniki* *govorjat,* *čto k tebe* *lično*
boss.NOM.PL.M say.IPFV.PRS.3PL that to 2SG.DAT personally
pretenzij *net,* *no u nich* **restrukturizacija,**
complaint.GEN.PL.F NEG.EXIST.PRS but by 3PL.GEN restructuring.NOM.SG.F
reorganizacija, **optimizacija,** *eščë* **kakaja-to**
reorganization.NOM.SG.F optimization.NOM.SG.F yet/more some.NOM.SG.F
chrenizacija, *im* *nužno* *kogo-to* *uvolit'* […]
CHRENRED.NOM.SG.F 3PL.DAT is.needed someone.ACC.SG fire.PFV.INF
'The bosses say that they have no complaints against you but they are under-
taking a restructuring, reorganization, optimization, some other ***ization, so
they need to fire someone.'
[Web, <https://www.rulit.me/books/komnata-si-read-523670–2.html>]

Examples, as those cited in (47)–(49) above, are yet another piece of evidence,
showing to what extent the problem of gradience between syntax and morphology
is relevant for categorizing co-compounds and, specifically, for reduplication pat-
terns, based on taboo roots. As is well-known, uninterruptibility, i.e. the impossi-
bility of insertion of other lexical material between the constituents, is among the
major criteria of wordhood. For instance, consider (50) below, where the adjectival
general extender *pročij* 'other, further' is inserted between the noun *visjúľka* 'charm,
pendant, useless hanging object' and its *chren*-reduplication echo form *chrenúľka*,
explicitly connected by the coordinate conjunction *i*. What is interesting, the
co-compound *visjúľka-chrenúľka* is perfectly grammatical in Russian, whereas the
respective coordinate phrase **visjúľka i chrenúľka* without the general extender is
unacceptable. The construction clearly has a categorizing meaning, which is rein-
forced by the general extender *próčij*. However, the nominal base *visjúľka* occurs in
the singular and, just like other nouns in *-uľka*, such as *figúľka* 'small trifling object',
it has singulative semantics and is never used as a collective noun with plural ref-
erence in the singular (plural reference requires plural marking). The construction
forces the extension of the referential domain of the otherwise singulative noun and
is an instance of gradience between a morphological construction (a co-compound)
and a syntactic listing structure:

(50) *Esli* **"visjuľ'ka"** *i* *pročaja* **"chrenuľ'ka"**
if charm.NOM.SG.F and other.NOM.SG.F CHRENRED.NOM.SG.F
zakryvaet *obzor* *choť* *na santimetr,* *to*
cover.IPFV.PRS.3SG view.ACC.SG.M at.least on santimeter.ACC.SG.M then
ne *prosto normaľnyj* *srednij* *voditeľ,* *a*
NEG just normal.NOM.SG.M average.NOM.SG.M driver.NOM.SG.M but
ljuboj *zdravomysljaščij* *čelovek*
any.NOM.SG.M reasonable.NOM.SG.M person.NOM.SG.M
ponimaet, *čto èto* *– pomecha.*
understand.IPFV.PRS.3SG that this obstacle.NOM.SG.F
'If the **charm, and/or any other hanging stuff**, blocks the view, even a centimeter
of it, not only a normal average driver but any reasonable person realizes that
it is a distraction.' [Web, <http://rusdtp.ru>]

(51) is an excerpt from a speech by Natal'ja Poklonskaja, who was appointed prosecutor of Crimea after its annexation by Russia. The speaker uses the the *m*-reduplication *njáša-mjáša* [njaša.SG.F-MRED.SG.F], based on her popular nickname Njaša. It performs the categorizing function, simultaneously associated with a component of downgrading, directed against those who invented the nickname:

(51) *Zdes' ja* *prokuror.* *I* *poètomu* *nikakich*
here 1SG.NOM prosecutor.NOM.SG.M and that.is.why not.any.GEN.PL
njaš, **mjaš** *i* *tomu* *podobnoe*
nyasha.GEN.PL.F MRED.GEN.PL.F and that.DAT.SG.N similar.NOM.SG.N
ja *ne* *dopušču.*
1SG.NOM NEG tolerate.PFV.FUT.1SG
'I am a prosecutor here. That is why I'm not going to put up with any **nyashas, myashas or anything like that**.'
[Web, <https://www.youtube.com/watch?v=Wm95qSmru28>]

Natal'ja Poklonskaja is known to have become popular due to her supposed resemblance with an anime character. She was initially referred to in Russian social media as *njáša* 'sweetee; darling' (a term from anime slang, based on the Japanese onomatopoeic form *nyā* 'meow', denoting the property of cuteness). In Poklonskaja's speech, the *m*-reduplication refers to the *ad hoc* category [NYASHA AND OTHER STUPID CUTE STUFF] and has the pejorative function. The categorization effect is further reinforced by the general extender *i tomú podóbnoe* 'and similar things'. In morphological terms, the Russian word *njáša* is a female noun in *-a* of the first declension. However, as the above-cited utterance became viral, the genitive plural form with its zero inflectional suffix was reanalyzed as a masculine nominative singular noun of the second declension, thus becoming the new nickname of the politician. Furthermore, it was inserted in the rhyming pattern *njaš mjaš – Krym*

naš 'Njaš-Mjaš, Crimea is ours', a parody variation of the "Crimea is ours" nationalist slogan, which, in turn, rapidly spread as a meme. This on-going signification process highlights the importance of the ludic component in online meaning construction.

4.   Conclusions

In terms of word-formation patterns, two main types of reduplicative phenomena with a categorizing function exist in Russian, rhyming co-compounds, which in turn can be further classified on semantic, as well as formal, grounds (the appositive subtype, the prefixed subtype etc.), and imitative co-compounds, involving *m-, shm-* and taboo word reduplication. These phenomena rise issues regarding the boundary between syntactic phrases and compound words. In principle, corpus data suggest that there is a fuzzy boundary between syntactic (asyndetic non-exhaustive lists) and morpho-lexical means (co-compounding), and different criteria of wordhood themselves reveal this fuzziness.

When ad hoc categories are constructed, overcoding is a pervasive phenomenon. Hence, different means of categorization, involving, among others, general nouns, general extenders, and non-exhaustive listing structures, tend to cluster together within the same contexts. Background and contextual knowledge, as well as the co-text play a key role in guiding the selective information processing, fundamental in *ad hoc* categorization. In co-compounds, the abstractive categorizing process is supported by their morpho(no)logical structure, which is also intertwined with the evaluative function (irony or negative attitude). Thus, within the process of meaning construction, the categorization mechanism interacts with the evaluative component both in terms of semantics and pragmatics.

Formal (phonological and morphemic) similarity between the components of co-compounds allows to establish 'blurred similarity' and semantic relatedness between their meanings (the entities may pertain to the same frame, (Fillmore 1985) or be members of the same lexical semantic domain etc.). This mechanism widens the referential domain of lexical items, pertaining to the common lexical stock. Consequently, it enhances the construction of new context-dependent concepts and categories, as in the case with the appositive and prefixed subtypes of rhyming co-compounds, or triggers generalization and extension of the lexical meaning of the base, as in the case with imitative co-compounds.

Rhyming patterns typically present instances of playful language. They contribute to the creation of new lexical items, especially in case of lexical gap, and simultaneously create an effect of ludicity, which, in turn, often results in downgrading

the referent and conveying a negative connotation. Wiese and Polat's (2016: 243) observation regarding urban German that *m*-reduplication exhibits "constructional pejoration" of meaning of the referent, in other words, that it is pattern-based rather than associated with individual evaluative items, equally holds for the Russian data.

These mechanisms may act independently. Thus, for instance, not all rhyming co-compounds are associated with a negative connotation. Accordingly, imitative co-compounds that do not exhibit the categorizing function are also possible. However, prototypically, these two mechanisms co-exist. Hence, when co-compounding is exploited in *ad hoc* categorization and conceptualization, it is typically associated with an evaluative function.

Acknowledgements

References

Aarts, Bas. 2007. *Syntactic Gradience: The Nature of Grammatical Indeterminacy*. Oxford: OUP.

Arcodia, Giorgio F., Grandi, Nicola & Montermini, Fabio. 2009. Hierarchical NN compounds in a cross-linguistic perspective. *Italian Journal of Linguistics* 21(1): 11–33.

Armoskaite, Solveiga & Kutlu, Deniz Aysegul. 2014. Turkish *m*-reduplication: A case of similative plural. *Turkic Languages* 18(1–2): 271–288.

Baayen, R. Harald. 1993. On frequency, transparency, and productivity. In *Yearbook of Morphology 1992*, Geert E. Booij & Jaap van Marle (eds), 181–208. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-017-3710-4_7

Barsalou, Lawrence W. 1983. Ad hoc categories. *Memory and Cognition* 11: 211–277. https://doi.org/10.3758/BF03196968

Barsalou, Lawrence W. 1985. Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 11: 629–654.

Belikov, Vladimir N. 1990. Produktivnaja model' povtora v russkom jazyke (variant dlja obsuždenija) (A productive model of repetition in Russian (data for discussion)). *Russian Linguistics* 14(1): 81–86.

Benigni, Valentina. 2016. Le marche di lista in russo. Segnali riformulativi, estensivi, generalizzanti…*i vse takoe*. In *Le lingue slave tra struttura e uso*, Valentina Benigni, Lucyna Gebert & Julija Nikolaeva (eds), 17–43. Firenze: Firenze University Press. https://doi.org/10.36253/978-88-6453-328-5

Benigni, Valentina. 2018. *Ad hoc* categorization in Russian and multifunctional general extenders. In *Linguistic Strategies for the Construction of Ad Hoc Categories: Synchronic and Diachronic Perspectives*, Caterina Mauri & Andrea Sansò (eds). Special issue of *Folia Linguistica* 52(s39–1): 97–123. https://doi.org/10.1515/flih-2018-0003

Benko, Vladimír. 2014. Aranea: Yet another family of (comparable) web corpora. In *Text, Speech and Dialogue. 17th International Conference, Brno, Czech Republic, September 8–12, 2014*, Petr Sojka, Aleš Horák, Ivan Kopeček & Karel Pala (eds), 247–256. Cham: Springer. https://doi.org/10.1007/978-3-319-10816-2_31

Bisetto, Antonietta & Scalise, Sergio. 2005. The classification of compounds. *Lingue e Linguaggio* 4(2): 319–332.

Carston, Robyn. 2010. Metaphor: Ad hoc concepts, literal meaning and mental images. In *Proceedings of the Aristotelian Society* 110: 295–321. Oxford: OUP. https://doi.org/10.1111/j.1467-9264.2010.00288.x

Chirikba, Viacheslav A. 2008. The problem of the Caucasian Sprachbund. In *From Linguistic Areas to Areal Linguistics* [Studies in Language Companion Series 90], Pieter, Muyksen (ed.), 25–92. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.90.02chi

Corbett, Greville G. 1983. *Hierarchies, Targets, and Controllers: Agreement Patterns in Slavic*. London: Croom Helm.

Dalrymple, Mary & Nikolaeva, Irina. 2006. Syntax of natural and accidental coordination: Evidence from agreement. *Language* 82(4): 824–849. https://doi.org/10.1353/lan.2006.0189

Dressler, Wolfgang U. & Barbaresi, Lavinia M. 1994. Morphopragmatics in inflection. In *Morphopragmatics: Diminutives and Intensifiers in Italian, German, and Other Languages*, Wolfgang U. Dressler & Lavinia M. Barbaresi (eds), 58–83. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110877052.58

Fillmore, Charles J. 1985. Frames and the semantics of understanding. *Quaderni di Semantica* 12: 222–254.

Haspelmath, Martin. 2011. The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica* 45(1): 31–80. https://doi.org/10.1515/flin.2011.002

Janda, Laura A., Endresen, Anna, Kuznetsova, Julia, Lyashevskaya, Olga, Makarova, Anastasia, Nesset, Tore & Sokolova, Svetlana. 2013. *Why Russian Aspectual Prefixes Aren't Empty: Prefixes as Verb Classifiers*. Bloomington IN: Slavica.

Janda, Laura A. & Nesset, Tore. 2010. Taking apart Russian *raz-*. *The Slavic and East European Journal* 54(3): 477–502.

Josephson, Folke. 2015. Grammaticalization paths of verbal prefixes in Slavic and other Indo-European branches. *Scando-Slavica* 61(2): 283–292. https://doi.org/10.1080/00806765.2015.1109192

Krongauz, Maksim A. 1998. *Pristavki i glagoly v russkom jazyke: Semantičeskaja grammatika* (Prefixes and verbs in Russian: Semantic grammar). Moskva: Škola "Jazyki russkoj kul'tury".

Kukhto, Anton & Piperski, Alexander. 2017. *Xuj*-Reduplication in Russian. Paper presented at the 14th Old World Conference in Phonology (OCP14), Düsseldorf, Germany, February 20 –22, 2017.

Le Bruyn, Bert, Que, Min & de Swart, Henriëtte. 2012. The scope of bare nominals. In *Genericity*, Alda Mari, Claire Beyssade & Fabio Del Prete (eds), 116–139. Oxford: OUP. https://doi.org/10.1093/acprof:oso/9780199691807.003.0003

Lipec, Rachil' S. 1939. Vvodnaja stat'ja (Introductory article). In *Byliny M.S. Krjukovoj*, Èrna G. Borodina & Rachil S. Lipec (eds) 1–48. Moskva: Gosudarstvennyj literaturnyj muzej.

Masini, Francesca, Mauri, Caterina & Pietrandrea, Paola. 2018. List constructions: Towards a unified account. *Italian Journal of Linguistics* 30(1): 49–94.

Mauri, Caterina & Sansò, Andrea. 2018. Linguistic strategies for ad hoc categorization: Theoretical assessment and cross-linguistic variation. In *Linguistic Strategies for the Construction of Ad Hoc Categories: Synchronic and Diachronic Perspectives*, Caterina Mauri & Andrea Sansò (eds). Special issue of *Folia Linguistica* 52(s39–1):1–35. https://doi.org/10.1515/flih-2018-0001

Minlos, Philip R. 2004. Reduplikacija i parnye slova v vostočnoslavjanskich jazykach (Reduplication and pair words in East Slavic languages). Candidate's degree dissertation, Rossijskaja akademija nauk, Moskva.

Pacsai, Imre. 2013. Vostočnye èlementy russkoj narodnoj reči (leksika i derivacija) (The eastern elements in the russian folk speech (vocabulary and derivation)). *Folia Linguistica Rossica* 9: 43–56.

Plotnikova, Anna M. 2012. Parnye sočetanija v sovremennom russkom jazyke i ich kreativnyj potencial (Pair word combinations in Modern Russian and their creative potential). *Ural'skij Filologičeskij Vestnik* 2: 183–187.

Rozhanskiy, Fedor I. 2015. Two semantic patterns of reduplication. Iconicity revisited. In *The Why and How of Total Reduplication: Current Issues and New Perspectives*, Daniela Rossi (ed.). Special issue of *Studies in Language* 39(4): 992–1018.

Scalise, Sergio & Bisetto, Antonietta. 2011. The classification of compounds. In *The Oxford Handbook of Compounding*, Rochelle Lieber & Pavol Štekauer (eds), 34–53. Oxford: OUP.

Stolz, Thomas. 2008. Total reduplication vs. echo-word formation in language contact situations. In *Language Contact and Contact Languages* [Hamburg Studies on Multilingualism 7], Peter Siemund & Noemi Kintana (eds), 107–134. Amsterdam: John Benjamins. https://doi.org/10.1075/hsm.7.07sto

Tkačenko, Orest B. 1979. *Sopostaviteľno-istoričeskaja frazeologija slavjanskich i finnougrorskich jazykov* (Comparative-historical phraseology of Slavic and Finno-Ugric languages). Kiev: Naukova dumka.

Voinov, Vitaly. 2012. Rhyming reduplication in Russian paired words. *Russian Linguistics* 36(2): 175–191. https://doi.org/10.1007/s11185-012-9091-5

Wälchli, Bernhard. 2005. *Co-compounds and Natural Coordination*. Oxford: OUP. https://doi.org/10.1093/acprof:oso/9780199276219.001.0001

Ward, Dennis. 1973. Appositional compounds in Russian. *The Slavonic and East European Review* 51(122): 1–10.

Weiss, Daniel. 1993. Dvojnye glagoly v sovremennom russkom jazyke (Double verbs in Modern Russian). In *Kategorja skazuemogo v slavjanskix jazykax: Modaľnosť i aktualizacija* (The category of predicate in Slavic Languages: Modality and actualization) [Slavistische Beiträge 305], Francesca Fici Giusti & Simonetta Signorini (eds), 67–97. Munich: Otto Sagner.

Weiss, Daniel. 2013a. Dvojnye glagoly v russkoj razgovornoj reči v zerkale Nacionaľnogo korpusa: formy imperativa množestvennogo čisla (Double verbs in Colloquial Russianon on the basis of data from the National Corpus: Imperative plural forms). In *Schweizerische Beiträge zum XV. Internat. Slavistenkongress in Minsk, August 2013*, Ekaterina Veľmezova (ed.), 319–341. Bern: Peter Lang.

Weiss, Daniel. 2013b. Russian double verbs in the 1st pl Imperative. *Wiener Slawistischer Almanach* 85: 165–175.

Wiese, Heike & Polat, Nilgin Tanis. 2016. Pejoration in contact: *m*-reduplication and other examples from urban German. In *Pejoration* [Linguistik Aktuell/Linguistics Today 228], Rita Finkbeiner, Jörg Meibauer & Heike Wiese (eds), 243–268. Amsterdam: John Benjamins. https://doi.org/10.1075/la.228.11wie

Wilson, Deirdre & Carston, Robyn. 2007. A unitary approach to lexical pragmatics: Relevance, inference and ad hoc concepts. In *Pragmatics*, Noel Burton-Roberts (ed.), 230–259. Houndmills: Palgrave. https://doi.org/10.1057/978-1-349-73908-0_12

CHAPTER 13

Encoding ad hoc categories in Georgian

Three types of echo-word constructions

Zaal Kikvidze

Ivane Javakhishvili Tbilisi State University

The chapter is a discussion of Georgian echo-word constructions (non-prototypical reduplication) as a linguistic encoding of ad hoc categories. Based on the empirical data (published texts, corpora, elicited examples), I present a classification of Georgian echo-word constructions into three types with their respective formal and semantic features. With respect to possible parameters of base/reduplicant combinations, such as (a) free/bound form, (b) kind and direction of copying, (c) meaning of resulting formation, the following three types were identified: Type 1: Turkish-like *m*-reduplication; Type 2: Initial C/cluster deletion; Type 3: Fossilized echo-pairs. Notwithstanding a type, they all, to a certain extent, participate in the encoding of ad hoc categories in Georgian.

Keywords: ad hoc categories, Georgian, echo-word constructions, reduplication, base/reduplicant relation

1. Introduction

The principal objective of the present chapter is to provide a linguistic description of various types of echo reduplication in Georgian as means of the linguistic encoding of ad hoc categories. I will present a classification of Georgian echo-word constructions (non-prototypical reduplication) into three types with their respective formal and semantic features. Based on possible parameters of base/reduplicant combinations, such as (a) free/bound form, (b) kind and direction of copying, (c) meaning of resulting formation, the following three types were identified: Type 1: Turkish-like *m*-reduplication; Type 2: Initial C/cluster deletion; Type 3: Fossilized echo-pairs. They are dealt with here because, notwithstanding a type, they all, in some way or another, participate in the encoding of ad hoc categories in Georgian.

When Lawrence Barsalou provided the seminal analysis of ad hoc categories, alongside with cognitive psychological factors, he acknowledged the participation of linguistic means in their construal:

> Both conceptual and linguistic mechanisms appear central to forming ad hoc categories. Conceptually, people combine existing concepts for objects, events, settings, mental states, properties, and so on to form novel conceptual structures. Linguistically, people combine words in novel ways to index these concepts
>
> (Barsalou 2011: 86)

A cross-linguistic perspective of relevant linguistic means is definitely an effective way for making a more or less complete picture possible. It was Caterina Mauri who provided an outline of the means in question; she stated that the aforementioned perspective showed great variation in the types of strategies that might be used in this function, echo reduplication among them (Mauri 2017: 307). It is noteworthy that she went on to describe the strategies as being

> characterized by the explicit mention of one or more exemplars, used as a starting point to infer some high-order entity. Depending on the relationship between the exemplars, the inferential process can lead to the encoding of a *set*, if the exemplars are elements that co-occur in combination […], to the encoding of a *class*, if the exemplars are equivalent alternatives […], or to the construction of a [*narrative*] *frame*, if the exemplars are actions that occur within a narrative scheme
>
> (Mauri 2017: 302)

It should be emphasized that Mauri and her colleagues' endeavors in this direction, conceiving of linguistic entities not as just referring to or reflecting categorization but rather contributing to the encoding of categories, bore good fruit when a special issue of *Folia Linguistic Historica* (Mauri & Sansò 2018a) was dedicated to the problem in point. The aforementioned provides a beneficial background to researchers of linguistic strategies applied for the encoding of ad hoc categories in languages so far never been dealt with in terms to the phenomenon in question. There is another nonetheless favorable condition: reduplication, as a linguistic phenomenon, has been extensively investigated both in Georgian and cross-linguistically.

Certain features and uses of echo reduplication have been discussed within contexts which greatly resemble those associated with ad hoc categorization; hence, I can in no way neglect them as far as they are very bright illustrations of this kind of categorization. In his paper about echo-word constructions in Bhojpuri (an Indo-Aryan language) Kapil M. Tiwary describes the following situation:

> In this community of Bhojpuri speakers, particularly among the members of the upper caste, smoking in the presence of the parents, or of one's social seniors in general, is considered bad manners. In fact, it is not only that one does not smoke in front of the parents, one always tries to keep one's habit of smoking itself a secret. Consequently, in the presence of the parents, one might implicitly ask someone to buy cigarettes for him from the market by overtly asking him to buy /deslai-oslai/ matchbox, etc., since a matchbox is closely associated with with smoking
>
> (Tiwary 1968: 37)

Albeit not mentioning ad hoc categories explicitly, the passage provides a description of a typical case how one is encoded by means of echo reduplication.

Besides, the list-like property of echo-pairs has been observed in literature without special reference to their participation in the encoding of ad hoc categories:

> echo-word formation is often described as deprecative and/or list-like. Therefore, the copy is most often translated by circumscriptions such as *and the like*, etc. This list-like property of the construction makes it, at the same time, inherently plural independent of number marking on its constituents. For instance Turkish *kitap mitap* 'books and the like' contains two word-forms unmarked for number. *Mitap* then stands for referents which are similar though not identical to those of *kitap*"
>
> (Stolz 2008: 126)

In the present chapter, I will initially present a set of features on which the classification of Georgian echo reduplication rests and, hence, the classification itself. Consequently, each type will be described and exemplified with illustrations from published work; their structural and semantic peculiarities will be identified; meanwhile, it will be established whether an individual type is productive or not. The occurrence of echo reduplication in a language may in no way imply that echo pairs participate in the encoding of ad hoc categories in that language. Therefore, based on lexicographic, corpus and elicited data, the following paragraph discusses various Georgian echo constructions in terms of their possible contribution to the construal of the aforementioned kind of categories.

2.   Echo reduplication in Georgian: Features, types, and classification

Since August Friedrich Pott, who in 1862, for the first time in linguistics, divided *Doppelung* ('doubling') into two sub-types – *Gemination* and *Reduplikation*, viewing the former, as different from the present-day understanding of the term, as *Total Reduplication* ('Wiederholung im Ganzen'), and the latter as *Partial Reduplication* ('verkürzte und nur zum Theil, also bloß andeutungsweise vollzogene Wiederholung') (Pott 1862: 16), *Total Reduplication* and *Partial Reduplication* have been normally treated as prototypical sub-types within various classifying frameworks (see, for instance, Hurch & Mattes 2009; Inkelas & Zoll 2005; Moravcsik 1978; Raimy 2000; Rubino 2005; Stolz et al. 2011). A base and a reduplicant are the two whales on which reduplicated formations rest. A base is a segment which serves as a material to undergo copying, while a reduplicant is a product of this copying; with respect to completeness of a copied product, Total (or Full or Complete) Reduplication and Partial Reduplication are identified. Owing to the said approach, reduplicative patterns like *ablaut reduplication* and *echo reduplication* have been dealt with as "the phenomena that deviate far from the prototype and tend to the

opposite pole" (Urdze 2018: VII) – hence, the label *non-prototypical reduplication*. As far as in the present chapter I am dealing with echo constructions,[1] it should be necessarily mentioned that in the literature they are also referred to, among others, as *rhyming reduplication, rhyme-motivated reduplication, Vokalharmonische volle Reduplikation, heterogen (Reimbildungen)* (Bzdęga 1965), *ricochet words, reduplication with fixed segmentism* (McCarthy et al. 1999), *distortive reduplication with [m]* (Konstantinidou 2005), *total reduplication cum variation (=TRCV)* (Stolz 2008), *reduplication with m* ("m'li ikilime" = "mühleme") (Johanson 2002), *echo-compounding* (Steever 2009: 651), etc.

Various types of reduplication have been normally identified with respect to Base-Reduplicant relations; hence, initially, it is appropriate to define them as axial constituents:

> The *Reduplicant* R is the actual phonological projection of some reduplicative morpheme RED$_i$ which has a phonologically – unspecified lexical entry. […] The *Base* B is the phonological material to which the reduplicant is attached – for reduplicative *prefixes*, the following structure, and for reduplicative *suffixes*, the preceding structure   (McCarthy & Prince 1994: 339)

As for echo reduplication, which is a case in point in the present chapter, it does not, in my opinion, belong to either type (Total and Partial), and, hence, it should be considered a special case. Its very specific character is well described as follows:

> It consists of two immediately adjacent strings of segments which differ as to the quality of the filler of one of their slots, meaning: they share all segments but one. This partial phonological identity has another effect which jeopardizes the word-hood of the string of the segments which constitutes the copy in the pair of strings. This string which displays one segment which is not copied from the original string often is a non-word and thus only exists in combination with the original string (i.e. it is bound to the co-occurrence of the original and appears only in this particular construction). The non-word itself has no autonomous semantics
> (Stolz 2008: 109)

The phenomenon of reduplication in Georgian and in Kartvelian (South Caucasian) languages at large has been discussed in a number of works (for instance, Datukishvili 1990; Gersamia et al. 2016; Kikvidze et al. 2018; Neisser 1953; Sanikidze 1968, 1976; Topuria 1979); however, concerning echo reduplication, no clear-cut and strict

---

1. The term 'echo-word' was borrowed "from a discussion of certain noun-formations found in many, if not most, of the languages of India, Indo-Aryan, Dravidian, and Munda, a discussion which was published in the Circulars of the Linguistic Society of India in 1928" (Emeneau 1938: 553).

approaches have been set forth for the sake of their adequate treatment (nothing to say about their association with ad hoc categorization).

As long as the base and the reduplicant are two axial segments of any reduplicated construction, I will design my classification of Georgian echo-pairs based on relationships between these entities. The set of the classificatory features includes the following: (i) Identification of a base and a reduplicant (normally, a base is a free form and a reduplicant is a bound form); (ii) Kind of copying (a consonant (mostly, labial); for the most part, -*m* is taken on by a vowel-initial base (prothetic *m*-) or replaces its initial consonant, or is deleted as an initial element of a base); (iii) Copying direction (a reduplicant copies either from the left edge of a base (left-to-right copying) or from its right edge (right-to-left copying) (see Kikvidze 2018 for details of copying direction in the Kartvelian languages)); (iv) Meaning conveyed by a resulting entity.

In accordance with the aforementioned features, I identified the following types of reduplicative echo-constructions in Georgian:

Type 1:   Turkish-like *m*-reduplication
Type 2:   Initial C/cluster deletion in the pre-echo twin
Type 3:   Fossilized echo-pairs

2.1   Type 1

I refer to Type 1 as 'Turkish-like *m*-reduplication;' the set of its characteristic features is as follows:

i.   Base/Reduplicant: a base is a free form and a reduplicant is a bound form
ii.   Kind of copying: the *m*- is either taken on by a vowel-initial base (prothetic *m*-) or replaces an initial consonant
iii.   Copying direction: a reduplicant copies the left edge of a base (left-to-right copying (L-R))
iv.   Meaning conveyed by the resulting entity: 'X and such,' 'X and the like,' etc., based on extension of a notional domain; similative plural; "loose kind of plurality" (McCarthy & Prince 1996: 84)

In this paragraph, I will provide some of the illustrations already attested in published works:

(1)   *bodiši-modiši* 'empty excuses'
        Base: *bodiši* 'apology'
        Reduplicant formation: *m+[b]odiši*

(2) *oxer-moxrebi* 'rabble, layabouts'
    Base: *oxeri* 'orphan, urchin, waif; outcast; forsaken, derelict, destitute; hateful, abominable'
    Reduplicant formation: *m+ox[e]r-eb-i*

(3) *xili-mili* 'fruit and such'
    Base: *xili* '(edible) fruit'
    Reduplicant formation: *m+[x]ili*

(4) *xurda-murda* 'trivial, a triviality'
    Base: *xurda* 'small (change/coins)'
    Reduplicant formation: *m+[x]urda*

The following three ((5), (6), (7)) are also frequently cited ones (Neisser 1953: 52; Niepokuj 1997: 43; Sanikidze 1976; Southern 2005: 83); however, they should be disclaimed for not satisfying some of the aforementioned conditions:

(5) *are-mare* 'around, here and there, surroundings'
    Base: *are* 'area, place, environment'
    Reduplicant formation: **m+are*
    Comment: *are-mare* seems to be a loan-word.

(6) *axlo-maxlo* 'close by, not far off'
    Base: *axlo* 'near, close'
    Reduplicant formation: **m+[m]axlo*
    Comment: *maxlo* is not a full-fledged bound form (hence, it cannot be assumed as a non-word); for instance, it occurs in the following word-forms: *maxlobel(ad)* 'near/close by'; *maxlob(e)li* '**1** near, close; **2** approaching;' **3** close relative/friend; **4** neighbour'; *maxloblad* 'close by, near by'; *maxlobloba* '**1** vicinity; **2** close relations' (Rayfield 2006: 859).

(7) *cwanili-mwanili* 'vegetables and such'
    Base: **cwanili*; actually, *mc'vanili* 'salad greens, green herbs'
    Comment: **cwanili-mwanili* is a non-existent lexeme; moreover, the actual one *mc'vanili* cannot be a base for *m*-reduplication as far as it is an *m*-initial word.[2]

---

2. "One restriction on /m/-reduplication in Georgian derivational morphology is that it is dissimilative: the initial consonant, if there is one, of the root must not be itself a labial. The same restriction appears to apply to the formation of lexoids by /m/-reduplication" (Tuite 2015: 33). I would add that this also applies to words with /m/ as (a non-initial) constituent of the onset structure, e.g. *gmiri* 'hero.' Besides, one should bear in mind that there are plenty of labial-initial prefixes (both inflectional and derivational) in the language. Another restriction is that normally the reduplicant must not homonymize with an already existing word.

Obviously enough, *m*-reduplication is definitely a contact-induced phenomenon:

> A cursory look at languages in the neighborhood of Turkish (including former possessions of the Ottoman empire) reveals that patterns which closely resemble *mühleme* can be found to the south, east, north and west, in languages for which relatively close contacts with Turkish are almost self-evident or historically well documented. These languages belong to different phyla and macrophyla. Moreover, they are also typologically quite different from each other   (Stolz 2008: 116)

Irrespective of its clearly contact-induced nature, similar phenomena have been evidenced in Old Georgian prior to language contact with Turkish; in the same chapter, Stolz (2008: 114) is not certain about "[w]hether Turkish patterns had a reinforcing effect on on the autochthonous ones." Judging from frequent occurrences of *m*-reduplicative constructions in colloquial Georgian as a means of ad hoc categorization, it can be explicitly stated that Turkish patterns did reinforce the already existing ones whether autochthonous or not.

It is particularly significant in terms of the construal of ad hoc categories that the aforementioned phenomenon, applying the borrowed mechanism (*m*-reduplication), is a productive means for echo-word formation in present-day colloquial Georgian.

2.2   Type 2

I prefer the label 'Initial C/cluster deletion in the pre-echo twin' for Type 2 (the label occurs in Southern (2005: 88)). Here is a set of its characteristic features:

i.   Base/Reduplicant: a base is a free form and a reduplicant is a bound form
ii. Kind of copying: a consonant or a consonant cluster is detached from the base and thus added to it as a reduplicant from its left side
iii. Copying direction: a reduplicant copies the right edge of a base (right-to-left copying (R-L))
iv. Meaning conveyed by the resulting entity: extension of a notional domain

This kind of echo-formation is rather rare cross-linguistically; however, it is evidenced in all the Kartvelian (South Caucasian) languages and it can hardly be assumed as a contact-induced phenomenon.[3] Its Georgian illustrations are the following:

(8) *ava-dava* 'flaring quarrel'
     Base: *dava* 'dispute, argument'
     Reduplicant formation: [*d*]*ava*

---

3.   This does not imply that such echo-pairs do not occur elsewhere; they are well attested, for instance, in Indo-Aryan languages (D'souza 1991: 295–296).

(9) *ali-k'vali* 'the spitting image of'
Base: *k'vali* 'trace, imprint'
Reduplicant formation: [*k'v*]-*ali*

(10) *ač'ia-bač'ia* 'senseless talk, stupid twaddle'
Base: *bač'ia* 'rabbit'
Reduplicant formation: [*b*]-*ač'ia*

(11) *ial-k'iali* 'scurrying clouds'[4]
Base: *k'iali* 'radiating, shining, brilliance'
Reduplicant formation: [*k'*]-*iali*

(12) *otlo-totlo* 'soft-boiled (*egg*)'
Base: *totlo* 'soft'
Reduplicant formation: [*t*]-*otlo*

Type 2 is unproductive formation; hence, respective echo constructions represent a closed set. It is also important to note that, owing to their lexical meanings, not all of them can be successfully used to encode ad hoc categories (for instance, (12) *otlo-totlo*). However, some of them seem to be rather appropriate for this kind of categorization; this will be exemplified below (see Section 3).

2.3   Type 3

I label Type 3 as 'Fossilized echo-pairs.' It is identified owing to the fact that pertaining reduplicatives cannot be assigned to either of the aforementioned ones. Its dimensions are the following:

i.   Base/Reduplicant: a base (?) and a reduplicant (?) are both bound forms
ii. Kind of copying: undetermined
iii. Copying direction: undetermined
iv. Kind of meaning conveyed by the resulting entity: extension of a notional domain? similative plural?

As long as both constituents of such an echo-pair are bound forms, that is, nonwords, it is hardly possible, at least, synchronically, to establish which of them are a base and a reduplicant; hence, both the kind of copying and the copying direction are undetermined. This may be due either to the fact that they are borrowed

---

4.   Dictionaries (see, for instance, Rayfield 2006: 727; Tschenkeli 1960: 519) usually refer to its parallel forms such as

(11a) *ial-ciali* 'scurrying clouds'
(11b) *ial-t'iali* 'scurrying clouds'

echo-pairs and/or to the loss of the meaning of a base through time (this is why I chose to refer to them as *fossilized*).

The following illustrations are organized distinctly from those of the preceding types as far as bases and reduplicants are not identifiable (hence, they are not glossed). On the other hand, each construction is provided with a translation from *A Comprehensive Georgian-English Dictionary* (Rayfield 2006); these translations may tell much about the opportunities of (fossilized) echo-pairs for the encoding of ad hoc categories.

(13) *avan-čavani*
      1. factotum, right-hand man, secretary; 2. ins and outs, vicissitudes, secrets going on   (Rayfield 2006: 23)

(14) *ala-bala*
      1. sb who doesn't stop walking; sb who walks with big strides; 2. fuss, ado
                                                                    (op. cit.: 39)

(15) *alan-talani*
      topsy-turvy; *kveqʼana alan-talani xom ar aris!* "The world isn't for up for grabs!"
                                                                    (op. cit.: 40)

(16) *alian-čaliani*
      all sorts of things, gossip, this and that; miambe šeni *alian-čaliani* "Tell me all your news"   (op. cit.: 42)

(17) *ančʼi-mančʼi*
      weaver's reed, comb and shuttle   (op. cit.: 79)

(18) *aruk-maruki*
      everything good or bad   (op. cit.: 91)

(19) *aruk-paruki*
      spoilt, precious, namby-pamby   (op. cit.: 91)

(20) *aur-zauri*
      1. row, ado, turmoil; 2. impropriety   (op. cit.: 106)

(21) *aqʼal-maqʼali*
      quarrel, scuffle, brawl   (op. cit.: 128)

(22) *ača-bača*
      various things   (op. cit.: 133)

(23) *inči-binči*
      the slightest; kartuli inči(-binči)c ar icis "doesn't know a word of Georgian"
                                                                    (op. cit.: 752)

(24)   *c'ara-mara*[5]
     a scatter-brained; indolent; pointless
    a.   *c'ara-marad*
        adv all the time; again and again
    b.   *c'aramaraobs*
        iv talks rubbish/gibberish   (op. cit.: 1582)

(25)   *ic'ilo-bic'ilo*[6]
    1. chant in, and name of, game: children sit with legs out, one child counts legs on eeny-meeny miny mo principle; last child counted folds legs; game ends when all legs are folded; **2**. drawn-out boring procedure   (op. cit.: 774)

(26)   *ok'o-bok'o // ok'ro-bok'ro*
    1. n a uneven/rough (ground); **2**. messy, disordered   (op. cit.: 1054)

Once again, irrespective of the fact that both constituents of echo-word constructions pertaining to Type 3 are nonwords, resulting entities are meaningful. The question to be answered here is whether they can contribute to the encoding of ad hoc categories in Georgian. In the following paragraph I will try to find evidence in order to answer the question adequately for all the three types.

3.   Georgian echo-word formations as contributors to the encoding of ad hoc categories

In the present paragraph, I will try to find out whether various Georgian echo-pairs actually contribute the encoding of ad hoc categories or not. This can be achieved based on relevant data from dictionaries, corpora, and elicited illustrations.

    As it has already been noted, *m*-reduplication is a productive phenomenon in present-day colloquial Georgian; however, what is particularly significant in terms of the present frame of reference is whether it is a productive device for the encoding of ad hoc categorization in the language. The answers are lying in individual examples.

    Take, for instance, (3) *xili-mili* 'fruit and such.' If we assume *xili-mili* to refer to only 'various kinds of fruit; tutti-frutti,' then it is, of course, just a conventional category. Actually, *xili-mili* may also refer to various combinations of assorted fruits,

---

5.   The only author who refers to *c'ara-mara* as an echo-pair is Rudenko (1940: 269).

6.   Naturally enough, here I am concerned with meaning 2.: 'drawn-out boring procedure.' As for children's finger-counting rhymes, as well as spells and glossolalia, they too occur as echo-pairs; their numbers are considerable in Georgian (Kikvidze 2016) and Kartvelian languages, at large (Tuite 2015). However, they hardly have anything to do with ad hoc categorization; hence, they are not included in the list of illustrations.

berries, cucurbits (resp. melon and watermelon), traditional Georgian confection-
aries (such as *Churchkhela*, dried fruit *Chiri*, sweet dried roll *Tkbili kveri*, etc.)
served as dessert or snacks:

(3) *xili-mili*
 p'arask'ev-s sayamo-s vaxšam-i
 Friday.DAT evening.DAT dinner.NOM
 šabat-i dila-Ø sauzme-Ø
 Saturday.NOM morning.NOM breakfast.NOM
 šabat-i sayamo-Ø vaxšam-i
 Saturday.NOM evening.NOM dinner.NOM
 k'vira-Ø dila-Ø sauzme-Ø
 Sunday.NOM morning.NOM breakfast.NOM
 k'ide, **xili-mili** *cucnaoba, parask'ev-s ro ča-Ø-val-t,*
 Still more, fruit and the like, nibbling, Friday.DAT PREV.1.arrive.PL,
 sanam k'ap'it'alur-i vaxšam-i ikneba da šabat-s
 till full-course.NOM dinner.NOM be.FUT and Saturday.DAT
 sauzmesa da vaxšams šoris
 breakfast.DAT and dinner.DAT between
    (source: http://off-road.ge/forum/index.php?showtopic=19437&st=80)
 Friday evening – dinner
 Saturday morning – breakfast
 Saturday evening – dinner
 Sunday morning – breakfast
 Moreover, **snacks**, nibbling, when we arrive on Friday, before a full-course
 dinner, and on Saturday between breakfast and dinner

The following is a construction which has not been yet evidenced in any of the sources;
needless to say, my choice has fallen on it because it encodes an ad hoc category:

(27) *sašvi-mašvi* 'entry pass and the like'
 Base: *sašvi* 'entry pass'
 Reduplicant: *m+[s]ašvi*
   q'velanairi sašvi-mašvi gaukmebulia
   all-kinds-of pass.and-the-like annulled.be.PRES3
   'All kinds of passes are annulled.'
 Ad hoc category: 'all kinds of documents granting access to the building
   of Tbilisi City Council'
 Context: Due to a special occasion, the Tbilisi City Council decreed
   to deactivate all kinds of documents granting access to the
   building, and this is how one of the opposition politicians
   referred to them (source: http://www.newposts.
   ge/?l=G&id=153583-საკრებულო,%20ზურაბიშვილი)

The following two are elicited examples at various locations in the city of Tbilisi:

(29) *sark'eebi-mark'eebi* 'mirrors and such'
    Base: *sark'e-eb-i*
                              mirror-PL-NOM
                              'mirrors'
    Reduplicant formation: *m+[s]ark'eebi*
        roca   dagč'irdeba   sark'eebi-mark'eebi,   mašin
        when PREV.S2.need.O3 mirror.PL.NOM.and-the-like.PL.NOM, then
        ras   izam?
        what.DAT do.S2FT
        'What will you do, when you need mirrors and the like?'
    Ad hoc category: Furniture and equipment for the barber's shop
    Context: a conversation between a barber and a provider at the barber's shop in Paliashvili Street, Tbilisi

(30) *čalağaži-malağaži* 'loin and such'
    Base: *čalağaži* 'loin'
    Reduplicant: *m+[č]alağaži*
        unda gaq'ido   čalağaži-malağaži
        must PREV.S2.sell loin.NOM.and-the-like.NOM
        'You have to sell all cuts of beef.'
    Ad hoc category: Loin and other cuts of beef
    Context: a conversation between two butchers about the best-selling cuts of beef at the Dezerter Bazaar in Tbilisi

I must admit that I have been lucky to have had an opportunity of eliciting the above-cited illustrations as far as they usually appear exclusively online and in specific contextual situations. However, it should be borne in mind that not all occurring *m*-reduplicatives can be assumed to be able to contribute to ad hoc categorization.

As for the echo-pairs pertaining to Type 2, neither field work nor corpus searches yielded any results relevant to our study. The two corpora that I mined (Georgian National Corpus (GNC) and Georgian Dialect Corpus (GDC)) provided illustrations in which some of them occurred either in their conventional meanings or as children's counting games. However, it is their conventional meanings that prompt their potential in terms of ad hoc categorization; for instance,

(9) *ali-k'vali*
    bavšvi alik'vali mama aris
    child   spitting image father be3SG
    The child is just like his father.

(10) *ač'ia-bač'ia* '

The Database of Georgian Idioms and Proverbs provides the following translation for it:

Lit: To tell stories about a rabbit

Fig: Foolishness, folly, stupidity.   (DGIP 2017)

Hardly can an argument be found to support the stance that Georgian echo-pairs, pertaining to Type 2, are contact-induced. Irrespective of the fact that this kind of formation is no longer productive, related echo-pairs (though a closed set, as already noted) can serve as a potential resource for the encoding various ad hoc categories which may emerge owing to particular contexts (illustrations will be provided below). With respect to the type of the relationship between their exemplars, the inferential process, involving the Georgian Type 2 echo-pairs, will not lead to the enconding of *sets*, but rather of *classes* and *frames*.

If we examine corpus data associated with the items pertaining to Type 3, it will become obvious that their lexical meanings are apt to be applied as raw material for the encoding of ad hoc categories. Some of them (for instance, (17) *anč'i-manč'i*) seem to have been formed as such but were conventionalized later). I searched the Georgian National Corpus for (GNC) for 14 fossilized echo-word constructions (13–16). It is noteworthy that a few queries yielded no hits at all (for instance, (14) *ala-bala*, (15) *alan-talani*, (17) *anč'i-manč'i*, etc.). On the other hand, others (for instance, (20) *aur-zauri* – 717 hits, (21) *aq'al-maq'ali* – 193 hits) occurred quite extensively: all of them had conventional meanings in prototypical contexts.

Notably enough, (23) *inči-binči* 'the slightest' yielded 12 hits and all of them were associated with knowing something (not knowing anything, specifically, languages, diplomacy, mathematics, demography, etc.).

As for (25) *ic'ilo-bic'ilo*, of the 34 hits in a GNC query, it occurs only two times as children's counting game; in the rest 32 ones, it is a certain narrative frame for referring to various drawn-out boring procedures.

And finally and most interestingly,

(26) *ok'o-bok'o // ok'ro-bok'ro*

1. n a uneven/rough (ground); **2.** messy, disordered (Rayfield 2006: 1054)

The GNC query yielded 15 hits for *ok'ro-bok'ro* (its alternative version *ok'o-bok'o* was not attested). Of them, eleven were associated with the notions of road, way, location; one referred to one's messy steps (*ok'ro-bok'ro nabiǯit*), another to uneven written characters (*ok'ro-bok'ro asoebit*), and still another to one's messy Georgian (*ok'ro-bok'ro kartulit*); finally, there is one context in which *ok'ro-bok'ro* occurs as an attribute to *čogburti* 'tennis.'

*Serenasi, romelmac p'irvel or c'reši sak'maod **ok'ro-bok'ro** čogburt-i ačvena*
Serena.ɢᴇɴ which.ᴇʀɢ initial two round.in rather **disordered** tennis.ɴᴏᴍ showed

Of Serena [Williams] who, in the initial two rounds, demonstrated rather disordered tennis.

This is a good example to illustrate how a Type 3 echo construction can contribute to the encoding of an ad hoc category (disordered, uneven, unstable game).

It should be noted that similarly effective illustrations do not abound when echo-pairs pertaining to Type 3 are concerned. This can be due both to their fossilized character (hence, an unproductive set) and to the common difficulty of attesting and eliciting of echo-word constructions as 'encoders' of ad hoc categories.

4. Conclusion

A question that may normally arise when dealing with the phenomenon in point is whether the linguistic items under investigation satisfy the necessary conditions for the construal of ad hoc categories. Reduplication has been recognized as one of the strategies of ad hoc categorization across languages (Mauri 2017; Mauri & Sansò 2018b); echo reduplication, as a linguistic means for the encoding of ad hoc categories in Russian, has been discussed in (Benigni 2018). However, this kind of recognition does in no way exclude cross-linguistic variation; moreover, whenever reduplication phenomena and, specifically, echo reduplication, are concerned, another question should be asked: do they (reduplicative constructions) occur as part of a strategy for ad hoc categorization in a particular language?

In order to answer the question with respect to Georgian, at the initial stage, I provided a classification of Georgian echo-word constructions based on a set of structural and semantic features. Among the identified three types, 'Turkish-like *m*-reduplication' (Type 1) appeared to be most productive in terms of ad hoc categorization and least difficult to elicit (notwithstanding the fact that the very mechanism is a likely borrowing). Normally, all of the categories, based on Type 1 echo-reduplicatives, have been sets.

Reduplicatives, pertaining to Type 2 ('Initial C/cluster deletion in the pre-echo twin'), represent a non-productive pattern, that is, they are a closed set of entities. Structurally, they can potentially occur as linguistic items adequately fit for the encoding of ad hoc categories (namely, sets) owing to the fact that their base, a meaningful word, can act as an exemplar. Judging from the empirical data, more exactly, lack of the empirical data, we can speak of them only as a potential resource.

As we saw it with Type 2, as an unproductive linguistic phenomenon, pertaining linguistic items have very reduced opportunities for their participation in ad hoc categorization. As for Type 3 ('Fossilized echo-pairs'), their applicability seems to be much more limited owing to the fact that, besides being unproductive, both of their constituents are bound forms, that is, they are nonwords; hence, a base (a potential exemplar) and a reduplicant are not identifiable. This may be due to either their being borrowed echo-pairs and/or to the loss of the meaning of a base through time. Moreover, judging from corpus queries, some of them occur rather infrequently even in their conventional meanings, nothing to say about ad hoc categorization. However, I retrieved a couple of cases (for instance, (26)) when they contributed to the encoding of an ad hoc category. Occurrence of such instances tells us that this is possible (the constructed entity, yielded as a result of the inferential process, will be a narrative frame).

As long as the present chapter has been concerned with echo reduplication as a possible strategy of ad hoc categorization in Georgian, one of the Kartvelian (South Caucasian) languages, it can serve as a model for the investigation of similar phenomena in other Kartvelian and Caucasian languages, at large, as well as in co-territorial and neighboring idioms.

References

Barsalou, Lawrence W. 2011. Ad hoc categories. In *The Cambridge Encyclopedia of the Language Sciences*, Patrick C. Hogan (ed.), 86–87. Cambridge: CUP.

Benigni, Valentina. 2018. *Ad hoc* categorization in Russian and multifunctional general extenders. In *Linguistic Strategies for the Construction of Ad Hoc Categories: Synchronic and Diachronic Perspectives*, Caterina Mauri & Andrea Sansò (eds). Special issue of *Folia Linguistica* 52(s39–1): 97–123. https://doi.org/10.1515/flih-2018-0003

Bzdęga, Andrzej Z. 1965. *Reduplizierte Wortbildung im Deutschen*. Poznan: Poznanskie towarzystwo przyjaciól nauk.

Datukishvili, Ketevan. 1990. Redup'lik'acia kartulši (Reduplication in Georgian]). PhD dissertation, Georgian Academy of Sciences.

DGIP. 2017. The database of Georgian idioms and proverbs (with their English, Russian, German, French, Turkish and Arabic equivalents). <http://idioms.tsu.ge/?p=1994> (14 February 2020).

D'souza, Jean. 1991. Echos in a sociolinguistic area. *Language Sciences* 13(2): 289–299. https://doi.org/10.1016/0388-0001(91)90019-W

Emeneau, Murray B. 1938. An echo-word motif in Dravidian folk-tales. *Journal of the American Oriental Society* 58(4): 553–570. https://doi.org/10.2307/594370

Gersamia, Rusudan, Kikvidze, Zaal, Lomia, Maia, Mamiseishvili, Nona & Sagliani, Medea. 2016. *Ponosemant'ik'uri leksik'a kartvelur enebši* (Phonosemantic vocabulary in Kartvelian languages). Tbilisi: Inovacia.

Hurch, Bernhard & Mattes, Veronika. 2009. Typology of reduplication: The Graz database. In *Use of Databases in Cross-linguistic Studies*, Martin Everaert, Simon Musgrave & Alexis Dimitriadis (eds), 301–328. Berlin: Mouton de Gruyter.

Inkelas, Sharon & Zoll, Cheryl. 2005. *Reduplication: Doubling in Morphology*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511627712

Johanson, Lars. 2002. *Structural Factors in Turkic Language Contacts*. London: Curzon.

Kikvidze, Zaal. 2016. How to gloss and translate Georgian echo-paired words. In *General and Specialist Translation / Interpretation: Theory, Methods, Practice: International Conference Papers*, Artur Gudmanyan (ed.), 168–172. Kyiv: AgrarMedia Group.

Kikvidze, Zaal. 2018. Copying direction as a factor in the classification system of Kartvelian echo-word constructions. *Papers of BAS, Humanities and Social Sciences* 2: 135–143.

Kikvidze, Zaal, Gersamia, Rusudan & Lomia, Maia. 2018. Patterns of phonosemantic reduplication in Kartvelian (South Caucasian) languages. *Linguistics Beyond and Within* 4: 92–107. <http://lingbaw.com/2018/Zaal-Kikvidze,Rusudan-Gersamia,Maia-Lomia> (14 February 2020).

Konstantinidou, Magdalene. 2005. Degenerative lexikalische Reduplikation durch [m] im Anlaut: Eine eigenartige rhetorische Figur im Mundartgriechischen und Türkischen. In *Sprache und Multikulturalität. Festschrift für Professor Käthi Dorfmüller-Karpusa*, Eleni Butulussi, Evangelia Karagiannidou & Katerina Zachu (eds), 253–274. Thessaloniki: University Studio Press.

Mauri, Caterina. 2017. Building and interpreting ad hoc categories: A linguistic analysis. In *Formal Models in the Study of Language*, Joanna Blochowiak, Cristina Grisot, Stephanie Durrleman-Tame & Christopher Laenzlinger (eds), 297–326. Berlin: Springer. https://doi.org/10.1007/978-3-319-48832-5_16

Mauri, Caterina & Sansò, Andrea (eds). 2018a. *Linguistic Strategies for the Construction of Ad Hoc Categories: Synchronic and Diachronic Perspectives*. Special issue of *Folia Linguistica Historica* 52(s39–1).

Mauri, Caterina & Sansò, Andrea. 2018b. Linguistic strategies for ad hoc categorization: theoretical assessment and cross-linguistic variation. In *Linguistic Strategies for the Construction of Ad Hoc Categories: Synchronic and Diachronic Perspectives*, Caterina Mauri & Andrea Sansò (eds). Special issue of *Folia Linguistica* 52(s39–1): 1–35. https://doi.org/10.1515/flih-2018-0001

McCarthy, John J., Alderete, John, Beckman, Jill, Benua, Laura, Gnanadesikan, Amalia, & Urbanczyk, Suzanne. 1999. Reduplication with fixed segmentism. *Linguistic Inquiry* 30(3): 327–364. https://doi.org/10.1162/002438999554101

McCarthy, John J. & Prince, Alan. 1996. Prosodic morphology (1986) [Linguistics Department Faculty Publication Series 13]. <https://scholarworks.umass.edu/linguist_faculty_pubs/13> (14 February 2020).

McCarthy, John J. & Prince, Alan. 1994. The emergence of the unmarked; Optimality in prosodic morphology. *Proceedings of the North East Linguistics Society* 24: 333–379.

Moravcsik, Edith A. 1978. Reduplicative constructions. In *Universals of Human Language, Vol. 3: Word Structure*, Joseph H. Greenberg (ed.), 297–334. Stanford CA: Stanford University Press.

Neisser, Friedrich. 1953. *Studien zur Georgishen Wortbildung* [Abhandlungen für die Kunde des Morgenlandes 31(2)], hrsg. von Gerhard Deeters. Wiesbaden: Deutsche Morgenlandische Gesellschaft/Steiner.

Niepokuj, Mary. 1997. *The Development of Verbal Reduplication in Indo-European* [Journal of Indo-European Studies. Monograph Series 24]. Washington DC: Institute for the Study of Man.

Pott, Adolf F. 1862. *Doppelung (Reduplikation, Gemination) als eines der wichtigsten Bildungsmittel der Sprache, beleuchtet aus Sprachen aller Welttheile*. Lemgo/Detmold: Meyer.

Raimy, Eric. 2000. *The Phonology and Morphology of Reduplication*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110825831

Rayfield, Donald. 2006. *A Comprehensive Georgian-English Dictionary*. London: Garnett Press.

Rubino, Carl R. 2005. Reduplication. In *The World Atlas of Language Structures*, Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie (eds), 11–29. Oxford: OUP.

Rudenko, Boris T. 1940. *Grammatika gruzinskogo jazyka* (A grammar of the Georgian language). Moscow: Academy of Sciences of the USSR.

Sanikidze, Latavra. 1968. Puʒegaorkʼecebuli kʼmpʼozitʼebi kartulši (Stem-reduplicated compounds in Georgian). *Macne* 3: 144–172.

Sanikidze, Latavra. 1976. Gaorkʼeceba kartulši (Reduplication in Georgian). *Annual of Ibero-Caucasian Linguistics* 3: 35–86.

Southern, Mark R. V. 2005. *Contagious Couplings: Transmission of Expressives in Yiddish Echo Phrases*. Westport C): Praeger.

Steever, Sanford B. 2009. Tamil and the Dravidian languages. In *The World's Major Languages*, Bernard Comrie (ed.), 635–652. London: Routledge.

Stolz, Thomas. 2008. Total reduplication vs. echo-word formations in language contact situations. In *Language Contact and Contact Languages* [Hamburg Studies on Multilingualism 7], Peter Siemund & Noemi Kintana (eds), 107–132. Amsterdam: John Benjamins. https://doi.org/10.1075/hsm.7.07sto

Stolz, Thomas, Stroh, Cornelia & Urdze, Aina. 2011. *Total Reduplication: The Areal Linguistics of a Potential Universal*. Berlin: Akademie Verlag. https://doi.org/10.1524/9783050050973

Tiwary, Kapil M. 1968. The echo-word construction in Bhojpuri. *Anthropological Linguistics* 10(4): 32–38.

Topuria, Varlam. 1979. Kartvelur enata sitʼqʼvacʼarmoebidan, VI, kʼompozitʼi (From word-formation of Kartvelian languages, VI, Compound). In *Works*, Vol. 3, Varlam Topuria, 103–131. Tbilisi: Mecniereba.

Tschenkeli, Kita. 1960. *Georgisch-Deutsches Wörterbuch*. Zürich: Amirani-Verlag.

Tuite, Kevin. 2015. *The semiotics and poetics of South Caucasian nonsense vocabulary*. Ms. <http://www.mapageweb.umontreal.ca/tuitekj/publications/Tuite%20SEMIOTICS%20AND%20POETICS%20OF%20SOUTH%20CAUCASIAN%20NONSENSE%20VOCABULARY.pdf> (14 February 2020).

Urdze, Aina. 2018. Preface. In *Non-Prototypical Reduplication*, Aina Urdze (ed.), vii–xi. Berlin: Walter de Gruyter. https://doi.org/10.1515/9783110599329-011

French type-noun constructions based on *genre*

From the creation of ad hoc categories to ad hoc categorization

Wiltrud Mihatsch
University of Tübingen

Colloquial French *genre* is highly polysemous with a variety of functions such as comparison, approximation and the fully pragmaticalized functions as a mitigator, a quotative marker, an exemplifier, a marker of narrative progression, explanation and focussing. So far, the focus has been on the various discourse functions, however, several studies also point out possible paths of pragmaticalization. The aim of this paper is to analyse in greater detail the construction at the root of all these functions, i.e. "X of the same type as Y", a construction originally used for construing ad hoc categories. This rather learned construction became established in the 19th century for French and transformed into a backgrounded instruction for sloppy categorization and the other derived colloquial functions listed above.

Keywords: pragmaticalization, pragmatic marker, ad hoc categories, approximation, mitigation, colloquial French, 19th century

1. Introduction

Categories are a fundamental principle of knowledge organization, long recognized and today at the core of research in cognitive psychology. I will follow the definition proposed by Smith (1995) defining a category as a set of individuals grouped together on the basis of common properties, which in turn form a concept – a concept being defined as the mental representation associated with a category. A closer look at the psychological research shows that, implicitly or explicitly, categories tend to be identified with linguistic labels. These, in turn, have been shown to stabilize

categories in semantic memory,[1] make them manipulable – and are possibly also fundamental in the emergence of categories (for overviews see Gelman & Kalish 2006: 303; Waxman & Gelman 2010: 111). The distinction between well-established categories typically expressed by lexemes such as *animal* and ad hoc categories such as "things to sell at a garage sale", as investigated by Barsalou (1983) and the analysis of the respective linguistic expressions may shed light on the complex interaction between category types and the role of language.

The goal of this paper is to analyse in greater detail the particular construction at the root of the pragmatic marker *genre,* i.e. the construction X *du genre de* Y ('X of the same type as Y'), a construction originally used for construing ad hoc categories, but which is subsequently pragmaticalized as a marker indicating hedging or ad hoc categorization.

This source construction will be analysed in the context of other constructions with *genre*, where *genre* equally loses its status as a taxonomic noun, notably in the postdeterminer (as in *all kinds of* NP) and the attributive modifier construction (4.2.1 and 4.2.2) (as in *a strange kind of* NP), the subcategory construction leading to a hedge expressing ad hoc categorization (4.3.) (*a kind of* NP), and a construction creating ad hoc categories by establishing a link to preceding category members (*that kind of thing*) (5.1), which is closely related to general list extenders. All these constructions show the potential of taxonomic nouns to take over a variety of discourse functions leaving behind taxonomic hierarchies.

The first reduced constructions of the type *X genre Y* appear in the middle of the 19th century (Mihatsch 2010: 205). I have therefore created a subcorpus of FRANTEXT including the preceding years for a detailed qualitative and quantitative corpus search (all texts from 1840 to 1869, 25 604 619 words in 419 texts), altogether 2938 occurrences of the singular *genre.* In my counts I have excluded more specific readings of *genre* meaning '(biological) genus' or 'style' (not always easy to distinguish from the general categorial reading), as well as collocations such as *genre de vie* 'way of life', *genre humain* 'humankind' and others. This subcorpus will be the main empirical basis of this study, complemented by contemporary as well as older data.

---

1.   Semantic memory, i.e. the subtype of long-term memory which stores general information dissociated from an individual person and from particular autobiographic events as in the case of episodic memory. Both semantic and episodic memory are declarative, thus accessible to consciousness. The mental lexicon is stored in the semantic memory. The third type of long-term memory is procedural memory, which stores routines or procedures, not concepts, this is where grammar is stored, see (Ullmann 2004) on the role of the distinct memory types for lexicon and grammar.

2. Well-established categories and linguistic labels

Common or well-established categories (Barsalou 1983) such as *dog* or *animal* are stored in the semantic memory, i.e. the declarative memory containing general (shared) knowledge. Common, well-established categories (cf. Barsalou 1983) tend to be tied to lexical units and their signifiers ("labels" in the psychological literature) as in the case of *dog* or *animal*.

It is not clear in how far there are common concepts which are not labelled by lexical expressions, and if so, whether they may be expressed by grammatical morphemes. Berlin (1992: 190–194) observes a lack of lexical items meaning 'plant' and 'animal' in Tzeltal. However, he shows indirect evidence for the entrenchment of these concepts among Tzeltal speakers, such as distinct names for parts, results of categorization tasks and the existence of grammatical morphemes showing a conceptual distinction, i.e. a classifier *-tejk* for plants and *-kojt* for animals (Berlin 1992: 192ff.), while lexical labels for these concepts seem to be a recent feature in Tzeltal (Berlin 1972: 78).

Since grammatical items belong to the domain of procedural memory and give processing instructions rather than express accessible and manipulable concepts, it is possibly not adequate to speak of grammatical categories or concepts in the sense of the above definition.

According to Waxman and Gelman (2010), lexical items establish cognitive structures allowing mental processes which go beyond inductive reasoning and which support theorizing, albeit with word class particularities. Experimental studies on categories focus on nominal expressions, see Waxman and Gelman (2010) on the particular link between concepts for kinds and nouns. Nouns seem to be prototypically linked to categories, being particularly good labels (for an overview cf. Mihatsch 2009). This might be related to the reifying power of nouns and the fact that nouns are organized in taxonomic networks to a far larger extent than verbs or adjectives. Taxonomies, in turn, seem to be tied to well-established common categories according to Barsalou (1983: 212).[2]

All in all, nouns thus seem to be more than just labels for inductively established categories, since they offer a kind of abstract scaffolding for concepts perceived as particularly stable and allowing for generalization beyond purely perceptual properties (Waxman & Gelman 2010). In the following section, I will show that constructions creating ad hoc categories also draw on this scaffolding.

---

2. Interestingly, however, even nominal taxonomies are rather limited as to the number of hierarchical levels in everyday language, in particular above basic level, where meronymic principles are stronger and where we rather find collective nouns and object mass nouns such as *clothes* or *furniture* rather than hyperonymic count nouns such as the highly marked *garment* (Mihatsch 2016b).

3.   Ad hoc categories

Ad hoc categories, i.e. categories created on the fly in discourse and only stored temporarily such as "things to sell at a garage sale" (Barsalou 1983), can be created on the basis of very heterogeneous strategies on utterance level. For instance, they may arise in the shape of novel combinations of taxonomic adjectives (or nouns reinterpreted as adjectives) and nouns as in French *industrie instagram*[3] and other types of noun modification such as endocentric compounds, nouns modified by relative clauses, prepositional phrases and others, leading to a more specific category in a hyponymic relation to the (well-established) base noun. However, from a semantic point of view, the by now classic papers by Barsalou and others as well as the analyses of the LEADhoC project (notably Mauri 2017, and http://www.leadhoc.org/) restrict ad hoc categories to superordinate categories such as "things to sell at a garage sale" on a more general level like the well-established categories designated by *plant, garment* or *tool* rather than, say, *hammer, drill*, or *jumper*. According to this restriction, ad hoc categories create categories on the basis of several different kinds of exemplars (or rather concepts, categories or referents) belonging to different basic level categories in the sense of Rosch et al. (1976). This might be related to the more obvious categorizing function of the superordinate level, which groups together subkinds, as well as the more evident "negotiable" semantic content of hyponymic and meronymic superordinates (often oscillating between the two as in the case of Spanish *ropa* 'clothes, garment', see Mihatsch 2016b) as opposed to basic level items. Superordinate categories tend to offer more than one principle of classification, cf. *outfit* vs. *ensemble, laundry* vs. *clothes.* This also explains in part why nouns above the basic level are more varied and more instable than the rather stable basic level lexemes, which are largely determined by inherent permanent and typically holistic visual information such as shape. On superordinate level, we therefore find count superordinates such as *garment*, collectives such as *outfit* or *ensemble*, and the in-between category of object mass nouns such as *clothes* or *clothing*, which are more entrenched than collectives (Mihatsch 2016b).

Keeping in mind this intuitively convincing decision of delimiting ad hoc categories, in what follows I will stick to ad hoc superordinates in this sense, including sets/aggregates and frames. From a cognitive point of view, ad hoc categories (in the sense of Barsalou 1983) are processed in working memory or, if relevant beyond the utterance, in episodic memory.

---

3.   Cf. Knittel (2009) for the semantic class of taxonomic adjectives creating subkinds of the noun they modify, while English and German tend to create N+N-compounds for novel subkinds.

From a formal point of view, Mauri (2017) restricts ad hoc categories to complex linguistic expressions constructed on the spot as in *clothing to wear while house painting*, as opposed to "ready-made linguistic labels" such as simplex nouns, but also excluding more complex but smaller constructions such as compounds or certain adjective-noun combinations leading to new categories. Plausibly, in this case (for example, rather recent combinations such as *Paleo foods*) the combination of a rather general head noun with a simple lexical modifier contributing *differentiae specificae* shows a more stable, predictable semantics than superordinates (see Hampton 1997, for principles of concept combination) – although these phrases may also be created ad hoc and the assignment of individuals may be ad hoc too (see, for instance, the many recent online lists of *Paleo foods*).

According to Mauri (2017: 302), there are three relevant semantic types of ad hoc categories:

1. Sets, i.e. co-occurring elements such as the ingredients of a recipe: *For smoothies you need fresh fruit, vegetables, herbs and so on,* typically with conjunctively joined general list extenders formally closing an open list.
2. Classes of exemplars as in [quiet activities that may appeal to a 6-year-old boy]: *You can read a book, make a drawing or something* (Mauri 2017: 302), typically with a disjunctively added general list extender.
3. Frames, i.e. actions and related elements linked to a scheme, these will not be the focus of the present paper.

Ad hoc categories can come in different shapes; according to Mauri (2017: 318–319) the following formal types are wide-spread and can be observed in many languages:

1. Inflection as in the case of associative plurals
2. Derivation such as collective suffixes as in *Berlusconame* 'Berlusconi and his supporters or entourage'
3. And more complex syntactic constructions which may be partially lexically specified as is the case of lists with general list extenders or the non-exhaustive disjunction of two (usually bare) noun phrases (Ariel & Mauri 2018), and possibly also equivalent conjunctive coordinations creating sets.

Mauri (2017) points out that while inflectional and derivational expressions construe the superordinate starting out from one lexical expression, complex constructions such as lists may contain two and more lexical nouns. She further observes that inflection and derivation tend to lead to ad hoc sets, while more complex constructions seem to create more varied types of ad hoc categories, including classes and frames.

The semantic strategy can start out from a more general superordinate, such as *clothing* as in *clothing to wear while house painting*, while the added information

gives clues as to the property of the ad hoc superordinate. However, the strategies Mauri studies in detail are based on one or more exemplars (or subcategories/subordinate concepts) in a hyponymic or meronymic relation to the ad hoc (superordinate) category. The subordinates are modified by an associative plural, a collective or aggregate suffix or occur in lists with list extenders as in *You can read a book, make a drawing or something*. The clues as for the implicit properties of these ad hoc categories are less obvious than for expressions such as *clothing to wear while house painting*.

However, although the extension of an ad hoc category is often not clearly delimited and leads to the impression of extensional vagueness, the category as such is bound together by one or a few implicit[4] common properties, often a situational or temporary functional aspect:

> The identity of the category is guaranteed by the contextually relevant property P, whose identification allows to discriminate between potential members and elements that should be excluded. In other words, a sufficiently unambiguous identification of the property P allows to delimit the borders of the category, and ultimately to identify in a sufficiently unambiguous way the category itself.
>
> (Mauri 2017: 305)

This definition will be crucial for the distinction between ad hoc categories and ad hoc categorization diachronically arising from ad hoc categories as in the case of French *genre*.

4.   Taxonomies and taxonomic nouns

So far, research on ad hoc categories has focused on exemplar (or subcategory) as well as superordinate-based strategies. However, taxonomic nouns (or "type nouns") have long been used in the ad hoc constructions of categories, too, not only when speaking explicitly about established conceptual hierarchies (see Mihatsch to appear a and b for systematic overviews over taxonomic nouns and their derived uses in Germanic, Romance and Slavic languages). Despite the learned origins of most type nouns in Romance languages (but also other languages) (Mihatsch 2016b: 139), the creation and manipulation of categories itself seem to be cognitively hard-wired or at least acquired very early on. Children manage to group similar referents or concepts at a very early age and distinguish between kinds and individuals by an age of 12 months and acquire generic expressions very early on. According to Gelman and Brandone (2010), generic readings might indeed be the

---

4.   This distinguishes the above-mentioned formal types from the cases of conceptual combination as in (rather) novel adjective-noun or noun-noun combinations creating new kinds as in *diet industry* with an explicit defining property.

default interpretation of nouns and help children to map object labels. Linguistic labels in turn are, at least for adults,[5] cues for categories. Gelman and Brandone (2010) show that already children use a fast-mapping process on the basis of an unspecified kind structure serving as a placeholder in the course of concept acquisition and early on identify some basic ontological types such as animals, artefacts etc. Children further assume the extendibility of category labels to new instances and count on the possibility to fall back on expertise (Gelman & Brandone 2010: 225).

The general taxonomic scaffolding in particular for nouns plausibly facilitates the subsequent acquisition of very varied and rich category distinctions – and, in my opinion, also leads to a very smooth process of construing and communicating ad hoc-categories in children and adults which build on these general structures. The fact that we communicate ad hoc categories with great ease is certainly related to the very early mastery of the concept KIND.

However, no explicit type nouns are needed for these operations – they may remain implicit, as generic NPs and taxonomic or sort readings in many languages show, not only in Western literate societies, as in the case of the sort plural (*They produce three wines*, cf. Corbett 2000: 84–87). The ease of transition between taxonomic or sort readings to reference to individuals is plausibly the underlying mechanism that leads to an easily available reinterpretation and pragmaticalization of type nouns.

4.1   The implicit and explicit reference to subkinds: Taxonomic readings

Generic NPs with the definite article directly refer to categories as in *The dodo went extinct in 1681*. NPs may also directly refer to subkinds, in particular with mass nouns (*He likes a particular cheese, they sell about twenty wines*), but also with count nouns, although here the default reference to individuals has to be overridden. In the following passage I have underlined the explicit taxonomic NPs, the NPs with unmarked subkind readings are in bold face (all bold face in this paper is mine):

(1) In theory, the two federal agencies responsible for implementing the ESA–the USFWS and the National Oceanic and Atmospheric Administration–aren't supposed to favour <u>one type of animal</u> over another. Still, **larger animals** dominate the list. There are <u>234 foreign bird species</u> listed as threatened or endangered, and even **more foreign mammals**. By comparison, the act protects only **nine foreign amphibians, four insects, three plants**, and **one snail**.
<https://www.audubon.org/news/how-can-endangered-species-act-protect-wildlife-lives-outside-usnce> (20 February 2019)

---

5.   Cf. Waxman and Gelman (2010), who argue for an early ontogenetic link between labels and categorial status, while Sloutsky (2010) argues in favour of a later association between labels and categorial status.

Such uses look back to a long tradition, here is an equivalent 16th century example from French:

(2) **D'un oyseau de proyë qui voit la nuict**, nommé en Grec Phinis, et en Latin Ossifragus. (FRANTEXT, Belon, Pierre (1555): *L'histoire de la nature des oyseaux*)

'On **a bird of prey that can see at night**, called Phinis in Greek, and in Latin Ossifragus.'

Nouns introduce categories, not only with generic and sort readings particularly referring to categories, but also generally, while adjectives single out properties, verbs actions (cf. Gelman & Kalish 2006). NPs may be predicates, for instance in copula constructions, and then assign a category (as in *John is a painter*), or express arguments and refer (*John met a painter*). However, even when they refer to individuals as in *a painter* NPs categorize at the same time and show a default conflation of categorization and reference (roughly paraphraseable as 'John met a person, and this person is a painter'). This hybrid function of nouns becomes more obvious when taxonomic readings are combined with referential uses, i.e. when type and instance readings conflate as in *I bought three cheeses*. However, we may also use taxonomic nouns which render explicit subkind readings, but may be combined with referential uses. These cases will be discussed in the next section.

4.2 Preposed and postposed taxonomic nouns and the conflation of type and instance readings

Taxonomic nouns explicitly express subkind readings in the construction *a type of* NP. Such subkind uses look back to a venerable tradition, going back at least to Greek philosophy. This is also where the Aristotelian differentiation between genus and species originates, i.e. nouns designating categories with a specialization on different levels of generalization, which can be observed today in the particularly fine-grained terminology of taxonomic levels in biology.

In French, subkind constructions are attested from the 14th century onward for *sorte, espèce*, but also from the 12th century on for *genre* according to the TLFi (Trésor de la Langue Française informatisé). *Genre*, traditionally the superordinate level above species, is also established as a more general taxonomic noun, not only as a superordinate of species.

Taxonomic nouns tend to be learned nouns, but have now been well established even in colloquial language in a great variety of contexts, in particular when explicit reference to categories is required. They are then near synonyms of NPs with a sort reading. Just as sort readings of NPs without taxonomic nouns, complex NPs with explicit taxonomic nouns creating subkinds often show conflation of categorization

and reference as in *I bought three kinds of cheese*. Here, the argument of *buy* is not *kind*, but *cheese*. This seems to explain why *This kind of tree* appears to be synonymous with *A tree of this kind* (cf. OED (*Oxford English Dictionary*), s.v. *kind*), a very inconspicuous but probably very frequent case of "dual reference" (following Ward & Birner 1995: 732). Ward and Birner use the term dual reference in the case of reference to a new (indefinite) instance of a known (or definite) type or category as in *There was **the usual crowd** at the beach today* (Ward & Birner 1995: 732). If the taxonomic noun follows the other noun as in *A tree of this kind*, the duality is evident in the two determiners (*a* and *this*), while the indefinite article is implicit in referential uses of *This kind of tree*:

(3) (…) loblolly pine tree (parents **planted this kind of tree** in front of the house they lived in when I was born).   <https://www.sandiegoreader.com/photos/galleries/tattoo-you/26772/#> (12 June 2018)

In such uses, the taxonomic noun easily loses its head status, a strong piece of evidence is the frequent agreement with the second noun (also see Mihatsch 2016a: 142–146), not only in Italian as in (4):

(4) a. *Questo tipo di persone {è / sono}*
         'This type of persons {is /are}'
     b. *Le persone di questo tipo {*è / sono}*   (Zamparelli 2000: 103)
         'The persons of this type {*is / are}'

However, the restrictions on the interchangeability of the variants *an N of this kind* and *this kind of* NP in (5) are still unclear:

(5) a. *There is a student of that type in the class. She does better than most.*
     b. *There is that type of student in the class. ??She does better than most.*
                                                              (Zamparelli 2000: 90)

Apart from these restrictions, the two constructions *a kind of X* and *an X of this kind* often seem to be equivalent. For French *genre* Rosier also mentions the equivalence of both constructions, i.e. Fr. *ce genre de dossier ~ un dossier de ce genre* (Rosier 2002: 82). However, the equivalence does not seem to work with all taxonomic nouns in contemporary French, unlike in English:

(6) Fr. *cette espèce de dossier* / ?? *Un dossier de cette espèce.*[6]
     'this kind of file / ?? A file of this kind'

---

6.   However, this is somewhat difficult to test, since *espèce* de is now predominantly used as an approximator and in this interpretation impossible here as well. Both constructions are acceptable and common with English *kind* and *sort*, which are more general than *espèce*.

In fact, the two constructions are not symmetrical, but express different taxonomic relations. For *ce genre de dossier*, I propose a pseudo-partitive expressing subordination, while *un dossier de ce genre* rather stresses a superordination relation, placing *dossier* into a superordinate category. The definition of *genre* in logics and philosophy in the TLFi refers to the particular use for superordination based on isolated properties: "Idée générale ou classe d'êtres ou d'objets qui possèdent un ou plusieurs caractères communs", i.e. 'general idea or class of beings or objects which have one or several properties in common'. Here, the extraction of one or more isolated common properties seems to be rather related to the assignment of superordinate categories rather than subdivision, a use TLFi also assumes for everyday uses of *genre*: "L'accent est mis sur l'ensemble et sur l'appartenance d'un être ou d'un objet à l'ensemble". This explains why nouns such as *espèce* specialized in establishing subkind readings are today impossible in this construction, while *genre* is perfect, evident in Example (7):

(7)   Il y eut encore une scène de révolte et d'emportement qui fut tout ce qu'il est possible d'imaginer **en ce genre** de plus puéril et de plus charmant
                    (FRANTEXT, Sandeau (1848): *Mlle de La Seiglière*)
      'There was again a scene of revolt and anger that was the most childish and the most charming that can be imagined **of that kind**'

Taxonomic nouns not only express explicit categorical structures, they may render some constructions more expressive, underline lexical choices or help out in situations where speakers have problems with lexical access or need to refer to unknown concepts. Typical contexts will be presented in the following sections.

4.2.1   *Conflation of type and instance readings: Taxonomic nouns supporting adjectival modification*

A particular construction type showing a conflation of type and instance readings is the one with an NP specified by a taxonomic noun, which, in turn, is modified by an adjective[7] (also see Brems & Davidse 2010, for English; Mihatsch 2016a: 148–151, on Romance languages):

---

7.   See Denison (2002) and Brems and Davidse (2010), who further distinguish two subtypes, the attributive modifier use stressing rather unexpected or unusual selected features as in *This is but a scandalous sort of an Office*. (PPCEME, Farquhar, 1707) and the classifying semi-suffix uses such as in *David knew nothing about this mincing, half-and-half, milk-and-water Sort of religion*. (CLMETEV, Booth, 1880, both examples in Brems & Davidse 2010: 187). (13) is a case of attributive modification.

(8) (…) je compte bien, de mon côté, avoir l'honneur et l'agrément de **relations d'un genre plus intime**, madame (…)

> (FRANTEXT, Duranty, Louis (1860): *Le Malheur d'Henriette Gérard*)
> 'I intend, for my part, to have the honour and pleasure **of a relationship of a more intimate kind**, Madam'

The subcorpus of Frantext I selected for the analysis of the 19th century constructions with *genre* contains 75 occurrences (50 of which occur with abstract nouns) of this type with *genre*, most of which seem to be equivalent to direct adjectival modification of the non-taxonomic noun (*relation* in the example above) and can be replaced by an NP directly modified by an adjective, but showing a sort reading. However, the taxonomic noun adds some explicitness of categorization, which underlines individual-level modification and suggests the existence of stable subtypes.

4.2.2 *Conflation of type and instance readings: Postdeterminer uses*

A further type of use typically leading to a conflated type and instance reading is the use of *genre* with identifying[8] and quantifying expressions (negative quantifiers, universal quantifiers, indefinite quantifiers), which can be subsumed under the label postdetermination (Brems & Davidse 2010; De Smedt, Brems & Davidse 2007, for an in-depth corpus analysis of *kind, sort* and *type,* and Mihatsch 2016a, for an analysis of the Romance equivalents). The taxonomic noun reinforces quantification or determination (see Mihatsch 2016a: 142–144) and emphasizes or suggests (or pretends) a conscious and careful categorial identification undertaken by the speaker, both with the preposed and the postposed variant as in (9) and (10):

(9) **aucun genre** de chute (FRANTEXT, Sand, George (1843): *Consuelo*, t. 19)
'**no kind** of fall'

(10) des privations de **tout genre** (FRANTEXT, Collectif (1869):
> *Recueils de textes d'histoire, 4. L'Époque contemporaine (1789–1870)*)
> 'hardships of **all sorts**'

In this construction, the taxonomic noun becomes an enclitic of the preceding determiner, with which it then forms a postdeterminer (Denison 2002). Again, the taxonomic reading can become backgrounded as a consequence of the frequent conflation of type and instance readings. This transition is particularly easy since the type readings with quantifiers and identifying uses as well as the adjectival ones can be reinterpreted as an equivalent instance reading via implicature. For example, the reference to all kinds as in (10) expressively implicates in dual

---

8. For instance, for *même, autre, tel, quel* etc., I have not included the lexical adjective *nouveau* here, although it shares the identifying function with some expressions of this type.

reference contexts the reference to all individuals (as in 'they suffered hardships of all sorts'). In my FRANTEXT subcorpus, in all the postdeterminer uses, abstract nouns prevail (266 occurrences), followed by artefact nouns (24) and human nouns (15). Constructions with postposed *genre* as in (10) are far more frequent than the preposed variant as in (9), for instance, 200 of the postdeterminer uses with abstract nouns correspond to the postposed variant.

These constructions do not create ad hoc superordinates. However, uses with demonstratives and definite articles as well as similative deictics such as *un tel* can point to preceding subkinds or individuals and thus create an implicit ad hoc superordinate as in *Le matin, j'ai été à <u>la chasse aux scorpions</u> avec un gentleman adonné à **ce genre de sport**.* 'I went <u>scorpion hunting</u> with a gentleman who was into **this kind of sport**' as in Example (18) discussed in 5.1.

Not only *genre* is used in these constructions, but also other taxonomic nouns, although less frequently (Mihatsch 2016a: 151–152). These uses are also common in the other Romance languages, but also English (Brems & Davidse 2010) and German.

4.2.3   *From subkind specification to approximative ad hoc categorization*
Although the construction I discuss in this section also shows uses conflating type and instance readings, the outcome is different. Here, the taxonomic noun and its reference to a category does not just fade away, but is reinterpreted as an approximator. Taxonomic nouns have long been used in contexts where speakers or writers lack more specific expressions and where they only manage to give a (rough) assignment to a superordinate. In such cases, the differentiae specificae remain implicit as in Example (11):

(11)   La ville est ouverte, et presque dans l'impossibilité d'avoir une enceinte de murailles. Elle offre une surface trop immense. Il faudroit **un genre de fortifications particulier**.
            (FRANTEXT, Mercier, Louis-Sébastien (1782): *Tableau de Paris*, t. 4)
    'The town is open, and almost impossible to have a wall enclosure. It offers too much surface area. There would have to be **a particular kind of fortification**.'

The taxonomic noun refers to a hierarchical organization, which, however, is not fleshed out. It points to a general scaffolding, thereby suggesting that there are well differentiated and possibly established subtypes of fortifications in the case of Example (11).

Remarkably, in many (perhaps all?) languages the unspecified and unmodified subkind construction (X *is*) *a kind of* Y, where the differentiae specificae remain implicit, are easily reinterpreted as imprecision or approximation markers. In French, three taxonomic nouns have been clearly reinterpreted in this construction, *genre*

de, espèce de and *sorte de* (Fleischman 1998; Rouget 1997, DHLF). (*Un*) *type de* and *forme de* appear occasionally in contexts of approximation, *classe* does not (Mihatsch 2010: 146).

Another means of signalling explicit subkind readings is the use of a similative deictic like *such*, which is often equivalent to the construction with a taxonomic noun. Unlike these constructions, however, *such an* N does not necessarily suggest a stable subcategory, but a similarity relation established on the fly.

Reinterpretation plausibly starts out from the very useful innovative subcategorization of peripheral subcategories or referents (X), which are classified as albeit marginal subcategories of a superordinate Y. If we are at the margin of the category Y, the relation between X and Y then seems closer to cohyponymy. Particularly in everyday language, same-level similarities with a prototype closely associated with the superordinate may be more salient than more general properties shared by cohyponyms and defining the superordinate.

In (12), *poree*[9] is used as a superordinate, comprising a subkind *poree commune*. Remarkably, we have both an assignment to a superordinate and a comparison with a cohyponym (*poree commune*):

(12) **Une espece de poree** que l'en dit espinars et ont plus longues feuilles, plus gresles et plus vers, que poree commune.
              (FRANTEXT, Anonymous (1394): *Le Menagier de Paris*)
      '**A kind of leek** which is called spinach and which has longer, finer and greener leaves than common leek.'

Ambiguities arise in particular with abstract nouns as in Example (13):

(13) Je crois les adoptions bien préférables aux anoblissemens faits par l'état. Elles feroient revivre des familles illustres, dont les descendans languissent aujourd'hui dans la plus étroite pauvreté. Elles rendroient la noblesse chère au peuple, et le peuple cher à la noblesse. Il faudroit que le privilége de les conférer devînt **un genre de récompense** pour les nobles eux-mêmes.   (FRANTEXT, Bernardin de Saint-Pierre, Henri (1784): *Études de la nature,* t. (3)
      'I believe adoptions are much better than the ennobling done by the state. They would bring back to life illustrious families, whose descendants are now languishing in the deepest poverty. They would make the nobility dear to the people, and the people dear to the nobility. The privilege of conferring them should become **a kind of reward** for the nobles themselves.'

---

9. *Poree* 'leek' also had a general meaning 'vegetable' according to Godefroy (1881–1902, s.v. *poree*), possibly, here we have an intermediate level between leek and vegetable.

From marginal or ad hoc subclassification, we thus arrive at approximation (or hedging) transcending strict taxonomic structures. Then, the reinterpreted taxonomic noun just establishes a similarity relation, the target concept and the noun following the taxonomic noun do not belong to one common superordinate. Such hedging uses are attested at least from the 16th century for *espèce*, from the 17th century for *sorte* (TLFi and DHLF)). For *genre* TLFi mentions that it is often used "avec une valeur d'approximation",[10] although less established and probably younger, since the everyday subkind reading might be partially blocked due to the learned genus reading.

Clear recent examples of approximation with *genre* appear even in specialized texts as in (14), in the example below in the context of a comparison between a 19th century project to a law introduced in the first half of the 20th century, thus establishing a vague relation of similarity:

(14)　La première [société industrielle] cité date de 1852 ; et cette société, continuant à être précurseur avait organisé **un genre de « loi Loucheur »**[11]　(FRANTEXT, Becquet, Yvonne (1939): *L'Organisation des loisirs des travailleurs*)
'The first [industrial company] cited dates back to 1852; and this company, continuing to be a forerunner, had organized **a kind of 'Loucheur law''**

In colloquial language, the functions go beyond semantic approximation and may signal the unexpected use of metaphors as in (15):

(15)　Ça moutonnait le long des boutiques. Le tramway, **un genre de girafe obèse**, il dépassait les bicoques, il laminait la cohue, il godaillait dans les vitres…
(FRANTEXT, Céline, Louis-Ferdinand (1936): *Mort à crédit*)
'people pushed along the shops. The tramway, **a kind of obese giraffe**, passed the shacks, it toppled the crowd, it dug into the shop windows'

---

10. In Mihatsch (2010: 161), I give an example for an early hedging use, which I would now rather classify as a taxonomic use, with *comme* taking over the comparative function used as a hedge (interestingly, the Latin text does not contain a hedging expression):

"Genus est mortis male vivere. Ovidius, De Tristibus. Veult dire Ovide que vivre mauvaisement n'est pas droite vie, mais si comme genre de mort"　(FRANTEXT, Pisan, Christine de (1412): *Livre de la Paix*)
'Living badly is a type of death. Ovidius. De Tristibus. Ovide means that living badly is not a righteous life, but like a type of death'.

11. The "loi Loucheur" was voted in 1928 and was intended to develop housing schemes for the working class.

The hedge or approximator *genre* (as well as the other taxonomic nouns reinterpreted as hedges) operate on the semantic level, since they are truth-conditional. They express ad hoc categorization by flexibilizing category boundaries of a lexical expression in the case of referents which do not fit neatly into an established category. They do not, however, create ad hoc categories, since there is no (implicit) property establishing an albeit ephemeral category. In addition to their semantic function, they tend to take over further, more pragmatic functions such as the expression of metalinguistic distance or word-finding problems or mitigation on the illocutionary level (Caffi 2007), although these uses are less established and still more restricted for these French hedges than the more pragmaticalized English *sort of* and *kind of* (Aijmer 2002: 175–209, for a synthesis see Mihatsch 2007, 2010: 155–160, 174–178).

5.   Constructions with *genre* building ad hoc categories

Conflation of type and instance readings also play a central role in the constructions leading to ad hoc superordinates, which get established in the 19th century. These will be described in greater detail in the following sections. Number variation will not be addressed in this paper since in the domain of ad hoc categorization with *genre*, the singular prevails (for some language-particular trends in Romance languages see Mihatsch (2016b)).

5.1   Ce genre de choses *and related anaphoric constructions*

Taxonomic nouns may refer to well-established categories and taxonomies. Especially in everyday language, they also create ad hoc superordinate categories, which are held together by an implicit common property inferred from world knowledge or the context. In this section I will present a particular phoric construction creating ad hoc superordinate categories by deictically pointing to a usually preceding non-exhaustive list of subkinds or individuals or even just one subkind.

As shown above, the constructions (with a demonstrative) *this/the kind of* Y and *a* Y *of this/the kind* are not equivalent at first sight. While the first construction typically establishes a subcategory, the second construction expresses an assignment to a superordinate category, although the overall reference of both constructions is the same, namely a specification of Y pointing to further information in the context or co-text. Both variants are commonly used to establish ad hoc superordinates by pointing to subkinds or exemplars usually preceding in the text or discourse. The subcategories or individuals are underlined in the following examples. In (16)

fileuses 'spinners' and *ménagères* 'housewives' are chosen to characterize a more general implicit superordinate category, roughly "rather dull, but solid weddable women", interestingly with an additional general explicit characterization *des êtres bons, propres et sages* 'good, tidy and decent beings':

(16)   Mettez ces femmes-là derrière un rouet et appelez-les nânon, et vous aurez leur encadrure et une harmonie. <u>Ce sont des fileuses, des ménagères, des êtres bons, propres et sages</u>, **le genre de femmes** qui convient peut-être le mieux à l'homme, pour la consuetudinem vitae, comme dit le droit romain…
          (FRANTEXT, D'Aureville, Barbey (1858): *Memorandum* (*Quatrième*))
          'Put these women behind a spinning wheel and call them donkey, and you will have their framing and harmony. <u>They are spinners, housewives, good, clean and decent beings</u>, **the kind of women** who may best suit the man, for the consuetudinem vitae, as Roman law says'

In (17) one particular and rather unusual type of hunting is given, which serves to create a superordinate ad hoc category "eccentric /exotic and dangerous type of sport":

(17)   J'ai, l'autre jour (en allant à Utique), couché dans un douar de Bédouins, entre deux murs faits en bouse de vache, au milieu des chiens et de la volaille ; j'ai entendu toute la nuit les chacals hurler. Le matin, j'ai été à <u>la chasse aux scorpions</u> avec un gentleman adonné à **ce genre de sport**. J'ai tué à coups de fouet un serpent (long d'un mètre environ) qui s'enroulait aux jambes de mon cheval.
          (FRANTEXT, Flaubert, Gustave (1860): *Correspondance (1858–1860)*)
          'The other day (on my way to Utica), I lay in a Bedouin douar, between two walls made of cow dung, among the dogs and the poultry; I heard the jackals screaming all night. In the morning, I went <u>scorpion hunting</u> with a gentleman who was into **this kind of sport**. I whipped dead a snake (about a meter long) that was wrapped around my horse's legs.'

In (18), the subkinds have to be construed on the basis of more loosely related aspects such as typical behavior, here a long stretch of preceding discourse, characterizing a particular category of woman:

(18)   <u>Ce caquetage fut sifflé d'une voix si doucement moqueuse, si mignonne, avec des mouvements de tête si coquets</u>, que d'Arthez, à qui **ce genre de femme** était totalement inconnu, restait exactement comme la perdrix charmée par le chien de chasse.   (FRANTEXT, Balzac, Honoré de (1844): *Les Secrets de la princesse de Cadignan*)
          '<u>This chatter was whistled in a voice so softly mocking, so cute, with such pretty head movements</u>, that Arthez, to whom **this kind of woman** was totally unknown, remained exactly like the partridge charmed by the hunting dog.'

These uses are still frequent today. Example (19) shows both positional variants pointing to preceding information on an individual event illustrating the ad hoc superordinate, i.e. the concert given by Noir Désir for *ce genre d'évènement* and *concert de ce genre*, plausibly creating an ad hoc category referring to a concert by an internationally renowned and famous band:

(19) A la fin de la soirée, les Yéménites interviewés ont tous insisté sur le succès [sic] <u>du concert</u> et ont exprimé leur joie de découvrir pour la première fois le groupe français. Certains ont même évoqué l'importance de **ce genre d'évènement** pour l'ouverture sur les autres cultures du monde. Un Français qui travaille au Yémen et qui a fait trois heures de route pour assister au concert de Noir Désir, était encore sous l'émotion quand il a déclaré : « J'ai été emballé ! Un **concert de ce genre**, c'est unique au Yémen ! ».   <https://musique.rfi.fr/musique/20020411-noir-desir-yemen> (22 February 2019)

'At the end of the evening, the Yemenis interviewed all insisted on the success of <u>the concert</u> and expressed their joy at discovering the French band for the first time. Some even mentioned the importance of **this type of event** for opening up to other cultures in the world. A Frenchman who works in Yemen and who drove three hours to attend the Noir Désir concert was still moved when he said: "I was thrilled! **A concert of this kind** is unique in Yemen!" '

Remarkably, these uses, possible with both variants, correspond to an interpretation equivalent to 'such a', although suggesting a more salient and rather stable ad hoc category.

Both variants also typically appear as general list extenders with general nouns –not in the focus of this paper, however–, but based on the same mechanism, i.e. pointing to a non-exhaustive list of subkinds, or just one subkind, or aspects vaguely suggesting subkinds:

(20) Attends encore un peu, je te prie. Veux-tu prendre <u>un café-crème</u> **ou quelque chose de ce genre?**   (FRANTEXT, Duhamel (1938): *Cécile*)

'Please wait a little longer. Would you like to have <u>a coffee</u> **or something of that kind**?'

The properties defining the superordinate tend to go beyond immediately perceptible properties and rather refer to character traits when speaking of people, or frames or contexts as in the above example, here probably referring to typical, standard items of a quick order at a café.

Both formal variants, the one with *genre* preceding and the one following another NP, can be found in the 16th century, although the early uses refer to well-established kinds, a particular kind of blackbirds in (21) and pear trees in (22):

(21)   (…) Aristote ha parlé du Merle noir et du blanc, voila comme il dit. Il y ha encor **un autre Merle de ce genre**, peu moindre que le Merle noir, et qui seroit semblable au noir (…).

(FRANTEXT, Belon, Pierre (1555): *L'histoire de la nature des oyseaux*)

'Aristotle has told about the black and the white blackbird, that's how he calls it. There is **another blackbird of this kind**, a bit smaller than the black blackbird, and resembling the black blackbird'

(22)   L'espace entre les poyriers doit estre de trente piedz. **Ce genre d'arbre** pour le bien proffiter, doit souvent avoir l'eau, et estre besché à l'entour (…)

(**FRANTEXT,** Darces, Jean (1554): *Les treize livres des choses rustiques de Palladius Rutilius Taurus Æmilianus*)

'The space between the pear trees must be thirty feet. If you want the take best advantage of **this kind of tree** you have to water it often and dig up the soil around it'

Figure 1 shows the absolute number of occurrences in my 19th century subcorpus (this analysis does not require normalized numbers, since I only look at one subcorpus, see (Section 1). I have analysed the type of category established by the construction with the singular *genre*[12] and have analysed three variants separately, distinguishing between the dominant use with a demonstrative, other (mostly definite) determiners which can take over the equivalent function, and the very infrequent uses with an indefinite article:

As Figure 1 shows, both variants (*genre* preceding and following an NP) overwhelmingly appear with demonstratives. These uses also comprise functionally equivalent variants with other prepositions (notably *dans*, but also *en*, by using the spatial preposition the NP is explicitly sorted into a superordinate category) and they appear mostly characterizing abstract nouns. Abstract nouns are difficult to pin down as to their taxonomic position and tend to lack well-established inclusion relations. Thus, if categorization is required, this tends to be done via ad hoc categories. Remarkably, both variants are also employed to categorize human nouns, here often personality or character types, a prominent use in authors such as Balzac and Flaubert who aim at a classification of human types, but also artefact nouns, less frequently but considerably more frequently than say natural kinds, which are more strongly integrated into established taxonomies.

---

12. The singular leads to the pragmaticalized functions and is therefore focused here, although the plural also appears in the variant *ces genres de choses* and is worth studying. The plural does not appear in the construction NP *de ces genres*, since here, an individual or subkind is assigned to one superordinate. This points to the above-mentioned conceptual difference between the variants with the preposed and the postposed taxonomic noun (cf. 4.2).
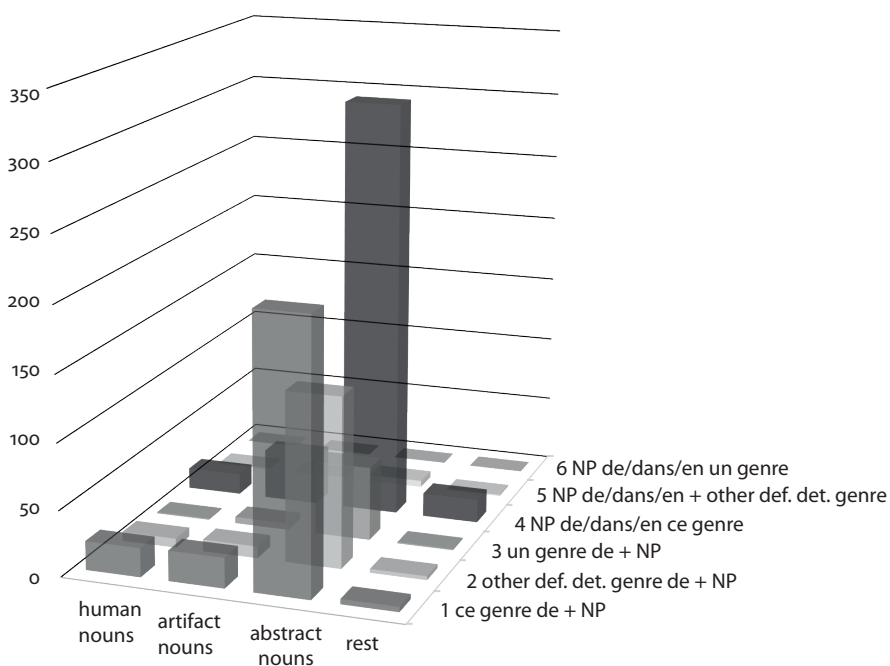
Figure 1. FRANTEXT 1840–1869, *ce/le/un genre de NP* and *NP de ce/du genre*

A rough comparative count of the forms in Figure 2 (only the ones with regular agreement and not attempting at a disambiguation of differing senses, which, however, are marginal for these nouns) gives the following results for *espèce, sorte, type, classe* and *genre* (*forme* has quite salient differing senses, such as 'shape'):
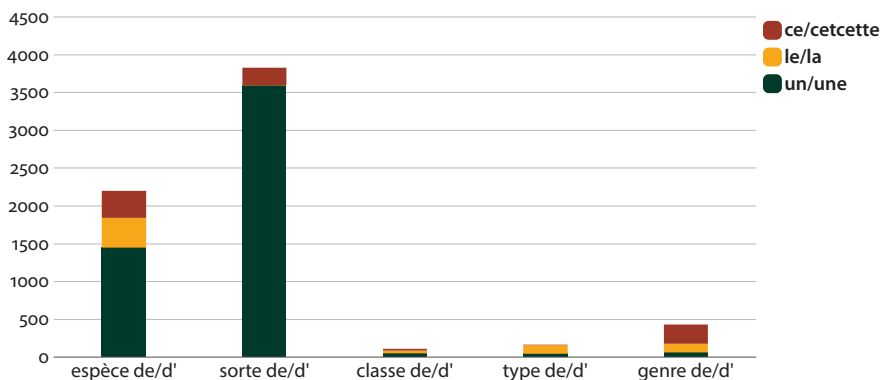
Figure 2. FRANTEXT 1840–1869, *a/these/the* TAXONOMIC NOUN *of* NP

Espèce and *sorte* appear predominantly in a subkind use with the indefinite article, with a high frequency compared to the other taxonomic nouns and, proportionally, far less frequently with the demonstrative than *genre*, although both the preposed and the postposed phoric uses are attested in the subcorpus as in Example (23):

(23)   Quant à l'autre histoire du capitaine Aguirre, je n'ai rien à vous dire, sinon que **les accidents de cette espèce** sont communs à Lima, et que je ne puis les empêcher.   (FRANTEXT, Mérimée, Prosper (1857):
Théâtre de Clara Gazul : Le Carrosse du Saint-Sacrement)
'As for the other story of Captain Aguirre, I have nothing to tell you, except that **the accidents of this sort** are common in Lima, and that I cannot prevent them.'

Type is altogether far less frequent, but shows the same preference for the demonstrative as *genre*. This might point to a skewed distribution, *espèce* and *sorte* preferring subkind readings, *genre* and *type* superordination and plausibly creating ad hoc superordinates. The selection of taxonomic nouns in this construction seems to be language-particular and requires a certain degree of conventionalization, since in such phoric contexts in contemporary Italian, Portuguese and Spanish *tipo* prevails, in French *genre* (Mihatsch 2016b: 152). Such uses are also common in German and English.

Possibly, due to entrenchment the preposed variant (*ce genre de NP*) is, after all, semantically symmetrical to the postposed variant and expresses rather the assignment of a superordinate than a subkind differentiation. However, further studies will be needed to investigate this point.

5.2   'X of the same type as Y' creating ad hoc categories

Just like the anaphoric constructions *ce genre de* N/ *un* N *de ce genre* discussed above in 5.1 the construction X *of the same type as* Y (NP *du/dans le genre de* NP) also productively creates ad hoc categories. The construction can be paraphrased as X *is of the (same) category or kind as* Y, as in 5.1 based on the assignment of a subkind or individual to an implicit superordinate.

Before analysing this construction in more detail, it is important to note that it has to be distinguished from superficially identical constructions which differ structurally. These express appositions where Y designates the target category, thus naming the superordinate category explicitly. The apposition can be linked with the preceding NP via a preposition as in (24) or without a preposition as in (25):

(24)   Estant donc cest oyseau Ossifragus **du genre des Aigles**, il vit de chair, et ha
       l'ongle crochu (…)

         (FRANTEXT, Belon, Pierre (1555): *L'histoire de la nature des oyseaux*)
       'Since this bird Ossifragus is **of the genus of the eagles**, it lives on meat and
       has hooked claws'

(25)   (…) non seulement que cet animal était de la classe des mammifères (…), mais
       encore qu'il appartenait à l'ordre des carnassiers ou à celui des ruminants, au
       genre chat ou au genre cerf.   (FRANTEXT, Cournot, Antoine (1851):
                                 Essai sur les fondements de nos connaissances et sur
                                         les caractères de la critique philosophique)
       'not only that this animal was of the class of mammals (…), but also belonged
       to the order of predators or ruminants, to **the genus cat or to the genus stag**'

In the selected subcorpus FRANTEXT 1840–1869, we find 68 such appositive uses
of *genre*, 48 of which appear without the preposition. 31 are NPs designating hu-
mans, 11 artefacts, only 9 abstract nouns and 17 belong to animates, institutions
and other categories taken together.

    The construction creating ad hoc superordinates differs from the appositional
construction in the status of Y with regard to *genre*. Here, the relation between
genre and Y is a relational genitive expressing possession or relation, and can be
paraphrased as X *of the same category as* Y. Y can also be a deictic as found in
the earliest attestations of this construction type in a systematic corpus search in
FRANTEXT, as in *des infortunes **du genre de la mienne*** 'misfortunes **of the kind of
mine**' (FRANTEXT, Crébillon, Claude Prosper Jolyot de (1751): *Ah quel conte*!).
First uses with NPs following *genre* appear in the 18th century as in (26). The con-
cept or referent to be classified, here *un chemin* is assigned to an implicit ad hoc
category illustrated by a (typical) category member, *celui de Kander-Steg* 'the one
of Kander-Steg'. The NP *un chemin* 'a path' specified by the PP *du genre de celui de
Kander-Steg* is thus first assigned to an implicit superordinate in the text, helping
the reader to construct the category on his or her own, only then the existing su-
perordinate term (*galleries*) is given and explained to a readership who is probably
not familiar with Alpine infrastructure:

(26)   Du haut de l'escarpement qui offre cette vue, on descend dans le précipice **par
       un chemin du genre de celui de Kander-Steg**, et que l'on a fort bien nommé
       les galleries; c'est un long zig zag taillé dans le roc perpendiculaire.
                (FRANTEXT, Ramond de Carbonnières, Louis (1781): *Lettres de M.
         William Coxe à M. W. Melmoth sur l'état politique, civil et naturel de la Suisse*)
       'From the top of the escarpment which offers this view, one goes down into the
       precipice **by a path of the kind of that of Kander-Steg**, and that one has very
       well named the galleries: it is a long zigzag cut into the perpendicular rock'

This construction is now only possible with taxonomic nouns allowing superordination as in the case of *genre.* Unlike the appositional construction, this construction originally requires a preposition after *genre*, i.e. *de, dans* and *en.* More than one subkind or referent Y illustrating the implicit superordinate can be given as is the case in Example (27):

(27)   Je viens de faire en quatre jours une pièce champêtre en 3 actes **dans le genre du** Champi, Mare au diable, etc., un peu plus dramatique cependant.
                    (FRANTEXT, Sand, George (1850): *Correspondance*)
'In four days, I have just made a pastoral piece in 3 acts **of the kind of** the Champi, Mare au diable, etc., a little more dramatic though.'

The implicit superordinate is in most cases an ad hoc category, unlike the explanatory apposition, where the second NP corresponds to a defined category label as in (24) and (25). The corpus analysis shows that there are two subpatterns, one where *genre de* is followed by an NP, another one, where it is followed by a deictic specified by a relative clause (*du genre de celui/celle qui…* 'of the kind of that which…'). As Figure 3 shows, again, the uses introducing abstract nouns prevail, followed by human nouns, including uses with proper names following *genre,* then artefact nouns, all domains typically allowing multiple ways of classification and showing a weakly established taxonomic differentiation:
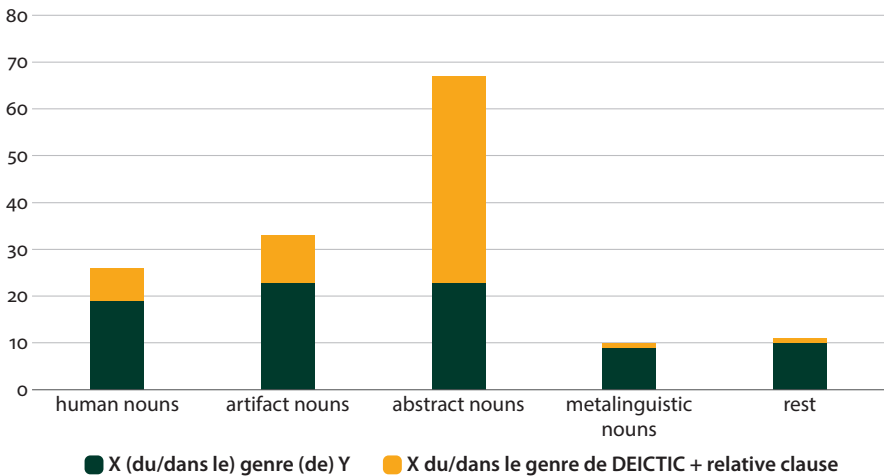
Figure 3. FRANTEXT 1840–1869, NP *of the same type as* NP
(anaphoric ad hoc superordinates)

Both variants are far less frequent than the phoric construction discussed in Section 5.1, particularly in the domain of abstract nouns:
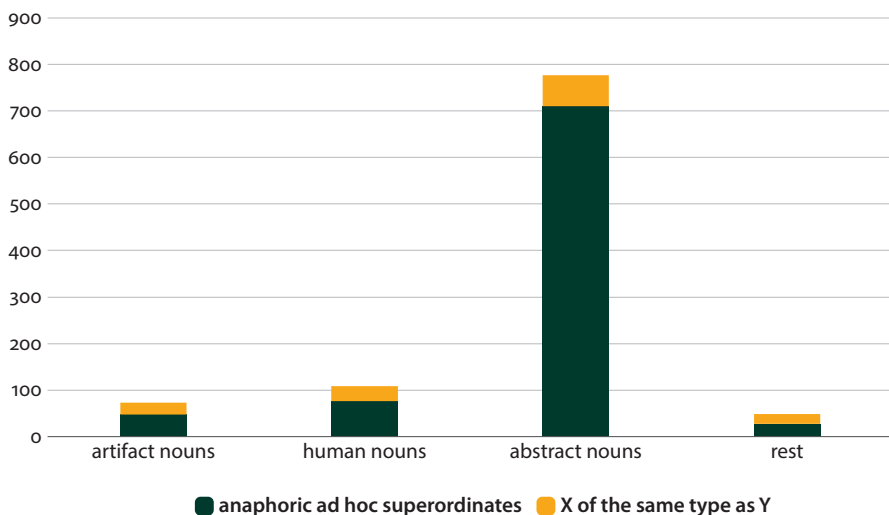
Figure 4. X *dans le genre de Y*, Anaphoric ad hoc superordinates
vs. NP *of the same type as* NP

Two particular uses in my subcorpus need to be commented on. In several cases, *genre* is followed by human proper names, often in appositional uses without prepositions, with the proper name, usually a famous figure, clearly established as a label for a whole category, such as in Example (28):

(28)   Figure-toi, Caroline, qu'il s'agit d'une petite fleur cueillie au coin d'un bois dans une promenade sentimentale, et qu'un monsieur **du genre Werther** avait juré de garder, qu'il fait encadrer, et qu'on lui redemande onze ans après

(FRANTEXT, Balzac, Honoré de (1850):
Petites misères de la vie conjugale, vol. (2)

'Imagine, Caroline, that it is a small flower picked from the corner of a wood in a sentimental walk, and that a gentleman **of the type Werther** had sworn to keep it, that he has it framed, and that we ask him to give it back eleven years later'[13]

Other cases, especially when linked by a preposition, clearly correspond to the construction containing a relational genitive, plausibly in the cases where the person is not notorious, for instance when referring to a figure in the same novel in Example 29):

---

13. I checked all translations with DeepL <https://www.deepl.com>, which generally gives very accurate results, but remarkably in these examples *du genre (de)* was translated by *like*, which is not perfectly equivalent to the 19th century use, although referentially identical.

(29) Les ménagères **du genre d'Agathe** ont un bon sens qui leur fait deviner ces sortes de tromperies politiques.

        (FRANTEXT, BALZAC Honoré de – *La Rabouilleuse* (1843))

'Housewives **of the sort of Agatha** have a good sense that makes them guess these kinds of political deceits'

Still today, proper names are frequent in the reduced construction (Chauveau-Thoumelin 2016), who observes a percentage of 31.5% among the different constructions with *genre*. In my 19th century subcorpus, of the 19 cases referring to humans, 16 are based on a proper name as in (29).

Another remarkable use (10 cases in my subcorpus) is the one where a whole utterance, or rather, a stereotypical utterance is given as a member of the ad hoc superordinate, leading to the contemporary function of *genre* as a quotative marker (probably converging with exemplification, see Section 6.3), such as in Example (30):

(30) Mme Piédefer, flattée dans sa fille, se permit aussi de dire **des choses dans ce genre** : « Ma fille, qui est une femme très supérieure, écrivait hier à Mme de Fontainetelles, telles choses. »

        (FRANTEXT, Balzac, Honoré de (1843) : *La Muse du département*)

'Mrs. Piédefer, flattered in her daughter, also took the liberty of saying **things like that**: "My daughter, who is a very superior woman, wrote to Mrs. de Fontainetelles yesterday, such things."'

In this example as well as the others found in the subcorpus, *genre* still functions as a taxonomic noun establishing a superordinate category, however, specifying a noun referring to a speech act (in 30) *des choses* 'things'). As soon as *genre* appears on its own without a verb of communication or a noun referring to metadiscursive entities it is clearly established as a quotative marker (see Cheshire & Secova 2018, on the current uses of quotative *genre*).

Appositions generally may appear with and without the preposition *de*, while the original construction with a relational genitive requires the preposition after *genre*. However, today a reduced variant coexists as an outcome of pragmaticalization, possibly in part converging with the prepositionless apposition. Appositions may have speeded up the process, since appositions and relational genitives are not always easy to distinguish, as in the case of some uses with proper names just discussed. Another use may have contributed to the reduction of the first preposition (leading from X *dans le/du genre (de)* Y to X *genre* Y), i.e. uses where X is followed by a comma, thus *(du/dans le) genre (de)* Y is used as an apposition of X, not infrequent in my data, as in Example 31):

(31) je comprends, dit-il enfin ; c'est pour un roman historique, **genre Dumas**.

（FRANTEXT, Nerval, Gérard de (1854): *Les Filles du feu*)

'"I understand", he finally said, "it's for a historical novel, **type Dumas**"'

When both prepositions are eliminated, *genre* directly links two elements and functions as a preposition itself (Danon-Boileau & Morel: 1997: 194–195). For French, such reduced constructions are attested from the 19th century onward (Mihatsch 2010a, 4.3.3), although the reduction of the second preposition seems primary, I have not detected any uses of the pattern X *genre de* Y, but X *du genre* Y and X *genre* Y. Here further studies will be needed.

Today, the reduced variant is by far the most frequent, at least in colloquial language. Chauveau-Thoumelin (2018: 182) observes for *genre* – in a balanced corpus – a dominance of the reduced variant NP *genre* NP in 81% of these constructions, while NP *du genre* NP is used in 18% of the cases, and only very marginally *du genre de* and *dans le genre de.* The overall frequency of this construction is highest in the rather informal internet forum doctissimo. In my 19th century-subcorpus the overall frequency of the full construction X *du/dans le genre de* Y is still the most frequent one (74%), compared to 15% of the type X *du/dans le genre* Y and 11% of the X *genre* Y-type (including cases with *genre* following commas). Figure 5 shows in some detail the more or less reduced variants and those with *de*, the source of the pragmaticalized marker, and *dans*:
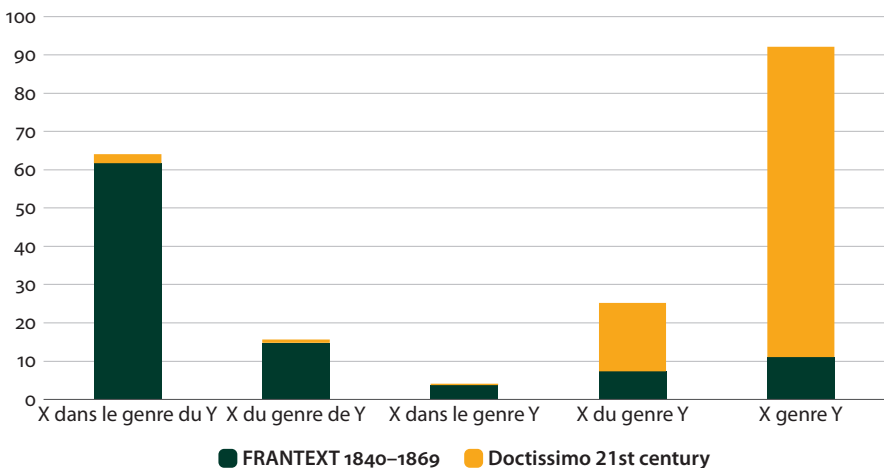
Figure 5. Variants of the X *genre* Y-construction in FRANTEXT 1840–1869 and Doctissimo (Chauveau-Thoumelin 2018: 182)

The three variants are now well entrenched with French *genre*, however, other taxonomic nouns can be found, notably *type* (in my 19th century-subcorpus 943 occurrences, roughly a third of the 2938 occurrences of *genre*). This is particularly important, since in many languages, cognates of *type* rather than *genus* appear in just this construction and become pragmaticalized in the same functions as contemporary French *genre* – namely Italian, Portuguese, some varieties of Spanish, Russian and Swedish (Kolyaseva 2018; Kornfeld 2013; Marques 2015; Mihatsch 2018; Pereira Lima-Hernandes 2005: 169; Rosenkvist & Skärlund 2013; Voghera 2013). Cognates of *type* are particularly suited for this construction since in scientific/learned taxonomies the type acquires a special function in biological taxonomies toward the end of the 19th century (Daston 2004). *Type* comes to designate a prototypical exemplar belonging to a category, or, alternatively, a specimen of a species, thus a prototypical subkind stabilizing a superordinate category. However, such uses of *type* in taxonomies appear even earlier, as can be seen in Example (32):

(32)   Ainsi, nos canards et nos oies domestiques retrouvent **leur type** dans les canards et les oies sauvages (…)   (FRANTEXT, Cabanis, Pierre (1808): *Rapports du physique et du moral de l'homme* 1)
'Thus, our ducks and domestic geese find **their type** in ducks and wild geese'

This procedure, i.e. the categorization on the basis of a prototype member or individual is widely accepted and employed by experts in the 19th century, in (33) an example from grammaticography:

(33)   Le mot fragrare « exhaler une odeur » est fort intéressant à notre point de vue, car dans **les formes du type frágro**, c'est le second groupe qui devait subir la dissimilation, tandis que dans celles du type fragráre c'est le premier.
(FRANTEXT, Grammont, Maurice (1895): *Dissimilation consonantique*)
'The word fragrare "exhale a scent" is very interesting from our point of view, because in the forms of the frágro type, it is the second group that had to undergo dissimilation, while in those of the fragráre type it is the first one'

Here, two interpretations are possible, either the prototype reading with an apposition directly naming the type which establishes the superordinate category, or the relational genitive indicating a subcategory assigned to the same implicit superordinate. Many uses of *type* further point to a reinterpretation, leading to the meaning 'category' and not prototype, showing that speakers and writers do not always distinguish neatly between horizontal categorization on the basis of a prototype and vertical categorization. The construction based on the relational genitive with *type* meaning 'category' then establishes an ad hoc superordinate via the procedure explained above.

In (34) below – after the ad hoc construction with *type* based on one individual illustrating the implicit superordinate –, two superordinate labels are given, Spanish

chatas and French *fausses maigres*, which the author does not assume to be known to the reader, again a case as in (26), where the writer first constructs an implicit ad hoc category by giving known examples, easily inferable by the reader, and then provides the reader with the established but perhaps not familiar expression:

(34) La « terrible Suzanne » ressemblait encore au portrait que Lautrec avait fait d'elle. La légende affirmait qu'elle avait été écuyère dans un cirque. C'était **une petite femme du type de Victorine**, qu'on appelle des « chatas » dans l'argot madrilène et, dans le langage familier, des « fausses maigres ».

<div align="right">(**FRANTEXT,** Carco, Francis (1941): *Nostalgie de Paris*)</div>

'The "terrible Suzanne" still looked like the portrait Lautrec had painted of her. Legend has it that she was a horsewoman in a circus. She was **a little woman of the type of Victorine**, called "chatas" in Madrid slang and, in colloquial language, "fake skinny ones"'

In French, however, unlike in other languages, *type* does not develop pragmaticalized uses, while *genre* does.

To sum up, several constructions containing *genre* attested in the 19th century (some of which are older) have been shown to serve different purposes beyond taxonomic organization:

a. Uses as adjectival support (4.2.1) and uses in combination with determiners and quantifiers (4.2.2) which suggest thoughtful categorization. Here, the taxonomic noun gets backgrounded, a reinterpretation facilitated by the frequent conflation of type and instance readings as discussed in 4.

b. A subcategory use as in *un genre de* X (4.3.) reinterpreted as a hedging expressing allowing flexibilized ad hoc categorization.

c. Two constructions establishing implicit ad hoc superordinate categories:
   - Anaphoric uses extracting implicit category properties from preceding category members (5.1), a construction related to general list extenders, but which does not show any further developments.
   - The "X of the same type as Y"-construction, which, although not the most frequent construction in my 19th-century subcorpus, is the source of the contemporary informal uses of *genre* as a pragmatic marker.

In the next section, I will show how this last construction is reinterpreted as a comparative construction, which in turn serves as a starting point for complex pragmaticalization processes. The expression leaves behind the creation of ad hoc categories with usually implicit common properties and arrives at a procedural meaning signaling ad hoc categorization via comparison. Ad hoc categorization gives instructions for the interpretation of given lexical expressions, but does not create concepts or categories.

6.   From ad hoc category to sloppy categorization: 20th and 21st century *genre*

The construction X (*du/dans le*) *genre* (*de*) Y is far less frequent in my 19th century corpus than the other uses such as the anaphoric contexts (cf. Figure 4), nevertheless, this construction leads to the highly frequent proceduralized uses of contemporary colloquial French. Further studies will be needed to investigate in which contexts the construction gains momentum. In what follows I will focus on the functional transformations of the construction discussed in the preceding section. I will argue that the complex mechanism of construing an ad hoc superordinate is reinterpreted as a simpler instruction for spontaneous, "non-canonical" categorization signaling imprecision and a number of derived functions. These are procedural in nature, i.e. expressing instructions which are stored in procedural memory, not directly accessible to conscious manipulation, as opposed to declarative semantic memory. Proceduralization characterizes grammaticalization processes in a wider sense, both grammaticalization in a narrow sense, reserved for grammatical morphemes such as inflection, and for pragmaticalization, i.e. the evolution of pragmatic markers.

So far, the focus in studies of French *genre* has been on the various discourse functions in contemporary colloquial French. However, several studies also point out possible paths of pragmaticalization (most authors adopt the term grammaticalization in a wider sense), mainly based on evidence from colloquial contemporary French (Chauveau-Thoumelin 2016, 2018; Cheshire & Secova 2018; Danon-Boileau & Morel 1997; Dufaye 2016; Fleischmann 1998; Isambert 2016; Mihatsch 2007, 2016; Vladimirska 2016, among others). One important argument is the formal reduction commented on in the preceding section, Rosier (2002) sees a close relation between the reduction of the construction containing *genre* and its reinterpretation as an approximator (2002: 80–84). Danon-Boileau and Morel (1997) point out the correlation between the selection of the reduced variants and spontaneous discourse. Further corpus analyses are necessary here, although intuitively this makes sense, since the more pragmaticalized functions are tied to colloquial language. Furthermore, the pragmatic marker *genre* itself has undergone phonetical reduction, as pointed out by Vladimirska (2016: 200), showing a flat intonational contour and a low intensity in all pragmaticalized uses (also see Hennecke/Mihatsch to appear) and often appearing in a short form [ʒã]. Similar processes have been observed for English *sort of/ kind of* (Dehé & Stathi 2016), the equivalent of *une espèce de*, but more strongly pragmaticalized (Aijmer 2002: 175–209).

I argue that *genre*, or rather, the particular construction containing *genre* analysed above (also see Chauveau-Thoumelin 2016, 2018), has gone through a process

or various processes of pragmaticalization, transforming *genre* into a procedural element giving background instructions for the interpretation of lexical units, but also larger discourse units. In what follows the focus will be on the functional changes this construction undergoes.

6.1   From ad hoc category to ad hoc categorization

The first and basic step in the process of transformation is the shortcut very plausibly arising when the construction is used in contexts where taxonomies and the explicit construal of categories become secondary. This is the case outside learned, scientific texts (which include the occurrences in Balzac or Flaubert and their explicit, fine-grained analysis of social types). Then the construction X *of the same type of* Y is reinterpreted as X *like* Y, eliminating the implicit superordinate and directly comparing two items. In both interpretations, however, the second part of the construction restricts the reference of the first NP. It is nearly impossible to tell whether speakers or writers using the construction now (but also in former times) take the mental shortcut via a comparison or not.

However, in the following examples taken from the FrTenTen12 corpus,[14] a construal of a superordinate is hard or impossible to imagine. This is the case of uses introducing metaphors (also see Chauveau-Thoumelin 2016, for a lucid discussion of such metaphorical uses), thus establishing a similarity between two different domains belonging to two different taxonomies as in (35):

(35)   Mais je vous promets pas que ce soit super compréhensible, la fiscalité française c'est **genre un labyrinthe aux murs qui bougent**….

(FrTenTen12, freegan.fr)

‘But I can't promise you that it's super understandable, the French tax system is **like a labyrinth with moving walls**’

Chauveau-Thoumelin (2016) also analyses two other telling uses, in (36), linoleum is not classified as belonging to the same category as tiling, but rather, as he convincingly points out, single properties are compared, thus meaning “a kind of linoleum which in one respect resembles tiling”:

---

14. Most colloquial contemporary examples of this section are from the web corpus FrTenTen12. This corpus contains internet texts collected in 2012, including highly informal written French. In order to exclude other varieties or learner varieties as far as possible I only searched French domains. The pragmaticalized functions of *genre* are translated by the English equivalent *like*.

(36) J'habite depuis 3 ans dans une maison ou le sol de la samme et en lino épais (**genre carrelage**) et j'en ai marre car mm en faisant le menage tous les jours j'ai l'impression que ma maison est tjs sale à causes des traces de chaussures, de chaises, de rayures etc…   (Doc., bricolage, in Chauveau-Thoumelin 2016)
'I have been living for 3 years in a house where the floor of the living room [*samme* is probably a typo of *salle*] is made of thick linoleum (**like tiling**) and I am fed up because mm even if I do the cleaning every day I still feel that my house is dirty because of the traces of shoes, chairs, scratches I….'

Again, no superordinate is referred to, *linoléum* and *carrelage* are not interpreted as belonging to one common superordinate.[15]

Similarly in (37), where the target, a particular casserole, resembles in some aspects (lid, handles…) a *cocotte,* but is not interpreted as belonging to one common category:

(37) Dans une casserole (**genre cocotte**), mettre à chauffer à feu vif l'huile d'olive (ou margarine).   (L' Est Républicain, in Chauveau-Thoumelin 2016)
'In a saucepan (**like a casserole dish**), heat the olive oil (or margarine) over high heat'

Uses with numerals also clearly leave behind the realm of ad hoc superordinates, since here imprecise readings of numerals are created. There is no imaginable superordinate based on common properties, but rather an imprecise zone of a numerical scale as in (38), although suggesting that the speaker ponders on the granularity of imprecision:

---

15. This is also true for cases where there is a metonymic shift such as from an object (a wet mouldy glove) to the effect it causes (a bad smell):

(a) Il y a une odeur épouvantable dans la salle de bain, *genre gant mouillé qui moisit depuis des semaines.*   (Doc., bricolage, in Chauveau-Thoumelin 2016)
'There is a terrible smell in the bathroom, like a wet glove that has been moulding for weeks'.

This use might be related to the prepositional use of *genre* already attested in the 19th century and related to 'manner'/'style', linking entities of different taxonomies as in (b), leading to a PP-status of the whole construction (see Section 5.2):

Au-dessus de la cheminée, une glace sans tain, historiée d'arabesques de couleur et de caractères persans, **genre café turc.**   (FRANTEXT, Goncourt (1863): *Journal*)
'Above the fireplace, a one-way mirror, decorated with coloured arabesques and Persian characters, of the type of a Turkish café.'

(38)   pensez vous qu'**un cailloux genre 1 kg** en bout d'aile peut aider a une eventuelle fermeture de bout d'ailes   (FrTenTen12, parapente-baronnies.superforum.fr)

'do you think that **a pebble of like 1 kg** at the tip of the wing can help an eventual closing of the tip of the wing?'[16]

Comparative uses expressing similarity, but not identity, are typically used when a speaker is unsure about labelling, often in combination with other expressions signaling metalinguistic doubts as in (39):

(39)   Ben en 'fait ce n'est pas un partenariat mais comment dire, **genre un jumelage** tout en conservant chacun son identité, genre 2 ententes potes tu vois ce que je veux dire ?   (FrTenTen12, blogit.fr)

'Well, it's not a partnership, but how can I say, **like twinning** while keeping each one their own identity, like two buddy agreements, you know what I mean?'

This explains why comparison markers frequently become approximators or imprecision signals.

I argue that ad hoc categories are conceptual in nature, since we can reflect on the implicit common properties, while comparison and approximation, i.e. the sloppy assignment of a category, are procedural in nature, giving rough instructions.

6.2   From comparison marker to approximation

As argued above, these conceptually simplified uses expressing similative comparison (see Haspelmath & Buchholz 1998) are roughly equivalent to *comme* 'like' and may be used in two different contexts in constructions linking two NPs. One context is the restrictive characterizing position as in *une maladie genre SIDA* 'disease of the type of AIDS/ like Aids' referring to a particular kind of disease, here construed with the help of an implicit ad hoc superordinate (sexually transmitted disease or a disease affecting the immune system, again, several superordinates can be imagined). Here 'disease' is specified. The second context is non-restrictive illustrating or exemplifying as in *une maladie, genre SIDA* 'a disease, like AIDS', referring to diseases in general, with one example.

Similative comparison is as such imprecise and approximative. However, the specialization as an approximator can be seen in the categorial change *genre* undergoes and which shows in its use in contexts that do not refer to comparisons of

---

16. Here, in the English translation the preposition is needed, since *like* introducing an approximative numeral is not prepositional, but adverbial.

two entities. This is also the case of French *comme* (Mihatsch 2009) and English *like*, although *genre* has evolved further than French *comme* and is more similar to English *like* (Andersen 2001), with which it shares all functions.

One plausible bridging context (in Heine's sense, cf. Heine 2002) is the use of the comparative preposition *genre* in copula constructions (see the arguments of Mihatsch 2009, for *comme* and its equivalents in other Romance languages), where the attribute can be nominal, adjectival or prepositional. In copula constructions prepositional *genre* may be easily reanalysed as an adverbial modifying a nominal attribute. The adverbial status is obvious when *genre* introduces NPs in argument position, a switch context in the sense of Heine (2002):

(40) on était nombreux et on avait genre **des horaires ou des tickets pour jouer**, c'était toujours douloureux (…) (FrTenTen12, white.wind.free.fr) 'we were many and we had **like schedules or tickets to play**, it was always painful'

However, today the older prepositional use as in (41) coexists with the adverbial use as in (40):

(41) Pour le reste le beau fixe, un divorce qui s'étire **genre guerre de tranchée**, mais tout va au mieux dans le meilleur des mondes. (FrTenTen12, Superforum) 'For the rest, everything's fine, a divorce that stretches **like trench warfare**, but all is well in the best of all worlds.'

As the examples above clearly show, approximation or imprecise categorization is often accompanied by subjective nuances such as a speaker signaling formulation problems, taking a distance, expressing terminological doubts, which in turn are reinterpreted as softening an assertion or petition, thus weakening the illocutionary force of an utterance. While approximation is still truth conditional, although with subjective overtones and diverse pragmatic effects, once *genre* is used in contexts not allowing approximation and where it only carries pragmatic functions, it does not affect the truth conditions of a proposition any longer (on the transition from approximation to mitigation and the intervening mechanisms of change, notably politeness principles, see Beeching 2007: 86–87).

Pure mitigation excluding approximation is typical for negative statements, which it softens, as in (42):

(42) Même quand ya des belles actions au milieu de tout ça on les savourent **genre moins**… (FrTenTen12, kenbogard.fr) 'Even when there are beautiful actions in the middle of all this we enjoy them **like less**….'

This reinterpretation arguably explains in part the great syntactic flexibility of *genre* today in diverse positions in a proposition, but particularly at the left sentence periphery and less frequently right periphery with broad scope. The following examples, taken from Vladimirska (2016: 198) and inspired by Dufaye (2016), may illustrate this point. While 43a) and b) are arguably cases of imprecision marking on a propositional level (possibly with more pragmatic overtones in 43a), example 43c) exclusively operates on the illocutionary level and mitigates the assertion or serves as a marker indicating exemplification, illustration or just narrative progress. I consider Example 43d) as an ambiguous case between a narrow hedging reading, hedging the quantity (10 euros), and a wide scope reading referring to the whole speech act:

(43) a. *Il l'a acheté genre pour 10 euros.* 'He bought it like for 10 euros'.
      b. *Il l'a acheté pour genre 10 euros.* 'He bought it for like 10 euros.'
      c. *Genre, il l'a acheté pour 10 euros.* 'Like, he bought it for 10 euros.'
      d. *Il l'a acheté pour 10 euros, genre.* 'He bought it for 10 euros, like.'

Quite often, in all uses, but particularly where we have subjective effects, *genre* may be separated from the host by commas or comma intonation, as in (44):

(44) Les iraniens sont un peuple inventé ? Typés européens, *pour des européens, c'est un peu,* **genre,** *normal.*   (Isambert 2016, http://forum.ados.fr/actu/
actualites/combattants-racontent-ciruler-sujet_24709_6.htm,
I have slightly enlargened the context)
'The Iranians are an invented people? European types, for Europeans, that's a bit, **like,** normal'

The detachment leading to parenthetical-like behaviour is typical of pragmatic markers, a spontaneous process Kaltenböck, Heine and Kuteva (2011: 874f.) call cooptation. Adverbial *genre* thus strongly differs from *un genre de*, which has not reached the same degree of pragmaticalization – unlike the English equivalents *kind of* and *sort of*, which show a similar process of detachment or cooptation.

  Mitigation in turn (possibly converging with approximation) may be reinterpreted as a marker of non-contrastive focus, since many uses of mitigators (but also approximators, maybe converging) appear in rhematic position. The focussing function is evident in contexts which do not usually allow approximation or attenuation, as is the case of intensified expressions as in (45) and (46):

(45) … il rajoute : l'Inquisition n'existe plus. Je lui fait alors remarquer que c'est **genre un truc un peu important** dans l'historique du jeu (…)
                                                              (FrTenTen12, free.fr)
'he adds: the Inquisition no longer exists. So I point out to him that it's **like something a bit important** in the game's history.'

(46)   Enfin quoique, il fait **genre super froid** dans une église, et vu que j'ai pas mangé
         depuis un bout bah, j'ai encore plus froid.   (FrTenTen12, Probb.fr)
         'Well, though, it's **like super cold** in a church, and since I haven't eaten for some
         time I feel even colder.'

It is not clear to me whether numerical approximation directly derives from se-
mantic approximation or rather from pragmatic effects such as distancing, this is
what a cross-linguistic analysis based on the idea of semantic maps suggests for
equivalent markers, which never show uses of numerical approximation without
having mitigating functions, but not the other way round (Mihatsch 2010: 250).

6.3   From comparison marker to exemplification marker to connecting uses

Exemplification is closely linked to comparison. Generally, markers of similative
comparison are typically used in constructions expressing exemplification,[17] i.e. in
non-restrictive or restrictive contexts adding an illustration via comparable items,
often accompanied by other exemplification markers as in (47) and (48):

(47)   Prenez un service public (**genre université**) avec 20% de techniciens payés 1800
         et 80% de cadres payés 2400 (…)   (FrTenTen12, politique.lefigaro.fr)
         'Take a public service (**such as a university**) with 20% of technicians paid 1800
         and 80% of executives paid 2400 (…)'

(48)   (…) confirmer quelque chose, **genre un brunch par exemple**
                                                     (FrTenTen12, grenouillebleue.fr)
         'confirm something, **like a brunch for example**'

Quite frequently, people are characterized via examples of typical or stereotypical
hypothetical statements, in my opinion this is a plausible bridging context in the
sense of Heine (2002) leading to clear uses as a quotative marker without verbs
of saying or metalinguistic nouns, as was still the case in the 19th century (cf.
Example (30)). Possibly, the illustrative, often explanatory function of exemplifi-
cation also leads to connector uses with an explanatory function and from there
generalizing to connectors marking a narrative break (Isambert 2016) or progress
or a prominent point in the narrative as in (49):

---

17.   The relation between ad hoc categorization and exemplification is bidirectional, Lo Baido
(2018) shows very convincingly for Italian how exemplification strategies are used to build ad
hoc higher-level categories.

(49)   hier on était allés au resto, et à ce resto y avait mon mec (moi j'le savais puisqu'on avait prévu de s'y voir) et **genre** dès qu'on arrive elle voit mon mec et elle crie …   (FrTenTen12, forum internet)[18]

'yesterday we went to the restaurant, and at that restaurant there was my boyfriend (I knew it because we had planned to meet there) and **like** as soon as we arrived she saw my boyfriend and she screamed'

6.4   A tentative semantic map for the pragmatic marker genre

In the preceding sections I argued in favour of diachronic relations between the different functions of *genre*, linked via implicatures and leading to reinterpretations from ad hoc category to comparison, and from there, to two paths of pragmaticalization:

1.   Restrictive comparative uses lead to uses as an approximator, subjective overtones further lead to mitigation and numerical approximation, the rhematic uses of mitigators may explain the evolution of a non-contrastive focusing function.

2.   Comparison may be used in non-restrictive contexts for exemplification, which in turn may lead to quotatives via illustrative stereotypical quotes, the change of perspective brought about by quotatives and exemplification may lead to a reinterpretation as a general connector expressing explanation, change of perspective, narrative rupture, or very generally structuration.

Further converging paths are possible, focusing can go back directly to approximation or to mitigation or both, since both are correlated with rhematic uses. The connecting function might derive from quotative uses, but also the explanatory,

---

18. Vladimirska (2016: 201) suggests as a general function of such connector uses a change in perspective, however the functions of some of her examples additionally have more specific functions such as quotation in (a) and exemplification in (b):

(a)   oui quand même oui **genre** ah comment t'es habillée moi tu serais ma petite sœur j'te laisserais pas sortir comme ça.   (CFPP2000, in Vladimirska 2016: 201)

'Yeah, well, yeah, **like** how you dress, if you were my little sister, I wouldn't let you out like that'

(b)   … avec la police j'ai eu des problèmes tout seul par exemple je me promenais avec mes amis et **genre** on me contrôlait moi tout seul enfin.

(CFPP2000, in Vladimirska 2016: 201)

'with the police I had problems just me, for example I was walking with my friends and **like** I was the only one being controlled then'

illustrative uses of examples. Numerical approximation might derive from the distancing effect of approximation or mitigation.

In the absence of reliable diachronic data for this rather recent phenomenon rooted in colloquial oral speech, semantic maps (see Haspelmath 2003 and van der Auwera 2013) may help to detect paths by comparing the functions of equivalent expressions in different varieties or languages showing patterns of polysemy which correspond to a coherent cross-linguistic map. Data from Spanish equivalents (Mihatsch in preparation) suggest diachronic changes as in Figure 6 from left to right and with an unclear status of numerical approximation, which does not seem to be tied closely to mitigation:
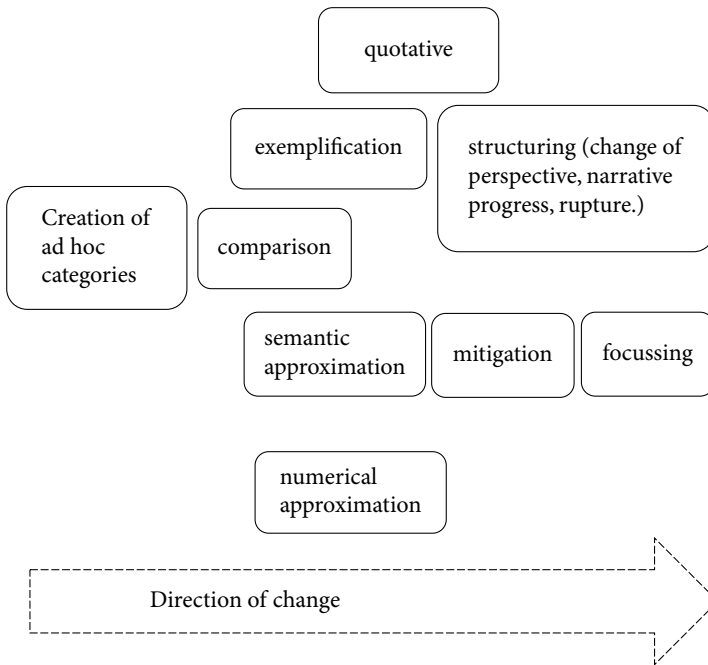
Figure 6. A tentative semantic map for the functions of *genre*

Further studies are needed to confirm this map, however, quite remarkably the same array of functions can be observed in several languages, albeit rather with *tipo* (see section above) as in Brazilian and European Portuguese, Italian, some varieties of Spanish, Russian and Swedish, starting between the 80s and 90s with young speakers (Fleischman 1998: 20; Kolyaseva 2018; Kornfeld 2013; Marques 2015; Pereira Lima-Hernandes 2005: 169; Rosenkvist & Skärlund 2013; Voghera 2013).

7. Conclusion

Apart from naming hierarchical levels taxonomic nouns take over a wide array of functions, which may lead to processes of grammaticalization or pragmaticalization. The ease with which taxonomic nouns are backgrounded and reinterpreted is due to the generally observable conflation of reference and categorization in noun phrases, with and without taxonomic nouns. Some taxonomic nouns appear in constructions which build ad hoc categories, notably the constructions *this type of Y* and *X of the type of Y*. The latter construction is the starting point for procedural functions of the French marker *genre*. The construction arises in the 19th century and is later reinterpreted as a cognitively simpler comparative construction, which, in turn, leads to functions such as approximation, exemplification, mitigation and further uses. Instead of building up ad hoc categories, *genre* then gives instructions determining the interpretation of other linguistic expressions.

Acknowledgements

References

Aijmer, Karin. 2002. *English Discourse Particles: Evidence from a Corpus* [Studies in Corpus Linguistics 10]. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.10

Andersen, Gisle. 2001. Pragmatic Markers and Sociolinguistic Variation. A Relevance-theoretic Approach to the Language of Adolescents [Pragmatics & Beyond New Series 84]. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.84

Ariel, Mira & Mauri, Caterina. 2018. Why use *or*? *Linguistics* 56(5): 939–993. https://doi.org/10.1515/ling-2018-0020

Barsalou, Lawrence W. 1983. Ad hoc categories. *Memory and Cognition* 11(3): 211–227. https://doi.org/10.3758/BF03196968

Beeching, Kate. 2007. A *politeness-theoretic approach* to *pragmatico-semantic change*. *Journal of Historical Pragmatics* 8(1): 68–108. https://doi.org/10.1075/jhp.8.1.05bee

Berlin, Brent. 1972. Speculations on the growth of ethnobotanical nomenclature. *Language in Society* 1: 51–86. https://doi.org/10.1017/S0047404500006540

Berlin, Brent. 1992. *Ethnobiological classification: Principles of Categorization of Plants and Animals in Traditional Societies*. Princeton NJ: Princeton University Press. https://doi.org/10.1515/9781400862597

Brems, Lieselotte & Davidse, Kristin. 2010. The grammaticalization of nominal type noun constructions with *kind/sort of*: Chronology and paths of change. *English Studies* 91(2): 180–202. https://doi.org/10.1080/00138380903355023

Caffi, Claudia. 2007. *Mitigation*. Amsterdam: Elsevier.

Chauveau-Thoumelin, Pierre. 2016. De l'exemplification à la catégorisation approximative: Étude de la construction [[X]SN genre [Y]SN]. *Congrès Mondial de Linguistique Française (CMLF 2016)*, Jul 2016, Tours, France. Actes du Congrès Mondial de Linguistique Française. <https://www.shs-conferences.org/articles/shsconf/abs/2016/05/shsconf_cmlf2016_12005/shsconf_cmlf2016_12005.html> (2 May 2017).

Chauveau-Thoumelin, Pierre. 2018. Exemplification and ad hoc categorization: The genre-construction in French. In *Linguistic Strategies for the Construction of Ad Hoc Categories: Synchronic and Diachronic Perspectives*, Caterina Mauri & Andrea Sansò (eds). Special issue of *Folia Linguistica* 52(s39–1):): 177–199. https://doi.org/10.1515/flih-2018-0004

Cheshire, Jenny & Secova, Maria. 2018. The origins of new quotative expressions: The case of Paris French. *Journal of French Language Studies* 28(2): 209–234.
https://doi.org/10.1017/S0959269518000029

Corbett, Greville G. 2000. *Number*. Cambridge: CUP. https://doi.org/10.1017/CBO9781139164344

Danon-Boileau, Laurent & Morel, Mary-Annick. 1997. *Question, point de vue, genre, style…*: Les noms prépositionnels en français contemporain. *Faits de Langues* 9: 193–200.
https://doi.org/10.3406/flang.1997.1155

Daston, Lorraine. 2004. *Type specimens and scientific memory. Critical Inquiry* 31(1): 153–182.
https://doi.org/10.1086/427306

De Smedt, Liesbeth, Brems, Lieselotte & Davidse, Kristin. 2007. NP-internal functions and extended uses of the 'type' nouns *kind, sort*, and *type*: Towards a comprehensive, corpus-based description. *Language & Computers* 62(1): 225–255.

Dehé, Nicole & Stathi, Katerina. 2016. Grammaticalization and prosody: The case of english *sort/kind/type of* constructions. *Language* 92(4): 911–946. https://doi.org/10.1353/lan.2016.0077

Denison, David. 2002. History of the *sort of* construction family. Paper presented at the Second International Conference on Construction Grammar, Helsinki, 6–8 September 2002.

DHLF = Rey, Alain. 1998. *Dictionnaire de la langue française*. Paris: Dictionnaires Le Robert.

Dufaye, Lionel. 2016. GENRE ou le scénario d'une grammaticalisation. *LINX: Bulletin de Centre de Recherches Linguistiques de Paris-X Nanterre* 10: 45–59.

Fleischman, Suzanne. 1998. Des jumeaux du discours: *Genre* et *like*. *La Linguistique* 34(2): 31–47.

FRANTEXT = Base textuelle *FRANTEXT* <https://www.frantext.fr/>

FrTenTen12 = French Web 2012 <https://www.sketchengine.eu/frtenten-french-corpus/>

Gelman, Susan A. & Kalish, Chuck W. 2006. Conceptual development. In *Handbook of Child Psychology, Vol. 2: Cognition, Perception and Language*, Deanna Kuhn & Robert S. Siegler (eds), 687–733. New York NY: Wiley & Sons.

Gelman, Susan A. & Brandone, Amanda C. 2010. Fast-mapping placeholders: Using words to talk about kinds. *Language Learning and Development* 6(3): 223–240.
https://doi.org/10.1080/15475441.2010.484413

Godefroy, Frédéric. 1881–1902. *Dictionnaire de l'ancienne langue française et de tous ses dialectes du IXe au XVe siècle*. Paris: F. Vieweg.

Hampton, James A. 1997. Conceptual combination. In *Knowledge, Concepts and Categories*, Koen Lamberts & David R. Shanks (eds), 133–159. Cambridge MA: The MIT Press.

Haspelmath, Martin. 2003. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In *The New Psychology of Language*, Vol. 2, Michael Tomasello (ed.), 211–242. Mahwah NJ: Lawrence Erlbaum Associates.

Haspelmath, Martin & Buchholz, Oda. 1998. Equative and similative constructions in the languages of Europe. In *Adverbial Constructions in the Languages of Europe*, Johan van der Auwera & Dónall Ó Baoill (eds), 277–334. Berlin: De Gruyter. https://doi.org/10.1515/9783110802610.277

Heine, Bernd. 2002. On the role of context in grammaticalization. In *New Reflections on Grammaticalization* [Typological Studies in Language 49], Ilse Wischer & Gabriele Diewald (eds), 83–101. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.49.08hei

Hennecke, Inga & Mihatsch, Wiltrud (to appear). From taxonomic to pragmatic uses of French *genre*: Degrees of prosodic prominence as indicators of pragmaticalization. To appear in: *Clear versus approximate categorisation: A crosslinguistic perspective*, Hélène Vassiliadou & Marie Lammert (eds.) Cambridge Scholars.

Isambert, Paul. 2016. *Genre*: Une mode récente mais qui vient de loin. *Journal of French Language Studies* 26(1): 85–96. https://doi.org/10.1017/S0959269515000538

Kaltenböck, Gunther, Heine, Bernd & Kuteva, Tania. 2011. On thetical grammar. *Studies in Language* 35(4): 852–897. https://doi.org/10.1075/sl.35.4.03kal

Knittel, Marie-Laurence. 2009. Taxonomic adjectives in French: A syntactic account. <https://hal.archives-ouvertes.fr/hal-00418115> (6 March 2019).

Kolyaseva, Alena F. 2018. The 'new' Russian quotative *tipa*: Pragmatic scope and functions. *Journal of Pragmatics* 128: 82–97. https://doi.org/10.1016/j.pragma.2018.03.004

Kornfeld, Laura Malena. 2013. Atenuadores en la lengua coloquial argentina. *Lingüística* 29(2): 17–49.

Lo Baido, Maria Cristina. 2018. Categorization via exemplification: Evidence from Italian. In *Linguistic Strategies for the Construction of Ad Hoc Categories: Synchronic and Diachronic Perspectives*, Caterina Mauri & Andrea Sansò (eds). Special issue of *Folia Linguistica* 52(s39–1): 69–95. https://doi.org/10.1515/flih-2018-0007

Marques, Aldina. 2015. "Tipo": Référentiation et modalisation dans des interactions verbales orales. In *Travaux et documents 60*, Maria Helena Araújo Carreira (ed.), 249–260. Paris: Université Paris 8.

Mauri, Caterina. 2017. Building and interpreting ad hoc categories: A linguistic analysis. In *Formal Models in the Study of Language: Applications in Interdisciplinary Contexts*, Joanna Blochowiak, Cristina Grisot, Stephanie Durrlemann-Tame & Christopher Laenzlinger (eds), 297–326. Berlin: Springer. https://doi.org/10.1007/978-3-319-48832-5_16

Mihatsch, Wiltrud. 2007. The construction of vagueness: 'Sort of' expressions in Romance languages. In *Aspects of Meaning Constructing Meaning: From Concepts to Utterances*, Günter Radden, Klaus-Michael Koepcke, Thomas Berg & Peter Siemund (eds), 225–245. Amsterdam: John Benjamins. https://doi.org/10.1075/z.136.15mih

Mihatsch, Wiltrud. 2009. The approximators French *comme*, Italian *come*, Portuguese *como* and Spanish *como* from a grammaticalization perspective. In *Grammaticalization and Pragmatics: Facts, Approaches, Theoretical Issues*, Corinne Rossari, Claudia Ricci & Adriana Spiridon (eds), 65–91. Bingley: Emerald.

Mihatsch, Wiltrud. 2010. *„Wird man von Hustensaft wie so ne art bekifft?" Approximationsmarker in romanischen Sprachen*. Frankfurt: Klostermann.

Mihatsch, Wiltrud. 2016a. Type-noun binominals in four Romance languages. *Language Sciences* 53(Part B): 136–159. https://doi.org/10.1016/j.langsci.2015.05.009

Mihatsch, Wiltrud. 2016b. Collectives, object mass nouns and individual count nouns: nouns between lexical and inflectional plural marking. In *Lexical Plurals and Beyond*, Peter Lauwers & Marie Lammert (eds). Special issue of *Lingvisticæ Investigationes* 39(2): 289–308. https://doi.org/10.1075/li.39.2.05mih

Mihatsch, Wiltrud. 2018. From ad hoc category to ad hoc categorization: The proceduralization of Argentinian Spanish *tipo*. In *Linguistic Strategies for the Construction of Ad Hoc Categories: Synchronic and Diachronic Perspectives*, Caterina Mauri & Andrea Sansò (eds). Special issue of *Folia Linguistica* 52(s39–1): 147–176. https://doi.org/10.1515/flih-2018-0009

Mihatsch, Wiltrud (to appear a). Germanic, Romance and Slavic taxonomic nouns and their functions: An overview. To appear in *Type Noun Constructions in Slavic, Germanic and Romance Languages – Semantics and pragmatics on the move*. Lieselotte Brems, Kristin Davidse, Inga Hennecke, Anna Kisiel, Alena Kolyaseva & Wiltrud Mihatsch (eds.) De Gruyter Series Trends in Linguistics. Studies and Monographs [TiLSM], 352. De Gruyter Mouton, URL: https://www.degruyter.com/document/isbn/9783110701166/html

Mihatsch, Wiltrud (to appear b). A panoramic view of Romance taxonomic nouns and their derivatives. To appear in *Type Noun Constructions in Slavic, Germanic and Romance Languages – Semantics and pragmatics on the move*. Lieselotte Brems, Kristin Davidse, Inga Hennecke, Anna Kisiel, Alena Kolyaseva & Wiltrud Mihatsch (eds.) De Gruyter Series Trends in Linguistics. Studies and Monographs [TiLSM], 352. De Gruyter Mouton, URL: https://www.degruyter.com/document/isbn/9783110701166/html

OED = Murray, James. 1884. *Oxford English Dictionary Online*. Oxford: OUP. <http://dictionary.oed.com> (2 May 2017).

Pereira Lima-Hernandes, Maria Célia. 2005. A interface sociolingüística / gramaticalização: Estratificação de usos de tipo, feito, igual e como. Sincronia e diacronia. PhD dissertation, University of Campinas.

Rosch, Eleanor, Mervis, Carolyn, Gray, Wayne, Johnson, David W. & Braem Penny Boyes. 1976. Basic objects in natural categories. *Cognitive Psychology* 8: 382–439. https://doi.org/10.1016/0010-0285(76)90013-X

Rosenkvist, Henrik & Skärlund, Sanna. 2013. Grammaticalization in the present – The changes of modern Swedish *typ*. In *Synchrony and Diachrony: A Dynamic Interface* [Studies in Language Companion Series 133], Anna Giacalone Ramat, Caterina Mauri & Piera Molinelli (eds), 313–338. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.133.13ros

Rosier, Laurence. 2002. *Genre*: Le nuancier de la grammaticalisation. *Travaux de Linguistique* 44(1): 79–88. https://doi.org/10.3917/tl.044.0079

Rouget, Christine. 1997. Espèce de, genre de, sorte de: Approximatifs ou sous-catégorisateurs? In *Psychomécanique du langage. Problèmes et perspectives. Actes du 7ᵉ Colloque International de Psychomécanique du langage (Cordoue, 2–4 June 1994)*, Paulo de Carvalho & Olivier Soutet (eds), 289–298. Paris: Honoré Champion.

Sloutsky, Vladimir M. 2010. From perceptual categories to concepts: What develops? *Cognitive Science* 34: 1244–1286. https://doi.org/10.1111/j.1551-6709.2010.01129.x

Smith, Edward E. 1995. *Concepts and categorization*. In *Thinking. An Invitation to Cognitive Science*, Vol. 3, Edward E. Smith & D. N. Osherson (eds), 3–33. Cambridge MA: The MIT Press.

TLFi = *Trésor de la Langue Française informatisé* <http://atilf.atilf.fr/>

Ullman, Michael T. 2004. Contributions of memory circuits to language: The declarative / procedural model. *Cognition* 92(1–2): 231–270. https://doi.org/10.1016/j.cognition.2003.10.008

van der Auwera, Johan. 2013. Semantic maps, for synchronic and diachronic typology. In *Synchrony and Diachrony: A Dynamic Interface* [Studies in Language Companion Series 133], Anna Giacalone Ramat, Caterina Mauri & Piera Molinelli (eds), 153–176. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.133.07auw

Vladimirska, Elena. 2016. Entre le dire et le monde: Le cas du marqueur discursif genre. In *Nouveaux regards sur l'approximation et la précision*, Silvia Adler, Hava Bat-Zeev Shyldkrot & Maria Asnes (eds), 195–209. Paris: Honoré Champion.

Voghera, Miriam. 2013. A case study on the relationship between grammatical change and synchronic variation: The emergence of *tipo*[-N] in Italian. In *In Synchrony and Diachrony: A Dynamic Interface* [Studies in Language Companion Series 133], Anna Giacalone Ramat, Caterina Mauri & Piera Molinelli (eds), 283–312. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.133.12vog

Ward, Gregory & Birner, Betty. 1995. Definiteness and the English existential. *Language* 71(4): 722–742. https://doi.org/10.2307/415742

Waxman, Sandra R. & Gelman, Susan A. 2010. Different kinds of concepts and different kinds of words: What words do for human cognition. In *The Making of Human Concepts* [Oxford Series in Developmental Cognitive Neuroscience], Denis Mareschal, Paul C. Quinn & Stephen E. G. Lea (eds), 99–129. Oxford: OUP. https://doi.org/10.1093/acprof:oso/9780199549221.003.06

Zamparelli, Roberto. 2000. *Layers in the Determiner Phrase*. New York NY: Garland.

CHAPTER 15

In a manner of speaking

The co-construction of manner in spoken Italian dialogues

Luisa Corona and Paola Pietrandrea
University of L'Aquila / University of Lille & CNRS UMR 8163 STL

We provide a functional, corpus-driven definition of the category of Manner and we observe its (co)-construction in interactional discourse. In the linguistic literature, Manner has been traditionally approached as a basic cognitive prime, expressing the way in which an action is performed. It has been traditionally considered as a category fulfilling a basically descriptive function, necessarily realized by a limited inventory of means essentially encoded either at the lexical level, on verbal roots, or at the syntactic level, on verbal adjuncts. In this paper, we propose a functional definition of Manner that allows identifying Manner constructions without prejudging their morphosyntactic realization. We present an annotation scheme of Manner constructions that allows for an exploration of the realizations of this category in a corpus of spoken interactional data. The analysis of 514 occurrences of Manner constructions shows that, far from being a static descriptive monological category, Manner can be regarded as an interactional, gradable category co-constructed by participants in discourse. Manner is used by speakers not only to describe reality but also to reach, in discourse and through discourse, a shared perception of reality.

Keywords: manner, co-costruction of meaning, spoken interaction, corpus-driven

1. Introduction

In this article we explore the construction and the co-construction of the category of manner in discourse. Manner has been traditionally approached as a basic cognitive prime expressing the way in which an action is performed and realized by a limited inventory of means essentially encoded either at the lexical level, on verbal roots, or at the syntactic level on verbal adjuncts and fulfilling a basically descriptive function.

We show in this article, on the basis of a corpus-driven analysis of spoken interactional data, that far from being a static descriptive monological category, manner can be regarded, when analysed through a proper methodology, as an interactional, gradable category co-constructed by participants in discourse. This category is used by speakers not only to describe reality but also to reach, in discourse and through discourse, a shared perception of reality.

The article is organized as follows: we first review the literature on manner in order to propose a semantic definition of the notion of manner and a formal representation of manner constructions (§ 2). Then we present the corpus that we analysed (§ 3), the annotation scheme that we developed to analyse manner on spoken data (§ 4) and the results of our empirical analysis (§ 5). In § 6 we discuss the relevance of our findings for a theoretical characterization of the construction of manner (as well as descriptive categories) in and through interaction. In § 7 we present some general concluding remarks.

2.   An ill-defined category

2.1   The traditional definition

Manner is a hotly debated topic in the linguistic literature. However, as Stosic (2019: 15) puts it "the vast majority of scholars dealing with manner do not provide any definition or semantic description of this concept". In spite of the numerous studies dealing with the lexical coding of manner, we still lack a suitable and satisfactory definition of manner that can be employed as a comparative concept in language description. Thus, an analysis of manner on corpora is particularly complex.

We will first briefly review the characterizations of manner provided by the grammatical tradition of Italian linguistic studies and more recent work conducted on this notion. In the *Dizionario della lingua italiana* edited between 1861 and 1869, Nicolò Tommaseo and Bernardo Bellini defined manner as "la qualità di procedere operando", 'the quality of proceeding by doing things'. In the same vein, some years later, Sechehaye (1926) defined manner as "l'idée de la qualité appliquée à des idées essentiellement verbales" 'the idea of quality applied to essentially verbal notions'. So, manner is conceived, in these early definitions, as part of the complex notion of quality of action and as such it is regarded as a category that is only relevant for verbs. Ray Jackendoff (1983) proposed a description of manner in the frame of cognitive semantics. According to Jackendoff, manner is an ontological category, i.e. a semantic prime (*primitive,* in his words) that cannot be further analysed. In Jackendoff's view, this innate prime, derived from the domain of spatial experience, can be combined with other semantic primes such as [THING], [PLACE], [DIRECTION], [ACTION],

[EVENT], [AMOUNT] to create more complex categories and to conceptualize entities (see Jackendoff 1983: 48–56; Haspelmath 1997). Manner has also received great attention in studies devoted to verbal semantics. Levin and Rappaport Hovav (1998, 2013) and Rappaport Hovav and Levin (2010) define Manner verbs as those verbs that "specify as part of their meaning a manner of carrying out an action": for example *nibble, rub, scribble, sweep.* Interestingly enough, Levin and Rappaport Hovav hypothesize that Manner cannot be encoded in the same verbal root with [RESULT], result verbs being those that "specify the coming about of a result state", as for example *clean, cover, empty, fill.* According to Levin and Rappaport Hovav, the complementarity between these two notions is due to a semantic difference that makes them conceptually "irreconcilable". Both Manner and Result verbs are dynamic; thus, they involve change. However, as pointed out by Rappaport Hovav and Levin (2010), there is a fundamental distinction between the two types of change involved: while result roots specify scalar changes, manner roots specify non-scalar changes. These two types of change are in complementary distribution in verbal roots: a verb can only lexicalize one of these two types of meanings.

Levin and Rappaport Hovav's hypothesis has been much debated (see Goldberg 2010; Beavers & Koontz-Garboden 2012; Franco & Lorusso 2020) and, as we will show, the analysis of spoken data can help to challenge it.

The tradition of studies stemming from Talmy's (2000) typology, based on the contrast between *Verb-Framed* (*VF*) and *Satellite-Framed* (*SF*) languages, is essentially centred around the cross-linguistic encoding of two main components in the conceptualisation of motion events: path, the core notion of translation, and the co-event, manner, defined as "the way in which the motion is performed".[1] In this tradition, a major impetus to the definition of manner came from Slobin's studies on the encoding of manner of motion (Slobin 2004, 2005, 2006 among others). Slobin, who characterized manner as "an ill-defined set of dimensions that modulate motion", confirmed the cross-linguistic validity of Talmy's hypothesis that manner is encoded on verbal roots in *SF* languages. He also found a clear correlation between Talmy's typology and the variety of expressions of manner in a given language: according to his results, *SF* languages have a larger and more diverse lexicon of manner verbs as compared to *VF* languages. Slobin explained this result by suggesting that the slot that *SF* languages devote to the expression of manner on the verbal root makes this notion more accessible to speakers and therefore more likely to be elaborated over time (see Slobin 2004: 250–252).

---

1.   The terminology used above refers to Talmy's (2000) work, which is the most organic exposition of his theories on the encoding of motion events. This model was preceded by a number of studies which represent the successive steps in his typological research. The most influential antecedents to Talmy (2000) are Talmy (1972, 1985, 1991).

The studies on manner just reviewed focus almost exclusively on the expression of this category on verbal roots: their authors consider that, since manner is defined as the way in which action is performed and since action is encoded on verbs, the privileged locus for the observation of manner must be the verb. This approach leaves no room for the analysis of manner at other levels of linguistic structure, and more generally for the exploration of *ad hoc* strategies or the co-construction of this category in discourse.

2.2   More recent approaches

A more recent approach to manner maintains that, far from being an ontological prime, this category is better regarded as a complex semantic category analysable in semantic features (see Moline & Stosic 2016; Stosic 2019, 2020).

According to Stosic (2020), a major difficulty in using manner as an analytical and descriptive category in linguistics is the proximity between manner and a set of neighbouring concepts, such as instrument (*mangiare col cucchiaio* 'to eat with the spoon'), quantity or intensity (*mangiare un sacco* 'to eat a lot'; *mangiare a dismisura* 'to eat out of all proportions'), quality (*mangiare sano* 'to eat healthy'), comparisons with referents that prototypically perform a given action (*mangiare come una bestia* 'to eat like a beast'). Stosic (2020) maintains that identifying the boundaries of the complex and "fuzzy" category of manner, from a syntactic and semantic point of view, is a challenge for those who intend to deal with it. This is especially true if one intends to study the expression of this category in discourse where, as we shall see, manner meanings are expressed by recruiting means of expression from neighbouring categories to construct extemporaneous, context-dependent, *ad hoc* representations of this notion.

Interestingly enough, Stosic (2019) also argues that, far from being just encoded on verbal roots, manner can be expressed by a variety of morphosyntactic structures: morphological items, syntactic constructions, and lexical units.

Stosic and Moline's characterization of manner as a complex category encoded at various levels of linguistic structure could be usefully employed in the analysis of spoken interactional corpora and would allow room for the identification of *ad hoc* constructions of manner in discourse. We regret, though, that the authors do not provide operational criteria to distinguish this notion from its contiguous categories such as instrument, quantity, and comparison. The vagueness of this characterization makes it impossible to use their approach in corpus-based research and to translate it into a replicable annotation schema.

2.3 Towards a functional definition of manner

In order to attain the objective of providing a solid definition of manner and at the same time observing its construction in discourse, we adopted a (i) functional and (ii) corpus-driven approach. In other words, (i) we decided not to limit our study to the analysis of verbal roots, rather, we sought to provide a theoretically meaningful definition of the functional domain of our analysis, and we then identified in corpora the constructions that express this domain, without prejudging their morphosyntactic realization. (ii) Both the definition of the domain of analysis and the identification of its relevant properties were established incrementally through progressive refinements of the traditional notions of manner in the light of corpus data.

In concrete terms, building on previous studies, we adopted the following definition of manner:

> **Manner is the explicit description of the way in which an action is performed**.

We propose to represent manner constructions as follows:

> **A manner construction *mc* consists of three elements: a scope *s* expressing an action α, a marker *m* expressing the way in which the action is performed and a relation *r* between *m* and *s*.**

Such a formalization[2] allows room for a consistent representation of any construction encoding manner, regardless of the morphosyntactic nature of its marker and whether the construction is interactional in nature or not. Let us analyse for example the following constructions:

(1) A: I will [(speak)s (in a low voice)m]r

(2) A: I will [(speak)s]
    B: yes but [(in a low voice)m]r, please

(3) A: I will [((whisper)s)m]r

In (1) the manner construction is realized by the monological linear relation between the scope *speak* and the marker *in a low voice*; in (2) the manner construction is realized by the dialogical linear relation between the scope *speak* and the marker *in a low voice*; in (3) it is realized by the relation consisting in the monological superposition of the marker and the scope in the verb.

It is important to highlight that from a semantic point of view, a manner construction *mc* should be regarded as a construction expressing an action α1:

---

2. For a theoretical justification of this formalization see Pietrandrea (2018).

$$mc = [m(s)]r$$
$$< \alpha1 >$$

By definition:

> the action α1 described by the construction mc realizes a hyponym, that is a more specific meaning, of the superordinate action α described by s.

So we can say for example that *slipping* IS *falling, frying* IS *cooking, whispering* IS *talking, speaking in a low voice* IS *speaking* and so on.

Going back to Moline and Stosic's studies which propose to regard certain occurrences of instrument (*with the spoon*), quantity (*enormously*), location (*at home*) and other contiguous categories as manner markers, the semantic characterization that we propose enables the occurrences of contiguous categories used in their own meaning to be distinguished from the occurrences of the same categories in which the central meaning is bleached and reanalysed as expressing manner. Let us compare for example, (4) and (5).

(4)   *ah la potresti mangiare col cucchiaio!*
'You could eat it with a spoon!' [talking to someone who is eating with their hands]

(5)   *c'era una mosca nel piatto, l'ho rimossa col cucchiaio*
'There was a fly in my dish, I whisked it away with a spoon'

In Example (4), the adjunct *col cucchiaio*, 'with a spoon' is not used to indicate specifically that the subject is asked to eat with a spoon rather than with another tool, but the construction is used in a broader sense to ask the subject to eat more politely than with her hands. The semantic contribution of the instrumental adjunct is not focalized *per se* and the action of eating with a spoon simply coincides with a hyponym of the superordinate action of eating. In (5), instead, the semantic contribution of the instrumental adjunct *col cucchiaio* 'with a spoon' is focalized *per se*; the speaker highlights that he whisked the fly away using a spoon (rather than with another instrument): so the action represented in (5) cannot simply be represented as a more specific meaning of the superordinate action of whisking away.

From an operational point of view, we decided to consider as desemanticized and absorbed in the manner construction all the occurrences of contiguous categories that could not be opposed to other elements of their paradigm. See Examples (6) and (7):

(6)   A: *mangia sempre al ristorante, per questo è così grasso!*
'He always eats in restaurants. That's why he is so fat!'
B: *??no, l'altra sera ha mangiato in pizzeria*
??No, last night he ate in a pizzeria

(7) A: *ho voglia di mangiare fuori a cena: andiamo al ristorante ?*
        'I'd like to go out for dinner: Shall we go to a restaurant?'
     B: *no, andiamo in pizzeria !*
        'No, let's go to a pizzeria!'

In these examples, it is clear that in (6) *al ristorante* 'in restaurants' is a locative adverbial used for encoding a sloppy or unhealthy manner of eating, whereas in (7) the same adjunct retains its locative meaning.

In our approach, we annotated as manner markers only those that expressed relations such as in (6), where *mc* encodes an action α1 that represents a hyponym of α.

3. The corpus

Once having defined manner both theoretically and operationally, we conducted an analysis of corpus data with the aim of identifying, without any pre-judgment, all the constructions that express manner in spoken Italian dialogues.

We focused our analysis on the VOLIP (De Mauro et al. 1993; Voghera et al. 2014), a corpus of spontaneous spoken Italian texts comprising 500.000 word tokens for 60 hours of recordings, collected in four different Italian cities (Milan, Florence, Rome, and Naples). The VOLIP corpus comprises five different textual typologies: face-to-face conversations; telephone conversations; interviews, debates, and classroom interactions; monologues (lectures, sermons, and speeches); radio and television broadcasts, as summed up in Table 1. We selected 40 dialogues, distributed as shown in Table 1, totalling 77.819 words.

By applying the definition of manner provided in § 2.3 and the operational criteria allowing the expression of manner to be distinguished from the expression of contiguous categories, we identified 514 manner constructions.

Table 1. Our corpus [A: Face-to-face conversations; B: Telephone conversations; C: Interviews, debates, and classroom interactions; D: Monologues (lectures, sermons, and speeches); E: Radio and television programs]

Cities	Text typology					tot.
	A	**B**	**C**	**D**	**E**	
Naples	10.377	2.640	4.484	1.207	1.000	19.708
Florence	6.266	1.931	2.745	2.680	5.425	19.047
Milan	7.076	2.592	4.093	1.174	4.296	19.231
Rome	11.466	2.200	2.149	1.182	4.018	19.833
						77.819

4. Annotation

We applied the annotation scheme represented in Figure 1 to the manner construc-
tions identified in the corpus.

Manner Marker	Morphosyntax	Verb Phrases
		Noun Phrases
		Adjectival Phrases
		Adverbial Phrases
		Prepositional Phrases
		Subordinate Clauses
		Gerundive Clauses
		Evaluative suffixes
		Onomatopoeias
		Quantifiers
		GM of Comparison
		Construction I.a
		Construction I.b
		Construction II
		Construction III
		Idioms
Manner Scope		
Manner relation	Linear Relation	M=S
		M>S
		S>M
		S-M-S
	Source	SS
		OS
	Manner Value	Modus
		Quality
		Degree
		Instrument
		Comparison
		Aspect
		Locative
		Domain
		Speed
		Quantity

Figure 1. Annotation scheme

As shown in Figure 1, the structure of our annotation scheme implements the
formal representation of manner constructions that we described in § 2.3: we an-
notated a scope, a marker and the relation between them.

```
Scope :      [predication]
             <depicted action α>
Marker :     [semantic predicate]
             <manner in which the action is performed>
Relation :   [relation between the marker and the scope]
             <hyponym of α α1>
```

4.1   The annotation of the marker

As for the marker, we annotated its morphosyntactic nature. As expected, we found that manner is encoded on:

(i)   verb phrases

    (8)   verbal root: *non brontolerò mai*
        'I will never grumble'

    (9)   phrasal verb: *l'avete buttato giù e rifatto*
        'You sketched it out and reworked it'

    (10) verb cluster:[3] *corre a visitare un malato*
        'He goes to visit a patient in a hurry'

    (11) light verb construction: *fa la scenata dell'altra volta*
        'She makes a scene like last time'

(ii)   noun phrases

    (12) *bisogna domarla bastone e carota*
        'You have to tame her carrot-and-stick'

(iii)   adjectival phrases

    (13) *poi richiama candido*
        'He calls back innocent'

(iv)   adverbial phrases

    (14) *lo faccio molto seriamente*
        'I do it very seriously'

(v)   prepositional phrases

    (15) *vivendo un po' di rendita*
        'Living off a trust fund'

---

3.   The verb cluster (or fr. *amas verbal*) is a pattern composed of one or more verbs, functional words and restricted lexical elements other than the verbs (Gerdes & Kahane 2006), as for example *ammazzarsi dal ridere* 'to laugh oneself to death'.

(vi)   subordinate clauses

    (16)   *lo firmava <u>senza sta' a rompe tanto i coglioni</u>*
            'She signed without busting my balls that much'

(vii)   gerundive clauses

    (17)   *vorrebbe il massimo <u>dando il minimo</u>*
            'He wants to get the maximum while doing the bare minimum'

(viii)   evaluative suffixes, whether affixed on the verb – as in (18) where the suffix *-ott-* in *parlottare* 'to mutter' modifies the verb *parlare* 'to speak'

    (18)   *e i colombi che <u>parlottano</u> sul terrazzo?*
            'And the pigeons muttering on the roof?'

or affixed on the noun, as in (19) where the suffix *-in-* modifies the noun *girata* 'quick ride'[4] < *giro* 'ride'

    (19)   *si va a fare <u>una giratina</u> in centro prima*
            'We go for a quick ride downtown first'

By defining our annotation scheme incrementally through a corpus-driven annotation procedure we found that manner is also encoded in:

(ix)   extemporaneous onomatopoeias, i.e. reproductions of sounds not encoded in the linguistic system[5]

    (20)   *questo fa <u>plin plin plin plin</u>*
            'To tinkle' (lit. 'to make plin plin plin plin')

(x)   quantifiers (21) and temporal markers grammaticalized as quantifiers (22) and (23)

    (21)   *vi sto facendo stare <u>un po'</u> in allegria*
            'I'm keeping you a little cheerful'

    (22)   *è una cosa che va fatta <u>un attimo</u> con ordine*
            'It's something we need to do in a rather orderly fashion [lit. an instant in orderly fashion]'

---

4. With deverbal nouns such as *giro* 'ride' (< *girare* 'to go around') or verbs such as *passeggiare* 'to stroll', the suffix *-ata* derives action nouns with a semelfactive aspectual meaning, as *girata, passeggiata*.

5. Since Slobin (2004) and Ibarretxe-Antuñano (2006), there has been growing interest in ideophones in manner encoding, especially in the motion domain.

(23) *bisogna anche* <u>*un momentino*</u> *adeguare le foto*
'You also need to adjust the photos a little bit [lit. a little moment the photos]'

(xi) grammatical markers of comparison such as *come* 'like', *stile* 'style' – already identified by Masini and Mauri (2020) – and *tipo* 'type' – already studied by Voghera (2017a; in preparation). In these constructions, manner is described by assuming as a basis for the comparison a referent which prototypically performs the action to be qualified

(24) *mangia* <u>*come una bestia*</u>
'He eats like a beast'

(25) *la trattava* <u>*tipo segretaria*</u>
'He treated her like a secretary'

(26) *fa l'arrogante* <u>*come tanti Bonito Oliva*</u>
'He gets uppity like a kind of Achille Bonito Oliva'

(xii) two constructions involving manner nouns such as *modo / maniera / forma, termine / stile*:

a. the construction [*in* + manner_NOUN + ADJ]

(27) *m'era sembrato che tu_ eh le facessi* <u>*in un modo strano*</u>
'I thought you were doing them in a weird way'

(28) *può anche scrivere* <u>*in maniera personalizzata*</u>
'You can also write in a personalized way'

(29) *lo voglio fare* <u>*in grande stile*</u>
'I want to do it in style'

b. the construction [*in* + DEM + manner_NOUN] in which the demonstrative DEM refers anaphorically (30) or cataphorically (31) to an expression of manner introduced in the context

(30) B: <u>*entro il dodici me lo paga me lo paga entro il sedici e per ora*</u>
A: *va bene*
B: *mi va bene insomma*
A: <u>*in questa maniera allora?*</u>
'B: I'll get paid by the 12th, I'll get paid by the 16th, and for now…
A: All right
B: I'm ok with that
A: So, in this way?'

(31) *allora io ho fatto* <u>*in questo modo ecco per esempio questa è una correzione nuova*</u>
'Then I did it in this way, here, for example this is a new correction'

(xiii) two constructions involving the aspectual progressive periphrasis "*stare* 'to stay' + representation of an action":

  a. [*stare lì/là in* + N] 'lit. to stay there in + N

  (32) *stanno lì in chiacchere*
      'They keep chatting over and over again [protractive aspect] / without purpose'

  b. [*stare lì/là a* + V]; [(*stare*) *lì/là che* + V] 'to stay there to + V'

  (33) *ragazzi di liceo e universitari stanno là a giocare con i bambini*
      (corpus ItTenTen16)
      'Students from high school and college are playing with the kids without purpose'

  (34) *lui tutto caruccio lei li' che civetta*
      'The guy, being all nice, and her, flirting back'

(xiv) the light verb construction [*fare* + NP], where the verb *fare* 'to do' is used in its meaning of 'to act, to behave', followed by a NP encoding a standard for the comparison. This construction can be more lexicalized as in (37) and (38) or more extemporaneous as in (35) and (36)

  (35) *allora vogliamo fare parte attiva o vogliamo fare i microfoni e gli altoparlanti di altra gente?*
      'So, do we want to be an active part or do we want to act as other people's microphones and speakers?'

  (36) *fanno il gatto e la volpe*
      'They act like the Fox and the Cat'

  (37) *fare il bastian contrario*
      'To get contrarian'

  (38) *fare il bello e il cattivo tempo*
      'To blow hot and cold'

(xv) idioms

  (39) *alla fine lei e' andata nel pallone*
      'She finally freaked out'
      andare nel pallone 'lit. to go into the ball > to freak out'

  (40) *facemmo un salotto_ io lei*
      'We made social chit-chat, me and her'
      fare salotto 'lit. to make a living room > to talk at length without purpose'

4.2 Annotation of the scope

As for the scope, we only identified it and left for future research the elaboration of a scheme of annotation for its morphosyntactic or illocutionary properties.

4.3   Annotation of the relation

We annotated three properties of the relation between the marker and the scope: their linear relation, their monological or dialogical status, and their semantic nature.

(i)   the linear relation between the marker and the scope can have one of the following values:

a.   The marker and the scope overlap

(41)   *friggere il proprio uovo*
       'To fry one's own egg'

(42)   *vagava un poco*
       'She wandered a little'

b.   The marker precedes the scope

(43)   *volentieri ti aiuto*
       'I would gladly help you'

(44)   *teoricamente, dovrei stare giù*
       'Theoretically, I should be sad'

c.   The marker follows the scope

(45)   *stiamo lavorando duramente*
       'We're working hard'

(46)   *pensavo di aver fatto una cosa da stracciare*
       'I thought I'd done something disastrous'

d.   The marker interrupts the scope

(47)   [*volevo*]$_s$ [*un pochino*]$_m$ [*relazionare*]$_s$ *su come va*
       'I wanted to report a little bit on how it's going'

(48)   [*togliendo*]$_s$ [*completamente*]$_m$ [*il traffico*]$_s$ *privato e pubblico*
       'Completely removing private and public traffic'

(ii)   concerning the monologic or dialogic status of the construction we distinguish between two cases:

a.   The marker and the scope are uttered by the same speaker

(49)   *le superfici_ eh degli ambulanti sono sono cali<brate> calibrate al millimetro_*
       'The surfaces of the peddlers are calibrated to the millimetre'

(50)   *il prodotto e' confezionato come come di norma*
       'The product is packaged as usual'

b. The marker and the scope are uttered by two different speakers, as in (51) and (52)

(51) A: *c'è invece chi aveva un appuntino suo mi pare che invece una persona l'abbia detto semplicemente_*
      B: *a voce*
      A: *a voce_ eh quindi questo*
      'A: Some people had a little note of their own, and I think that one person simply said it_
      B: by voice
      A: by voice_ yes, so this

(52) B: *e' leggermente piu' alto di te*
      I: *ah impercettibilmente*
      B: *cioe'?*
      I: *otto centimetri*
      'B: He's slightly taller than you
      I: Ah imperceptibly
      B: Meaning?
      I: Eight centimeters

(iii)   concerning the semantic nature of the manner construction we distinguished:
a. a purely manner value
    – MODUS (*viene preso di sorpresa* 'he's taken by surprise');
b. a manner value encoded through the expression of contiguous categories:
    – QUALITY (*non l'avete mai fatto in grande stile* 'you never do that in style');
    – DEGREE (*di solito parlo molto ma in certe situazioni parlo molto poco* 'I usually talk a lot but in some situations I speak very little');
    – INSTRUMENT (*l'ha fatto a macchina o con il computer* 'did he type it on the typewriter or on the computer?');
    – COMPARISON (*si presenta come il messia* 'he presents himself like the messiah');
    – ASPECT (*stanno lì in chiacchiere* 'they keep chatting over and over again" [protractive aspect]);
    – LOCATIVE (*mangiando nei barre o in questi troiai* 'eating [outside] in coffee shops or in shitty places like that');
    – DOMAIN (*abbiamo determinato in termini teorici il costo zero* 'we determined in theoretical terms the zero cost');

- SPEED (*lentamente rientrano dalla porta* 'they walk back through the door slowly');
- QUANTITY (*mi tocca andarci da sola* 'I have to go all by myself', *ci si andrà in gruppo* 'We'll go as a group').

We would like to highlight that, for theoretical reasons, the semantics of the construction is encoded on the relation rather than on the marker. Indeed, we argue that the manner meaning results from the combination of a manner (or non-manner) marker and its scope, rather than from the semantic nature of the marker only. Such a representation, already proposed and thoroughly theoretically justified by Pietrandrea (2018) for the analysis of epistemic constructions, allows us to justify the semantic shift that may occur in discourse when a marker from a contiguous category having an intrinsic meaning other than manner is reanalysed as encoding a manner meaning in a given construction.

5. Results

5.1   An extensive inventory of structures

The most striking fact in analyzing the results of our corpus analysis is the relatively low proportion of constructions in which manner is encoded on a verbal locus.

The encoding of manner on verbal phrases – which is, as already mentioned, the most widely studied manner construction – only concerns 14% of the constructions identified in our corpus: even more strikingly, the encoding of manner on verbal roots only concerns 5% of the manner constructions of our corpus. One might say that this is not surprising given the fact that Italian is a *VF* language. According to Talmy's (2000) and Slobin's (2004, 2005, 2006) hypothesis, indeed, *VF* languages prefer the encoding of manner on adjuncts rather than on verbal roots. However, the analyses of written corpora and lexical repertoires conducted by Cardini (2008) and Iacobini (2010) showed that Italian has a lexical inventory of manner verbs that is much richer than the inventories generally present in *VF* languages. What our analysis of spoken dialogues reveals is that manner is mostly encoded through prepositional and adverbial phrases: these markers represent 55.2% of our sample (32.9% of prepositional phrases and 22.3% of adverbial phrases). So, even though the Italian system can be considered as an exception among *VF* languages due to the richness of the inventory of verbal roots, it is true that Italian tends to use these lexical resources very infrequently in discourse.

But manner is also encoded through a variety of other markers as shown in Figure 2.
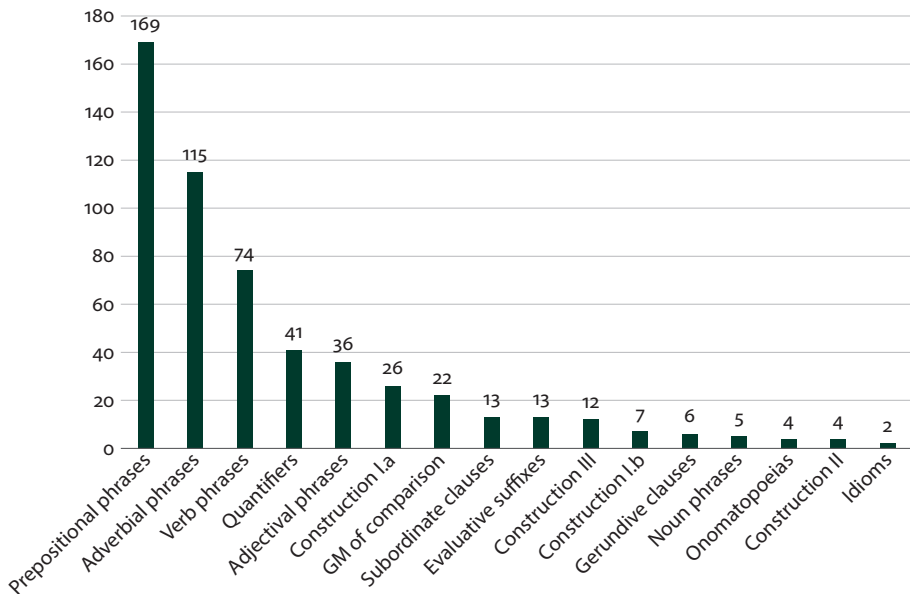
Figure 2. Morphosyntax of Manner markers
[Constructions I.a and I.b are described in point (xii) of the Annotation scheme;
Construction II is described in point (xiii) of the Annotation scheme;
Construction III is described in point (xiv) of the Annotation scheme]

Interestingly enough our data show that it is not infrequent for manner markers to cluster over one and the same scope (it happens in 103 constructions out of 514, that is 20% of the corpus). From a syntactic standpoint it is interesting to notice that such an accumulation of manner markers is realized through two strategies: either the manner marker$_2$ syntactically depends on manner marker$_1$, as in Examples (53) and (54) where we find a manner verb (*bollire* 'to boil', *fuggire* 'to run away') modified by a manner adjunct (Prepositional Phrase, *a fuoco lento* 'on a low heat', *in esilio* 'in exile'), respectively:

(53)   [[*bollire*]$_{m1}$ [*a fuoco lento*]$_{m2}$]$_{m3}$
        'To simmer'

(54)   [[*fuggire*]$_{m1}$ [*in esilio*]$_{m2}$]$_{m3}$
        'To bolt into exile'

or manner markers are piled up in list structures (Blanche-Benveniste 1990; Masini et al. 2018; Kahane et al. 2019), as in Example (55). We define list, following Masini et al. (2018: 50) as "the syntagmatic concatenation of two or more units of the same type (i.e. potentially paradigmatically connected) that are on a par with each other, thus filling one and the same slot within the larger construction they are part of".

Thus, in (55), the extemporaneous manner onomatopoeia *ta ta ta ta,* occurs in a list with the manner adjunct *in serie* 'serially', then with a repetition of the onomatopoeia *ta ta ta ta,* and finally with the manner adjunct *senza l'agenda davanti* 'without an organizer before one's eyes' to progressively construct the manner meaning 'very efficiently'

(55) [[*ta ta ta ta*]$_{m1}$ [*in serie*]$_{m2}$ [*ta ta ta*]$_{m3}$ [*senza l'agenda davanti*]$_{m4}$]$_{m5}$ *quello è lavorare*
'Ta ta ta ta, serially, ta ta ta, without an organizer before one's eyes, that's working!'

5.2   A complex semantics

Figure 3 shows the frequency distribution of the different manner meanings in our corpus. While manner is largely encoded in discourse, as well as in the linguistic system, by markers specifically dedicated to the expression of this notion (30%), the recruitment of contiguous (bleached) meanings for the expression of manner covers 70% of the occurrences of our corpus.
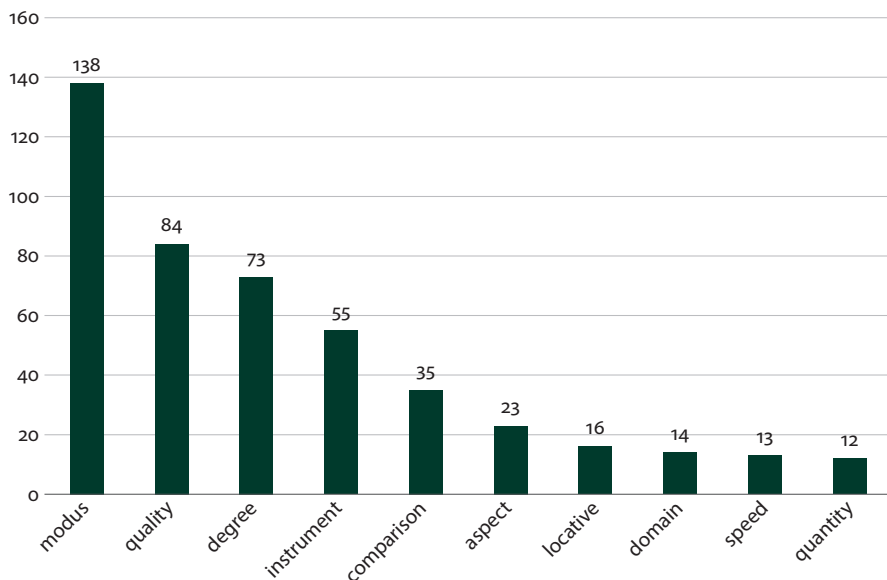
Figure 3. Manner values

We will show in § 6.2 through which discursive mechanisms the contiguous meanings are forced to be reinterpreted as manner markers.

6.   Co-constructing manner in and through discourse

The analysis of spoken interactional data not only revealed a surprising variety in the inventory of strategies to express manner *in* discourse, but it also allowed us to observe how the speakers construct and co-construct the description of manner in and *through* discourse.

We basically identified four discursive strategies the speakers use to (co)-construct manner through discourse and interaction: (i) progressive construction of manner meanings; (ii) negotiation of manner meanings; (iii) indexical construction of manner meanings; (iv) pragmaticalized use of (graduated) manner meanings.

6.1   Progressive construction of manner meanings

In order to observe the progressive construction of manner meanings through discourse we examine Example (55), reproduced below as (56):

(56)   [[*ta ta ta ta*]$_{m1}$ [*in serie*]$_{m2}$ [*ta ta ta*]$_{m3}$ [*senza l'agenda davanti*]$_{m4}$]$_{m5}$ *quello è lavorare*
'Ta ta ta ta, serially, ta ta ta, without an organizer before one's eyes, that's working!'

In this example, the speaker progressively constructs the description of the manner meaning he intends to express, i.e. 'very efficiently', by piling up in a list structure a number of constituents that contribute to the expression of the global meaning. In our view, the global construction made up of repetitions of constituents has a double function: it facilitates the lexical search for the target manner meaning but it also makes it possible to iconically represent the rapid succession of identical actions that characterizes efficient work. Another interesting example of this strategy is exemplified in (57):

(57)   *vivono_ eh* [[*in modo ancora un po' primitivo*]$_{m1}$ [*con* [*le slitte*]$_{m2}$ [*i cani*]$_{m3}$]$_{m4}$]$_{m5}$, *non sono moderni*
'They still live in a rather primitive way, with sleds, dogs, they are not modernised'

In this example, the speaker first utters a manner adjunct *in modo ancora un po' primitivo* 'in a rather primitive way' which is attenuated by the quantifier *un po'* 'a little'. The use of the attenuation seems aimed not at imposing but simply proposing a formulation that could be perceived as excessively blunt by the hearer. The speaker reformulates his first formulation by adding to it, in one and the same list structure, an instrumental adjunct *con le slitte* 'with sleds' and a comitative adjunct *(con) i*

cani 'with the dogs'. Occurring in a list with a manner adjunct, these two further adjuncts are reinterpreted quite easily as manner markers – shedding light, by the way, on the mechanisms that allow for the reinterpretation of the semantic focus of contiguous categories as manner markers that we described in § 2.3 and § 5.2. These two reformulative adjuncts are used to refine and justify the first formulation: their addition contributes to the progressive construction of a global manner meaning – that we indicated as m_5 – whose negative connotation is attenuated as compared to its initial formulation.

6.2 Negotiation of manner meanings

In the previous example we observed an attempt by the speaker to anticipate a reaction from the hearer that led him to reformulate and attenuate his initial formulation. We found examples in our corpus where the speakers explicitly negotiate the formulation of the manner meaning before introducing it in the common ground of a shared representation of reality. Example (58) is a case in point:

(58) A: *lo fa* [[[*da non residente*]$_{m1}$ [*come lo fareste adesso voi*] $_{m2}$ [*da non re<sidente>*] $_{m3}$] $_{m4}$
    B: [*come lo faremmo noi*] $_{m5}$
    C: [*residente*] $_{m6}$]]$_{m7}$
    A: *certo*
    'A: He does that as a non-resident, just as you would do'
    B: Ah! Just as we would
    C: Resident
    A: Sure'

In (58), speaker A proposes an initial formulation for the manner meaning through the adjunct [*da non residente*]m_1 'as a non-resident', he adds in the list structure a second adjunct [*come lo fareste adesso voi*] m_2 'just as you would do' – with probably a clarification intent – and then he repeats the first formulation [*da non re<sidente>*] m_3 'as a non-resident'. Globally he constructs, through a listing mechanism similar to the mechanism described above, a manner construction whose meaning has been made clearer to the hearers. Interestingly, the other participants in the discourse ask for confirmation for the meaning by repeating in the same list structure two of the adjuncts uttered by A, namely [*come lo faremmo noi*] m_5 'just as we would do' – uttered by B – and [*residente*] m_6 'resident' – uttered by C. All in all the negotiation of this manner meaning is expressed by the overall construction that we indicated with m_7. M_7 is realized by the syntactic list in which the occurrences of the adjuncts uttered by the three speakers pile up. Once this global construction

has been introduced in discourse, A confirms that all the interlocutors have a shared representation of the manner meaning by saying *certo* 'sure' and it can be assumed that that representation of manner becomes part of the common ground.

6.3   Indexical construction of manner meanings

Let us now examine Example (59). In this example we observe the extemporaneous construction of a manner meaning through an onomatopoeia reproducing the sound of suction (as can be heard in VOLIP MA15 10′34–38″)[6]:

   (59)   *con un phon le asciugava in modo vrouum*
         'with a hair dryer she dried it in a way that vrouum'

In this excerpt, a speaker is talking about the way in which the anchor woman Raffaella Carrà dried her pantyhose. An instrumental adjunct (*col phon* 'with a hair dryer') precedes the scope (*asciugare* 'to dry') and is followed by a manner marker realized by the onomatopoeia *vrouum*. This manner expression is totally anchored and is only interpretable in the speech context. Indeed, due to the absence of the gesture that – we hypothesize – accompanied the sound made by the speaker in the interaction, the meaning of this expression – besides not being transcribable – is not clearly interpretable.

     Slobin (2004) had already observed the gestural depiction of manner and he remarked that, due to lack of co-speech data, one cannot detect whether *VF*-language speakers provide gestural manner information at comparable rates to the lexical manner information provided by *SF*-language speakers.

     Let us add that, in interaction, gestures and sounds can be employed by speakers not only to accompany the expression of linguistic markers of manner, but also to build in context *ad hoc* manner meanings, which are not encoded in the linguistic system otherwise.

6.4   Pragmaticalized use of (graduated) manner meanings

We saw in (57) an example of graduation of the manner marker used to attenuate the evaluation it carried. In our corpus we identified 41 occurrences of graduated manner markers, i.e. manner markers occurring in the scope of a quantifier such as *un po'* 'little', *molto* 'a lot', but also temporal markers used metaphorically as quantifiers: *un attimo* 'an instant', *un momentino* 'a little moment'. This fact was quite unexpected since, as discussed in § 2.1, the traditional literature on manner

---

6. <http://www.parlaritaliano.it/index.php/it/visualizza-corpus?path=/Milano/MA15>

conducted on the lexical semantics of verb roots ruled out the possibility for manner to be considered a gradable category. However, in some recent studies, Amiot and Stosic (2011) and Stosic and Amiot (2014) showed that motion verbs can be modified by evaluative suffixes, in order to express manner values.

When observing the construction of manner in interaction, it is quite clear that the graduation and in particular the attenuation of manner markers is widely used with a general pragmatic objective, i.e., attenuating the illocutionary force of the utterance in which the marker occurs. Let us observe for Example (60), (61) and (62) – which reproduce (21), (22) and (23):

(60) *vi sto facendo stare* [[*un po'*]$_{m1}$ [*in allegria*]$_{m2}$]$_{m3}$
       'I'm keeping you a little cheerful'

(61) *è una cosa che va fatta* [[*un attimo*]$_{m1}$ [*con ordine*]$_{m2}$]$_{m3}$
       'It's something we need to do in a rather orderly fashion'

(62) *bisogna anche* [*un momentino*]$_{m1}$ *adeguare le foto*
       'You also need to adjust the photos a little bit'

In these examples, the quantifiers *un po'* 'a little', *un attimo* 'an instant' and *un momentino* 'a little moment' are used to attenuate not only the description of the manner meaning but also the injunctive illocutionary force of the utterances. Interestingly, we identified cases such as (62) in which the quantifier *un momentino* emancipates from the manner construction and scopes directly on the verb (rather than on the manner adjuncts) taking on a general manner meaning roughly corresponding to 'in a loose way'.[7]

7. Conclusion

We have presented in this article a corpus-driven analysis of the category of manner as it is expressed in and through interactional spoken discourse in Italian. Building on previous studies, we provided a rigorous functional definition of manner and a very abstract formal definition of manner constructions. Using these definitions we were able to identify a large number of manner constructions in our corpus without prejudging their morphosyntactic realization. The analysis of our corpus revealed two important facts: (i) manner is expressed through a large variety of

---

7. Voghera (2017b: 392–393) already pointed out for *un attimo / un attimino* a shift from the pragmatic strategy of attenuation to the expression of manner features: these markers indicate not only that the action is of short duration but also the relaxed and informal style in which the action is performed. Russo (in preparation) proposes to call this manner value CARELESSNESS.

constructions that can be identified at any level of linguistic structure, including discourse; (ii) manner is constructed online by speakers and it is often co-constructed through discourse.

The latter finding is particularly interesting and it deserves some commentary. We had already observed in previous work conducted on pragmatic categories (such as epistemicity in Pietrandrea 2018), that the participants in a conversation use a large inventory of discursive structures in making the effort to reach a shared opinion about reality. Quite interestingly the present analysis of manner – which is a descriptive, rather than a pragmatic category – shows that speakers deploy the same effort and the same discursive strategies to reach an agreement about the depiction of reality.

We can draw at least three orders of conclusion from our analysis: (i) manner is only a case study: it is entirely possible that other categories are co-constructed in and through discourse as manner is; (ii) the co-construction of the depiction of reality can be regarded as the confirmation of the inherently communicative nature of human language; (iii) the recurrence in interactional discourse of high level interactional discursive structures leads us to think that these abstract structures are likely to be encoded at some level of linguistic knowledge and that they deserve a thorough analysis in the general framework of a grammar of performance.

References

Amiot, Dany & Stosic, Dejan. 2011. *Sautiller, voleter, dansoter*: Évaluation, pluriactionnalité, aspect. In *Temps, aspect et classes de mots: Études théoriques et didactiques*, Eugenia Arjoca-Ieremia, Cécile Avezard-Roger, Jan Goes, Estelle Moline & Adina Tihu (eds), 277–297. Arras: Artois Presses Université. https://doi.org/10.4000/books.apu.7601

Beavers, John & Koontz-Garboden, Andrew. 2012. Manner and result in the roots of verbal meaning. *Linguistic Inquiry* 43(3): 331–369. https://doi.org/10.1162/LING_a_00093

Blanche-Benveniste, Claire. 1990. Un modèle d'analyse syntaxique "en grilles" pour les productions orales. *Anuario de Psicologia* 47: 11–28.

Cardini, Filippo-Enrico. 2008. Manner of motion saliency: An inquiry into Italian. *Cognitive Linguistics* 19(4): 533–570. https://doi.org/10.1515/COGL.2008.021

De Mauro, Tullio, Mancini, Federico, Vedovelli, Massimo & Voghera, Miriam. 1993. *Lessico di frequenza dell'italiano parlato*. Milano: Etaslibri.

Franco, Ludovico & Lorusso, Paolo. 2020. Verbi di maniera e predicazione elementare in una prospettiva interlinguistica. In *La semantica dello Spazio, del Movimento e della Maniera*, Luisa Brucale, Luisa Corona & Claudio Iacobini (eds). Special issue of *Testi e Linguaggi* 14: 242–258.

Gerdes, Kim & Kahane, Sylvain. 2006. L'amas verbal au cœur d'une modélisation topologique de l'ordre des mots. In *Ordre des mots et topologie de la phrase française*, Kim Gerdes & Claude Müller (eds). Special issue of *Linguisticae Investigationes* 29(1): 75–89. https://doi.org/10.1075/li.29.1.07ger

Goldberg, Adele E. 2010. Verbs, frames and constructions. In *Syntax, Lexical Semantics, and Event Structure*, Malka Rappaport Hovav, Edit Doron & Ivy Sichel (eds), 39–58. Oxford: OUP. https://doi.org/10.1093/acprof:oso/9780199544325.003.0003

Haspelmath, Martin. 1997. *Indefinite Pronouns* [Oxford Studies in Typology and Linguistic Theory]. Oxford: Clarendon Press.

Iacobini, Claudio. 2010. The number and use of manner verbs as a cue for typological change in the strategies of motion events encoding. In *Space in Language, Proceedings of the Pisa International Conference*, Giovanna Marotta, Alessandro Lenci, Linda Meini & Francesco Rovai (eds), 495–514. Pisa: ETS.

Ibarretxe-Antuñano, Iraide. 2006. *Sound Symbolism and Motion in Basque*. Munich: Lincom.

Jackendoff, Ray. 1983. *Semantics and Cognition*. Cambridge MA: The MIT Press.

Kahane, Sylvain, Pietrandrea, Paola & Gerdes, Kim. 2019. The annotation of list structures. In *Rhapsodie. A Prosodic and Syntactic Treebank for Spoken French* [Studies in Corpus Linguistics 89], Anne Lacheret-Dujour, Sylvain Kahane & Paola Pietrandrea (eds), 69–95. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.89.06kah

Levin, Beth & Rappaport Hovav, Malka. 1998. Building verb meanings. In *The Projection of Arguments: Lexical and Compositional Factors*, Miriam Butt & Wilhelm Geuder (eds), 97–134. Stanford CA: CSLI.

Levin, Beth & Rappaport Hovav, Malka. 2013. Lexicalized meaning and manner/result complementarity. In *Subatomic Semantics of Event Predicates*, Boban Arsenijević, Berit Gehrke & Rafael Marín (eds), 49–70. Dordrecht: Springer. https://doi.org/10.1007/978-94-007-5983-1_3

Masini, Francesca & Mauri, Caterina. 2020. Questione di *stile*. L'espressione analitica della maniera indessicale. In *La semantica dello Spazio, del Movimento e della Maniera*, Luisa Brucale, Luisa Corona & Claudio Iacobini (eds). Special issue of *Testi e Linguaggi* 14: 259–271.

Masini, Francesca, Mauri, Caterina & Pietrandrea, Paola. 2018. Lists: description, delimitation, definition. Special issue of *Italian Journal of Linguistics* 30(1): 41–48.

Moline, Estelle & Stosic, Dejan. 2016. *L'expression de la manière en français*. Paris: Editions Ophrys.

Pietrandrea, Paola. 2018. Epistemicity at work. A corpus study on Italian dialogues. *Journal of Pragmatics* 128: 171–191. https://doi.org/10.1016/j.pragma.2017.10.006

Rappaport Hovav, Malka & Levin, Beth. 2010. Reflections on manner/result complementarity. In *Syntax, Lexical Semantics, and Event Structure*, Edit Doron, Malka Rappaport Hovav & Ivy Sichel (eds), 21–38. Oxford: OUP. https://doi.org/10.1093/acprof:oso/9780199544325.003.0002

Russo, Gina. 2021. Percorsi di grammaticalizzazione nelle costruzioni di Maniera in Italiano. PhD dissertation, University of Salerno.

Sechehaye, Albert. 1926. *Essai sur la structure logique de la phrase*. Paris: Champion.

Slobin, Dan Isaac. 2004. The many ways to search for a frog. Linguistic typology and the expression of motion events. In *Relating Events in Narrative, Vol. 2: Typological and Contextual Perspectives*, Sven Strömqvist & Ludo Verhoeven (eds), 219–257. Mahwah NJ: Lawrence Erlbaum Associates.

Slobin, Dan Isaac. 2005. Relating narrative events in translation. In *Perspectives on Language and Language Development: Essays in Honor of Ruth A. Berman*, Dorit Ravid & Hava Bat-Zeev Shyldkrot (eds), 115–129. Dordrecht: Kluwer. https://doi.org/10.1007/1-4020-7911-7_10

Slobin, Dan Isaac. 2006. What makes manner of motion salient? Explorations in linguistic typology, discourse, and cognition. In *Space in Languages: Linguistic Systems and Cognitive Categories* [Typological Studies in Language 66], Maya Hickmann & Stéphane Robert (eds), 59–81. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.66.05slo

Stosic, Dejan. 2019. Manner as a cluster concept: What does lexical coding of manner of motion tell us about manner? In *The Semantics of Dynamic Space in French. Descriptive, Experimental and Formal Studies on Motion Expression* [Human Cognitive Processing 66], Michel Aurnague & Dejan Stosic (eds), 141–177. Amsterdam: John Benjamins. https://doi.org/10.1075/hcp.66.04sto

Stosic, Dejan. 2020. Defining the concept of manner: An attempt to order chaos. In *La semantica dello spazio, del movimento e della maniera*, Luisa Brucale, Luisa Corona & Claudio Iacobini (eds). Special issue of *Testi e Linguaggi* 14: 127–150.

Stosic, Dejan & Amiot, Dany. 2014. When evaluative morphology, pluractionality and aspect get tangled up: A case study of French suffixed verbs. In *Μελέτες αφιερωμένες στην Ομότιμη Καθηγήτρια Α.Π.Θ. Άννα Αναστασιάδη-Συμεωνίδη* – Mélanges offerts à Anna Anastassiades-Syméonides à l'occasion de sa retraite, Zoe Gavriilidou & Anthi Revithiadou (eds), 16–33. Kavala: Editions Saita.

Tommaseo, Nicolò & Bellini, Bernardo. 1861–1869. *Dizionario della lingua italiana*. Torino: Unione tipografico-editrice.

Talmy, Leonard. 1972. Semantic Structures in English and Atsugewi. PhD dissertation, University of California, Berkeley.

Talmy, Leonard. 1985. Lexicalization patterns: Semantic structure in lexical forms, In *Language Typology and Syntactic Description, Vol. 3: Grammatical Categories and the Lexicon*, 1st edn, Timothy Shopen (ed.), 57–149. Cambridge: CUP.

Talmy, Leonard. 1991. Path to realization: A typology of event conflation. In *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society*, Laura A. Sutton (ed.), 480–519. Berkeley CA: BLS. https://doi.org/10.3765/bls.v17i0.1620

Talmy, Leonard. 2000. *Toward a Cognitive Semantics: Typology and Process in Concept Structuring*, Vol. 2. Cambridge MA: The MIT Press.

Voghera, Miriam, Iacobini, Claudio, Savy, Renata, Cutugno, Francesco, De Rosa, Aurelio & Alfano, Iolanda. 2014. VoLIP: A searchable Italian spoken corpus. In *Complex Visibles Out There, Proceedings of the Olomouc Linguistics Colloquium: Language Use and Linguistic Structure*, Ludmila Veselovská, Markéta Janebová (eds), 628–640. Olomouc: Palacký University.

Voghera, Miriam. 2017a. La nascita delle costruzioni non nominali di *specie, genere, sorta e tipo*: Uno studio basato su corpora. In *Per la storia della formazione delle parole in italiano. Un nuovo corpus in rete (MIDIA) e nuove prospettive di studio*, Paolo D'Achille & Maria Grossmann (eds), 277–307. Firenze: Cesati.

Voghera, Miriam. 2017b. Quando vaghezza e focus entrano in contatto: Il caso di *un attimo*, anzi *un attimino*. In *Di tutti i colori. Studi linguistici per Maria Grossmann*, Roberta D'Alessandro, Gabriele Iannàccaro, Diana Passino & Anna M. Thornton (eds), 385–398. Utrecht: Utrecht University Repository.

Voghera, Miriam. In preparation. Vagueness expressions and approximate categorization: From comparison to focus?

Why it's hard to construct ad hoc number concepts

Mira Ariel

Lexical meanings are routinely adjusted in order to evoke ad hoc concepts. But number words pose a unique challenge. Carston (2002) discusses two relevant interpretative processes in this connection. Broadening (as in metaphorical uses) introduces an ad hoc concept by incorporating interpretations that fall outside the lexeme's linguistic meaning, and narrowing (e.g., interpreting *finger* as 'index finger') restricts the lexical meaning to a subset of its senses. I here argue that number words impose restrictions on the construction of such ad hoc concepts: (i) (nonround) number words cannot undergo narrowing, and (ii) when broadened (interpreted as 'approximately N'), number words typically require explicit marking (e.g., *about N*). Both restrictions stem from a single fact: Number words lack a prototype category structure (Lakoff 1972; Rosch & Mervis 1975). I support these claims with corpus analyses (The Santa Barbara Corpus of Spoken American English, The Longman Corpus of Spoken American English and the British National Corpus).

Keywords: number words, prototype category structure, ad hoc categories, broadening, narrowing, adaptors, rounders

1. Introduction

Lexical items, especially monomorphemic ones, typically stand for relatively stable cultural concepts (Barsalou 1983 and onwards). Now, these come in a very large variety, as testified by the existence of many partial synonyms. But according to the linguistic under-determinacy thesis, despite this rich variety there is always a "shortage" of lexical items: there are many more concepts speakers wish to refer to than there are distinct lexical items (Carston 2002; Wilson & Carston 2007). In fact, most of these concepts are ad hoc, namely ones the speaker instructs the addressee to construct online, having no intention for these concepts to have any long-term effect. The result is that lexical items are routinely mobilized for evoking ad hoc concepts.

A number of cognitive processes inform these context-sensitive meaning adjustments. One interpretative procedure involves broadening, which calls for encroaching on the meaning of a neighboring expression. Broadened meanings allow us to interpret some lexical item as close to a different, but relatively similar competitor in the lexical domain. *Red* in 1a is broadened to include a red-like color. And compare the differently broadened interpretations of *nice* in the following examples, where *nice* is interpreted as close to 'kind' in (b, but as close to 'attractive' in (c):

(1) a. MILES: I mean that looks kinda like a Black person.
            HAROLD: With **red** eyes,   (SBC:002)
      b. SHARON: You're too **nice**,
                  you're too **nice**,
                  your kids are gonna take advantage of you.   (SBC: 004)
      c. Sister: I've got a big date, Francis… Do I look **nice**??
            Francis: Depends. How **nice** are you supposed to look?
                                                          (*LA Times,* cartoon, June 3, 2000).

While these interpretations testify to the broadening of the lexical meaning of *nice*, for example, a second type of contextual adjustments involves the opposite cognitive process, namely, narrowing. Thus, the interpretation of graded adjectives is in principle compatible with any degree to which the adjective holds. *Nice*, for example, can denote any degree of 'niceness'. But it needs to be calibrated (narrowed) to fit the specific context. For example, in 1(b) 'niceness' (qua 'kindness') needs to be narrowed down to 'standard kindness towards kids'. In (c), 'nice' (qua 'attractive') needs to be narrowed into 'nice/attractive enough to impress a big date'. Francis' uncooperative response (1c), where he pretends to not be able to narrow down 'nice/attractive', only testifies how under the radar such narrowings are normally. Of course, 1(c) is taken from a humorous cartoon, where Francis is being a "wise-guy".

But narrowing is not limited to conventionally graded meanings. It is relevant to any category that has multiple denotations, which lexical items certainly do. Rosch and Mervis (1975) have drawn attention to the crucial distinction between central and peripheral category members, where central members are much more salient to us. The feminist protest against the automatic interpretation (narrowing in this case) of so-called gender unmarked nouns (e.g., *man, person, lawyer* etc.) as specifically males is a case in point. But narrowing to the stereotype is not the only default preference in selecting specific salient member/s. Narrowing is in addition context dependent. In fact, context may override the prototypical, salient sense. Consider the narrowed interpretations of *table* in the following examples:

(2) a. MILES: … I'm sitting down at a **table**, 002
     b. SHARON: every day,
                          I'd have em put their heads on the **table** for ten minutes,
                          ((3 LINES OMITTED))
                          with the- with their eyes closed.   (SBC: 004)
     c. So if we could add in picking up the chairs and the **tables**   (LSAC)

While *table* can refer to any type of entity belonging to the category of 'table', Miles' ad hoc table concept is specifically 'a dining room table' (a), and Sharon, a teacher talking about her students, is referring to 'tables in her school', presumably 'school desks' (b). In the context of planning a large gathering for lunch (c), where tables (and chairs) are to be rented out from a company, 'tables' must be narrowed to a different subtype of tables, namely 'folding tables'. When living room furniture is discussed a low, 'coffee table' is most likely intended by the word, etc.

But the interpretative processes called upon for the construction of ad hoc concepts need not be either broadening or narrowing. They may actually combine broadening with narrowing (Carston 2002). Such is the case in 1(b), in fact, where *nice* needs to broaden to 'nice/attractive' on the one hand, and to narrow to 'nice/attractive enough to impress a date' on the other hand. The same is true for *nice* in 1(c).

A special type of broadening+narrowing combination received quite a bit of attention in the literature on hedging. Some hedging devices, dubbed adaptors by Prince et al. (1982), create ad hoc categories where nonmembers of the lexical category, which are nonetheless relevantly similar to category members, are construed as marginal members of the category. At the same time, central category members are excluded. The result is a redrawing of the category boundaries, where the ad hoc category constructed includes both originally *marginal* category members and close-by *non*members. Such construals typically require overt marking:

(3)   She seemed **kind of nice**.   (LSAC)

The person referred to in (3) is specifically said to *not* be prototypically nice on the one hand, but on the other hand, she is asserted to be marginally so, where 'marginally nice' also covers attributes conventionally outside the category of 'niceness', provided they are relevantly close to it.

Ad hoc concept constructions based on broadening, as in 1(a), on narrowing, as in (2), and on broadening+narrowing, as in 1(b&c) and (3), all follow naturally from the assumption that lexical meanings form prototype categories (Lakoff 1987; Langacker 1987). Prototype categories have an internal structure, where (among other things): (i) Members are not equal in that some are more prototypical (representative, salient) of the category than others, and (ii) the boundaries surrounding the category are not rigid, thus allowing for them to be blurred and re-drawn. When

applied to lexemes, the nonequality between various meanings and/or denotations of the lexeme means that some of them are more easily accessed by default. In the same vein, some interpretations (not necessarily central ones) may be more relevant than others in specific contexts. The differential status of category meanings thus accounts for narrowing processes. Broadening too is a natural consequence of a prototype category structure. The fact that the boundaries surrounding lexical meanings are not rigid allows for the incorporation of meanings which are similar to the concept (within the specific context), despite the fact that they are not part of the lexical meaning. Opting for broadening+narrowing processes, interlocutors take advantage of both features of prototype categories.

So far the discussion focused on standard lexical items. Where do number words fit in this picture? What ad hoc categories or concepts can they give rise to? Do number words too constitute prototype categories just like standard lexemes, in which case they should give rise to the same interpretative processes with the same ease? According to Quirk et al. (1985) number words constitute a class of their own, because they have both open-class and closed-class characteristics. Just like open-class lexical items they come in a very large variety, as compared to the relatively small variety of closed-class items. But the relation between the meanings of different number words is different from the relation between the meanings of open-class words. The definition of number words involves mutual exclusivity.[1]

Now, Wilson and Carston (2007) have convincingly argued that broadening, narrowing and the combined broadening+narrowing are all manifestations of a unified process of context-dependent meaning adjustments, all geared towards the derivation of an optimally relevant interpretation of the utterance. This unification makes absolute sense in terms of the analysis of ad hoc interpretations, but my point here is that not all lexical items make available all such ad hoc adjustments. I here argue that unlike most lexical items, number words cannot easily serve for the construction of ad hoc broadened concepts, namely ones encroaching on a neighbor's meaning (as *nice* does in (1) (§ 2). And they cannot be narrowed at all (§ 3.1), which is why they cannot also be broadened+narrowed (§ 3.2), so as to exclude prototypical members while construing nonmembers as peripheral category members. I support these claims with corpus data from the Santa Barbara Corpus of Spoken American English (SBC, Du Bois et al. 2003; Du Bois et al. 2000–2005), The Longman Corpus of Spoken American English (LSAC) and the British National Corpus (BNC). I argue my point based on English number words, but my claim is about the inherent nature of numerals, and should therefore apply to any systematic +1 counting system.

---

1.   In addition, while they often modify nouns, number words hardly ever function predicatively, and they may occupy a different syntactic position from typical, adjectival noun modifiers

The reason for this special patterning of the number words, I propose, is that the meanings of number words do not constitute prototype categories. First, as mentioned above, in order to constitute such a structured category the meaning must denote a set of central and peripheral instantiations (members). But the number words, I claim, do not have multiple denotations. Each number word only denotes one meaning – a single member – 'N'.[2] Narrowing, which presupposes a multiplicity of members, is not an option, then. For the same reason, number words cannot undergo broadening+ narrowing processes, where again narrowing is (also) required. Finally, broadening entails the blurring of the boundaries surrounding the word's meaning. This is a challenge with number words, because unlike prototype categories, number word meanings have rigid boundaries. Hence, broadening processes are far more constrained for number words than they are for open-class lexemes. In order to create broadened ad hoc number concepts, where the speaker includes neighboring values of the number word as well ('approximately N'), *overt marking* is typically used with the number word.

A caveat is here called for. This paper focuses on nonround numbers, or rather on nonround interpretations of all number words (round ones included). I address round numbers in Ariel (in preparation). Briefly, I argue there that round numbers actually have a double membership. They may function just like open-class items, that is, as prototype categories, in which case they can give rise to broadened and narrowed ad hoc concepts, but at other times they function as nonround numbers, that is, they lack prototype category structure, in which case they cannot serve to build ad hoc narrowed concepts, and they need overt marking in order to broaden.

2.   Ad hoc construction of broadened (approximate) number concepts

According to Biber et al. (1999: 279) number expressions are most frequent in news and in academic writing, and least so in fiction. These differences are only natural. Precise quantities are often relevant in the more objective genres of news and academic writing, but not so in fiction. What about conversation? According to Biber et al., specifically, cardinal numbers are quite frequent in conversations (around 16,000 per million words), more frequent than quantifiers (just over 12,000 per million words). Since quantifiers mostly denote imprecise quantities (e.g., *some, most, a few*) one might think that the higher number of number words testifies to

---

2.   The classical Gricean analysis of the numbers assumed a lower-bound ('at least N') core meaning for number words, deriving their 'exactly N' meaning by adding on a Generalized Conversational Implicature of 'no more than N'. But more recent analyses (Ariel 2006; Breheny 2008; Bultinck 2005; Kadmon 2001) assume an 'exactly N' linguistic meaning for number words.

a significant role speakers attribute to precise quantities. But in fact, quite a few number words are mobilized to construct ad hoc approximate values.[3] Speakers opt for an approximate value for a number of reasons, one obvious motivation being a speaker's ignorance of the facts, which prevents her from committing to a precise number (Biber 1995). While an 'approximate N' value may not be as informative as a precise value ('N'), it is more specific (varies along a smaller range) than the quantity denoted by a quantifier.

Consider the following case of approximation:

(4) **Roughly round about** *nineteen twenty seven, twenty eight.* (BNC)

While (4) is an extreme case, it is not uncommon to find two approximators participating in the ad hoc construction of a quantity based on a number word:[4]

(5) REBECCA: I imagine it will be **about** .. *five* dollars **or so**. (SBC: 008)

The cognitive process which derives an approximate value based on a precise number word is an instance of the same broadening process that routinely applies to standard lexemes, such as *nice* above (Sperber & Wilson 2008). A redrawing of the boundaries surrounding the meaning of the lexeme, in this case 'N', creates an expanded range as an ad hoc category, which incorporates similar meanings, in this case neighboring numbers. As with standard lexemes, here too the process is governed by pragmatic considerations, which determine how large the derived range is, but this question will not be addressed in this paper (See Channell 1994: Chapter 3).

Here are three more examples, where only one boundary is to be blurred (the upper -boundary in a, the lower boundary in b and c):

(6) a. SHARON: HISD rules has it that you have to have **at least** *eighteen.*
                                                                         (SBC: 004)
     b. MONTOYO: but uh probably (H) **no more than** *two or three* if they
                        were lucky.   (SBC: 012)
     c. CONTEXT: The coach only admits girls six and up.
         Coach: How old is she?
         Mother: She's a **soft** *four.*   (Movie, "I, Tonya", 2018)

---

3. Solt et al. (2017) argue for a preference for approximation in discourse, and hence for a disproportionate number of round numbers. Our focus in this paper is on nonround numbers, which are also used as pivots to create ad hoc approximate concepts.

4. See Channell (1994: Chapter 3) for restrictions on such combinations.

Thus, speakers definitely have a use for ad hoc concepts built on broadened number words. The goal of the next section is to support my claim that since number word meanings do not constitute prototype categories, broadening them is not a trivial matter as it is with standard meanings, which do have a prototype category structure.

2.1 The crucial role of rounders

Note that the construction of ad hoc, broadened concepts based on the number word meanings in § 2 was triggered by explicit approximating expressions – rounders in Prince et al.'s (1982) terms. Rounders explicitly instruct the addressee to relax the boundaries surrounding the conventional lexical meaning, so that similar non-member meanings are included. This was not the strategy used with the standard lexemes briefly discussed in § 1, where there was no overt marking for boundary blurring. My claim is that the different strategies for broadening the items in § 1 and the ones in § 2 is not accidental. If the number words do not have a prototype category structure, their boundaries are not flexible, and they cannot be so easily abolished. If so, the speaker cannot simply count on the addressee to automatically relax the boundaries and incorporate neighboring values into the ad hoc constructed category. To ensure such broadening of the number word, the speaker is better off specifically instructing the addressee to do so. Hence the need for overt rounders. Tables 1 and 2 present the percentages of overt broadening expressions modifying four graded adjectives (Table 1) and five number words (Table 2):

Table 1. Rate of explicit broadening of *nice, blue, green* and *red**

Expression	Overtly broadened
Nice (143)	8 (5.6%)
Blue (137)	1 (0.7%)
Green (33)	1 (3%)
Red (40)	1 (2.5%)
Total (353)	11 (3.1%)

* Counts are based on SBC for *nice, green* and red, and SBC+LSAC for *blue*.

Now, compare the very low percentages of explicit broadeners in Table 1 (average 3.1%) with the much higher percentages observed for a few number words in Table 2 (average 24.7%):

Table 2. Rate of explicit broadening of *two, seven, ten, seventeen* and *thirty**

Number word	Overtly broadened
Two (89)	14 (15.5%)
Seven (66)	20 (30.3%)
Ten (64)	16 (25%)
Seventeen (56)	15 (26.8%)
Thirty (37)	12 (32.4%)
Total (312)	**77 (24.7%)**

* Counts are for every 5th hit of *two* in SBC, all SBC *sevens*, every other SBC *ten*, all SBC *seventeens* plus every 5th *seventeen* in LSAC, and all SBC *thirty*.

Still, the fact that Table 2 concerns number words is not the only difference between the marking (or lack thereof) of the broadened lexemes in the two tables. The lexical items in Table 1 have vague meanings, whereas the number words are precisely defined. The differential patterning could therefore be attributed to the difference in the linguistic meanings of the two types of expressions. It is only reasonable to assume that deriving an approximate concept out of a precisely defined lexical meaning should be more effortful. Indeed, Channell (1994: 52) notes that whereas *approximately* cannot modify vague quantities, such as *some*, it can modify the precisely defined *parallel*. However, Table 3 shows that nonnumber word lexemes which are precisely defined nonetheless do not pattern with the precise number words. Instead, their discourse profile with respect to broadening is quite similar to that of the vague terms, rather than to that of the precise number words:

Table 3. Explicitly broadening *middle straight* and *parallel**

Expression	Overtly broadened
Middle (110)	0
Straight (122)	1 (0.8%)
Parallel (15)	0
Parallel (100)	2 (2%)
Total (347)	**3 (0.9%)**

* Counts for *middle* are based on all SBC tokens and every 9th LSAC token. Counts for *straight* are based on all SBC hits and every 5th LSAC hit. Counts for *parallel* are based on all SBC+LSAC tokens and a sample of 100 BNC occurrences.

All in all, based on the lexemes I counted in Tables 1–3, it looks like on average, explicitly marking the broadening process is more than 10 times more frequent for the number words than for standard lexemes (24.7% versus 2.2%).[5]

---

5. 2.2% is the average calculated on the basis of Tables 1 and 3 taken together.

Now, the findings in Table 3 would actually be meaningless if it so happens that e.g., *middle* and *straight* are hardly ever explicitly hedged just because they need not be broadened in fact, that is, if their contextually appropriate interpretation is as 'exactly middle/straight'. The findings in Table 4 show otherwise:

Table 4. 'Exactly' versus 'broadened' interpretations for *middle* and *straight**

Expression Interpretation	Middle	Straight	Total
'Exactly'	3 (8.3%)	4 (13.8%)	7 (10.8%)
Broadened: 'Approximately'	28 (77.8%)	12 (41.4%)	**40 (61.5%)**
Broadened: Metaphorical	0	10 (34.5%)	**10 (15.4%)**
Unclear	5 (13.9%)	3 (10.3%)	8 (12.3%)

* Counts here are based on all SBC tokens of both expressions.

Despite the fact that more than 75% of the tokens of both *middle* and *straight* are contextually broadened, virtually no broadening marking was used in order to explicitly point the addressee to the required blurring of the boundaries. The speaker can count on the addressee to contextually broaden even precise meanings, provided they constitute prototype categories.

If I am correct, and unlike standard lexical items, number words require explicit rounders in order to broaden, we should witness a skewing in the propensity of approximators to modify number words. In fact we do. General rounders are often disproportionately used for broadening number meanings, and other linguistic means are actually specialized for broadening number meanings.

2.2 Skewed broadeners for number words

Comparing a few standard lexemes and a few number words in § 2.1, we saw that number words are far more likely to be overtly broadened than other terms. In § 2.2 we approach the question from the other direction, examining a few broadening expressions and noting what it is that they tend to broaden. The picture that emerges is quite similar. It seems that many broadeners have a marked preference for number words, co-occurring with number words much more than expected, given that there are approximately 1.6 number words per 100 words in natural conversations according to Biber.

Prince et al.'s first category of hedging markers is "Shields", where the expression lowers the speaker's overall commitment to the truth of her proposition. Here's such a case:

(7) LYNNE: .. And the first part of it,
                   .. is **like**,
        (H)   well we have lecture,
                   .. then we have la=b.
        (H)   .. And **like**,
                   the first part of it,
                   .. we just go into the classroom=m.   (SBC: 001)

However, once a shield has phrasal rather than sentential scope it may not only lower the speaker's commitment to the category, it may in addition blur its boundaries, and broaden the category, thus creating an ad hoc broadened concept, as in (8), which could either mean 'Wednesday, but with a lowered certainty', or 'around Wednesday':

(8) PATRICK: You gotta .. point her out .. **like** Wednesday.   (SBC: 045)

Here is the distribution of every 5th token of *like* in SBC:

Table 5. *Like* in SBC

Irrelevant ('similar', for 'example' 'say' etc.)	Sentential hedge	Phrasal hedge		Total
		Number words	**Nonnumber word**	
259	82	**30**	77	448

Excluding irrelevant *like* tokens, almost half of the hedging *like*s have sentential scope (43.4%). Interestingly, however, when phrasal hedging *like* is considered (107 tokens), as many as 28% of them (30) are *like N* cases. These broadened number words no doubt outnumber their corpus frequency.

　　Give or take shows an even higher proportion of broadened number words. There were no *give or take* tokens in SBC, and only 1 in LSAC (which indeed, modified a number word), so I turned to the BNC. 25/57 BNC tokens (43.9%) were used to specifically broaden number words. Now one would think that *more or less* would similarly specialize for number words, but that's not at all the case. Of the 49 hedging *more or less* in SBC + LSAC, only 3 tokens modified number words, as in:

(9) I drank a lot I drank like eighteen beers about **more or less** fifteen to eighteen beers   (LSAC)

The next table shows the distribution of *about, more or less*, and *roughly*, each in a sample of 100 BNC tokens:

　　Broadeners clearly vary as to how frequently they modify number words. *About* definitely leads, with 19/26 of its hedging uses (73%) reserved for number words

Table 6. A sample of 100 instances of BNC *about, around, more or less* and *roughly*

Expression/% approximator or hedge	*About*	*More or less*	*Roughly*	Total
+Hedge −Number word	7	**91**	**40**	138
+Hedge +Number word	**19**	0	**44**	63
−Hedge	73	9	16	98
Unclear	1	0	0	1

(and see § 3.2 below). More than half of hedging *roughly* tokens modify a number word (44/84, 52.4%). Again, *more or less* is not at all frequent with number words (here no co-occurrences at all).

Note that just like *give or take* and *more or less* show remarkably different patterns with respect to broadening number words, so do *X or something* and *X or something like that*. There are 49 *X or something* cases where X is hedged in SBC. Only 2 of them introduce a number word as X. But out of the 25 hedged *X or something like that* in SBC, almost a third were cases with a number word X (8). *Or so* is heavily skewed towards broadening number words:[6]

Table 7. The distribution of *or so*

Expression	Broadened by *or so*
Nonnumber word	80
Number word	108
Irrelevant	5
Unclear	2
Total	**195**

From X to Y and *between X and Y* may create approximate value ranges. Once again, number words are disproportionately represented in such constructions. A search for every 11th *from* [1–5 words] *to* yielded 104 hits in SBC. Of these, 24 created ranges, and in 12 (50%) of them X and Y were number words. A similar pattern emerged for every 5th hit of *between* [1–5 words] *and* Y. It yielded 99 hits where Y was a constructional argument. 51 of these created a range, and again, in half of them (26) X and Y were number words. The list could go on to single boundary-breakers such as *at least, at most* and many others. Here's a relatively innovative upper-boundary broadener (and see again the use of *soft* in 6c):

---

6. Data are based on all *or so* tokens in SBC and LSAC. Note that NPs such as *fifth grade*, a couple of weeks, *a year, last month* were not counted as numeric determiner phrases, because these do have a prototype category structure (see Ariel in preparation).

(10)   There will be 70+ exhibitions across Southern California.

(LACMA flier, September, 2017)

In addition to these hedges which are surprisingly often used to broaden specifically number words into ad hoc approximate values languages have quite a few numeral-special broadening constructions. *Odd N, some N, N something, N or N+1,*[7] *N N+1*[8] are such cases. It is worth noting that indeed an ad hoc category is constructed when speakers use these constructions. For example, in a questionnaire I ran subjects were asked to choose the more appropriate response (B_1 or B_2) to an utterance containing an *N or N+1* construction, such as (11) (see Ariel 2015 for information about the questionnaire):

(11)   A: Can you imagine! Linda had to walk **four or five** miles to get to school every day!
         B_1:   Wow, poor thing!
         B_2:   So, which is it, four or five?

B_1 constructs a single, approximate, higher-level category out of *four or five*. B_2 treats *four* and *five* as two separate alternatives such that a choice between them may be relevant. As it happens, 57/60 participants chose B_1.

The bottom line is that discourse findings show that number words are used remarkably often as mere pivots for creating broader ad hoc values–almost a quarter of their uses on average according to Table 2. This explains the skewed findings in 2.2, where many broadeners modified an unexpectedly high proportion of number words.

3.   No narrowing for number concepts

While broadening the meaning of number words is a marked interpretative step, which is then formally marked, my claim is that narrowing is simply unavailable for number words. Narrowing the denotation of some meaning presupposes that the meaning is compatible with multiple denotations. This is indeed true for meanings structured as prototype categories. But my point is that number words lack this structure, which is why they cannot give rise to ad hoc narrowed (3.1) nor broadened+narrowed (3.2) concepts.

---

7.   The subsequent number word in the disjunctive construction need not be exactly +1, but it must be an immediately successive number word in *some appropriate granularity* (we count by ones, pairs, fives, tens, hundreds, etc.). The reading associated with such constructions is the Higher-level category (Ariel 2015; Ariel & Mauri 2018).

8.   See *1927*, 28 in (4).

3.1   No narrowed ad hoc concepts for number words

Narrowers are explicit expressions that single out a subset of the members of a meaning category. Horn (2018) discusses a few such constructions, the discourse function of which is to coerce a partition between central and peripheral category members, and to focus on the central ones, while excluding noncentral ones. Such is the Clone construction where a NP (mostly) is reduplicated in order to instantiate only prototypical instances, as in:

(12)   We have muffins, and we have DESSERT desserts.   (Horn 2018: Example 1d)

Horn also cites Austin (1963) for an analysis of *a real X* construction as distinguishing between "real" (basically, prototypical) category members and marginal members. Now, if I'm right, and the number words lack a prototype category structure, specifically because they do not instantiate more than a single member, a number word should not be able to take part in constructions that impose a distinction between central vs. peripheral category membership. This is indeed the case.

But first, note that precisely defined meanings such as *flat*, may participate in the Clone construction. A flat tire is discussed in (13), and the question is aimed at ruling out the option of a nonprototypically flat tire ('just real low'):

(13)   Do you know if it's **flat, flat** or if it's just like real low?   (LSAC)

Let's now see how number words fare in such contractions. I searched SBC and LSAC for reduplicated *one* (129 cases), *three* (5 cases), and *ten* (16 cases). None of them carried the Clone function. Most were repetitions, or else cases where the same number word needed to be mentioned twice for different referential roles (e.g., *one one hundredth* – LSAC).[9] Next, there are 202 *real* instances in SBC, but only 15 function as 'central category member' indicators, as in (14),[10] where Kendra doesn't consider the icing on her birthday cake to be a good representative of the icing category:

---

9.   As Horn notes, the Nominal Clone construction is also used to insist on a literal interpretation, when competing with a salient derived interpretation. Under such circumstances one can (rarely) find the construction with number words. For examples, years ago, following an invitation I made to a friend for a get together at 9 o'clock, he asked me *Is that nine NINE*? (Originally Hebrew, where adjectives follow the modified nominal, hence the different stress pattern). What my guest wanted to know was whether the invitation was for 9 sharp, rather than the usual, Israeli 9 *o'clock*, interpreted as 9: 15–20 minutes later. But in fact, *nine* here is not a standard counting number word, but rather, an instance of social gathering time, in which case it does have a prototype category structure (See Ariel, in preparation), where in fact, 9:15 or so is the prototypical member, so the cloning insists on the literal meaning.

10.   The majority are intensifiers, 'nonfictional' or the lexeme *real-estate*.

(14)   KENDRA: what I like is,
                 like **real** icing.   (SBC: 013)

No *real N* was found in SBC. In BNC, where I searched for both *real* and *really* modifying several number words (3, 5, 7, 9, 18, 20, 25, 100), I did find 8 such strings, but again, none carried the narrowing function.

Typical has a similar function, to narrow the denotation of the concept it modifies to typical instances of the category, excluding nontypical ones from consideration. Again, if number word meanings only have single denotations, no distinction between sub-types is possible, so the prediction is that they cannot be modified by *typical*. This was borne out by the data – the 89 tokens of *typical* in SBC +LSAC. Thus, even when a *typical N* string was attested (1/89), *typical* did not in fact modify the number word, but rather, the nominal it is part of, as in the following:

(15)   I had a **typical** *ten* speed bike.   (LSAC)

I also found only 3 instances of *typical N* in BNC (I searched for all number words between *one* and *ten*). All patterned like (15). *Par excellence* similarly modifies a (highly) prototypical category member (Lakoff 1972). *Par excellence* is not a frequent marker, but still, out of the 93 tokens in BNC none showed *par excellence* modifying a number word. A web search for *eight par excellence* revealed the following example, where the number word does not fall under the scope of *par excellence*:

(16)   This set of *eight* "**Par *Excellence***" LED lighting fixtures is just what you need".
                 <https://manualzz.com/doc/10822231/pro-svl-2014a-spring>

The same is true for *technically speaking*, which does not modify *eight* in (17):

(17)   **Technically speaking**, it's *8* cubic feet of interactive, three-dimensional space.
                 <http://books.Google.co.?id=jz61CAAAQBAJ>

Here's a constructed example, where it might seem that the number word is narrowed to exclude competing values (in this case 'three'):

(18)   ~ **Technically speaking** she has *six* children. But *three* of them were put up for adoption.[11]

In a context where women are entitled to a monthly child support provided they 'have 6 children', (5) may be used to deny the woman child support, because she only 'has three children'. But actually, what is narrowed here is the meaning of 'having 6 children' – to 'giving birth to 6 children', as opposed to 'giving birth to and

---

11.   ~ indicates an unattested example.

raising 6 children'. Indeed, this is why the relevant factual number can be 'three', which is not at all a marginal member of *six*, nor a close-enough approximation of it.

Any is another category adaptor, this time, one where noncentral members are explicitly included under the denotation (although central category members are not excluded in this case):

(19)   a.   LAJUAN: Don't carry **any** excess baggage.   (SBC: 044)
     b.   SCOTT: I wonder if X .. if it would be,
            … **any** less .. abstra=ct … than the –   (SBC: 034)

Any excess baggage also denotes noncentral members of 'excess baggage' (*any* non-prototypical amount) and *any less abstract* denotes noncentral degrees of 'less abstract' as well. But note that when *any* modifies a numeric determiner phrase, no adjustment to the denotation of the number word is made:

(20)   ALAN: … You pick out **any** *four* of these you want.   (SBC: 060)

Any picking of the relevant entities in (20) involves exactly four items.[12] This is quite different from the minimally different constructed (21), where *four* is approximated:

(21)   ~Pick out **about** *four* of these.

Thus, all the narrowing adaptors examined in 3.1 fail to modify number words, because, I claim, number words do not have a prototype category structure. This is why speakers cannot construct narrowed ad hoc concepts based on number words.

3.2   No sorta and kinda ad hoc number concepts

In 3.1 we have seen that number words cannot be part of narrowing constructions, because these distinguish between central and peripheral meanings for the purpose of focusing on the central ones. We here examine other adaptors, which also draw a distinction between central and peripheral meanings, but for a different purpose. First, while they too narrow the denotation of the category, they focus on *peripheral* category members, rather than on central members. Second, at the same time, they also expand the category boundaries, incorporating nonmembers, ones relevantly similar to category members, into the ad hoc category constructed. *Sort of/sorta* and *kind of/kinda* are prime examples of such category adaptors:

(22)   a.   PETE: but then it would be sitting there,
            sort of .. decomposing,
            or doing its thing all over the lawn,   (SBC: 003)
     b.   LYNNE: so it was **kinda** review for us.   (SBC: 001)

---

12. (20) is the only relevant example in SBC, which includes 287 *any* tokens.

An examination of *kind of* and *kinda* in SBC revealed that as predicted, none of these adaptors modified a number word:[13]

Table 8. The distribution of *kind of/kinda* in SBC

Use/ Expression	Irrelevant (literal)	Hedge (−Number word)	Hedge (+Number word)	Unclear	Total
Kind of	21	37	0	0	57
Kinda	7	48	0	1	56
Total	**28**	**85**	**0**	**1**	**113**

A search for a sample of 100 *kind of* in the BNC revealed a similar picture, except that the hedging function is not as prevalent in this corpus (18% vs. 75.2% in SBC):

Table 9. The distribution of *kind of* in BNC

Use/ Expression	Irrelevant (literal)	Hedge (−Number word)	Hedge (+Number word)	Doubtful hedges	Total
Kind of	76	**18**	**0**	6	100

Note that the 6 doubtful hedgers do not modify number words either. There were in addition 140 *kinda* tokens in BNC, all hedgers, but none of them hedged number words. *Sort of* and *sorta* are predominantly used for hedging in SBC, 88.3%, and again, none of these hedgers construct ad hoc categories based on number words:[14]

Table 10. The distribution of *sort of/sorta* in SBC

Use/ Expression	Irrelevant (literal)	Hedge (−Number word)	Hedge (+Number word)	Doubtful hedges	Total
Sort of/ sorta	7	**68**	**0**	2	77

An examination of a sample of 100 *sort of* occurrences in BNC revealed no number word hedging either:[15]

---

13. Counts are based on all *kinda* and every third *kind of* in SBC.

14. Counts are based on all *sort of* and *sorta* tokens in SBC. There were only 10 *sortas*, so Table 10 collapses the results for the two expressions.

15. There were only 12 *sorta* tokens in the BNC, all hedgers, none modifying number words.

Table 11. The distribution of *sort of/sorta* in BNC

Use/ Expression	Irrelevant (literal)	Hedge (−Number word)	Hedge (+Number word)	Unclear	Total
Sort of	45	**50**	**0**	5	100

Tables 8–11 show a significant difference between American and British English. *Kind of* and *sort of* are used as hedges much more frequently in American English. There is also a marked difference between the use of the two hedging markers in British English 18% of *kind of*) and (50% of *sort of*).[16] But both expressions in both dialects show a clear absence of number word modification. This is just what is predicted for the number words, which do not have the prerequisite prototype category structure, and hence do not admit a central versus peripheral partition between denotations.

However, contra my claim here about the incompatibility, and hence non-cooccurrence of adaptors with number words, Gries and David (2007), find that unlike *kind of, sort of* does cooccur with number words. Gries and David focus on different discourse patterns associated with *kind of* and *sort of* in British English, an intriguing finding, given that the two expressions are synonymous. Specifically, Gries and David's comparison between the nouns modified by the two hedges is pertinent to our discussion. This comparison revealed only one major difference between the two hedgers: there is an *extremely large* number of *sort of* tokens modifying quantified expressions and number words, they find. This looks like a serious counter-example to my claim, and stands in contradiction to the findings presented in Tables 9 and 11 above for BNC.

To assess the significance of Gries and David's findings, I examined a variety of adaptors and broadeners modifying number words in the BNC. I chose 10 number words: 2, 3, 5, 7, 9, 18, 20, 25, 35 and 100, which followed any one of the following adaptors: *about, around, more or less* as well as *kind of* and *sort of*. Table 8 presents the results:[17]

---

16. And see Biber (1999) for register differences in both British and American English. Predictably, conversations show a much higher frequency of *kind of* and *sort of*. Biber finds no difference in the frequency of British and American use of *sort of*.

17. The searches were conducted for both alphabetical and Arabic number representations of the number words. For 100, I searched for *100, a hundred* and *one hundred*.

Table 12. The frequency of various adaptors for 10 number words in BNC

Expression/Number	Kind of	Sort of	About	Around	More or less
2	0	23	1519	158	2
3	0	0	1388	164	1
5	0	14	1130	154	0
7	0	6	390	55	0
9	0	6	305	43	0
18	0	1	235	30	0
20	0	8	1122	154	1
25	0	1	242	46	0
35	0	1	98	37	0
100	0	0	577	141	0
Total	0	60	5778	985	4

The first striking finding corroborates the picture in § 2.2, in that the frequency of different number word modifiers is extremely variable. I found 6827 hedge + number word combinations for the ten number words examined, but one of them, *about*, is overwhelmingly dominant: 5778/6827, 84.6%. The lion's share of the remaining cases is taken up by *around* (14.4%). *More or less*, as already noted above, rarely modifies number words.

But what about the two adaptors? As predicted, there were no *kind of* + Number word in BNC. But I did find 60 *sort of* ₊ N strings. While the difference between *kind of* and *sort of* is obviously extremely significant, supporting Gries and David's point, it's still the case that *sort of* constitutes less than 1% of the hedgers for the 10 hedged number words. But although numerically quite negligible, 60 *sort of* + N cases constitute 60 potential counter-examples to my claim about the incompatibility between adaptors and number words. Upon closer examination, however, these cases proved unproblematic for my claim. Most of the cases are ones where *sort of* does immediately precede the number word, but the number word itself is part of a complex NP, and it is that whole NP, rather than the number word by itself, that is modified by *sort of*, as can be seen in the following:

(23)   Obviously being a volunteer is a **sort of** *two way thing*   (BNC)

In other cases, *sort of* modifies a number word which is independently rendered approximate, as in:[18]

(24)   Y'know **sort of** *five or ten minutes or something like that* maybe

---

18. Such constructions do have a prototype category structure – see Ariel (in preparation). In addition, *sort of* may actually have sentential scope here.

Other cases are most likely not number word hedgers, as in:

(25)   In the last ten months of my period there, I did **sort of** *thirty five thousand miles*
       and a job as well.   (BNC)

It is more likely that the speaker in (25) hedges the whole *I did* sentence than the
35000 miles per se. In other words, what the speaker is saying is that s/he *more or*
less managed to perform both tasks. There were only a few cases where *sort of* did
unequivocally modify a number word, as in:

(26)   Just leave that unless you are totally stuck, and I don't mean **sort of** *five minutes*
       have a good **sort of** *maybe ten minutes or say about ten minutes of playing with*
       this.   (BNC)

Now, I've been talking about *sort of* possibly hedging number words in these last
few examples, but note that the effect *sort of* has here is different from its effect as
an adaptor. *Sort of* in (22) builds ad hoc concepts of 'nonprototypical decomposing'
(a) and 'nonprototypical review' (b). B0ut this is not what *sort of* does in (24)-(26).
In these examples *sort of* functions as a broadener, and not as an adaptor. In other
words, it instructs the addressee to blur the boundaries around the number words
and create an range, a broadened concept. It doesn't exclude the prototypical mem-
ber. Thus, *sort of five minutes* and *sort of ten minutes* in (24) are interpreted as 'about
5/10 minutes', which does not rule out the possibility that it might be '5/10 minutes'.

   These *sort of* + N cases are not counter-examples then. My claim is that adaptors
qua adaptors cannot be used with number words, which means that either they are
totally blocked (as in the case of *kind of*), or, if they marginally do cooccur with
number words they are not adaptors. Rather they turn into broadeners, just because
no partition between central and peripheral members can be drawn within the
number word meaning. Note that *sort of* is not the only marker which is originally
an adaptor that can cooccur with number words. The same pattern can be observed
for the suffix *-ish*. As a nonlexicalized suffix, *-ish* is an adaptor, so that *leftish* in 27(a)
is interpreted as 'nonprototypical leftist' and the same for *grayish* in (b):

(27)   a.   KEN: (TSK) And lots of people like trade unions,
                        ((2 LINES OMITTED))
                        and political,
             LENORE: M[hm],
             KEN: [you know],
                        left –
                        .. [left**ish**],
             JOANNE: [Unhunh].
             KEN: or liberal types,   (SBC: 015)
       b.   Um, it's a gray**ish** color.   (LSAC)

But when it modifies a number word, *-ish* becomes an approximator, as in:

(28)   She was thirty**ish**.   (LSAC)

Thirtyish does not rule out 'thirty' (although it may implicate that it's not 'exactly thirty'). Finally, a WEB search revealed the following *kinda + N* example:

(29)   It was **kinda** *a hundred* meters away.

<http://www.ontheroadwithrobert.com/p5-rica.html>

But just like *sort of + N*, this rare *kinda a hundred* is broadened (approximated), rather than adapted to an ad hoc category incorporating only marginal members.

So it looks like number words cannot give rise to ad hoc concepts which combine peripheral members with nonmembers. *Sort of N* strings do not create broadened+narrowed concepts. To the extent that *sort of* truly modifies a number word, it only broadens it.

4.   Conclusions

Speakers often use lexical items as mere starting points for the construction of ad hoc concepts. Most likely, such processes routinely and possibly equally operate on all open-class lexical items.[19] Following Carston (2002), I have discussed three types of ad hoc concept building: Broadening, narrowing and broadening+narrowing. What is pertinent to the research here reported is that while both standard lexemes and number words are frequently used to evoke ad hoc concepts, (i) the nature of these ad hoc concepts and (ii) the means to trigger them are remarkably different for the two lexical types.

Setting out from the assumption that the meanings of standard open-class items have a prototype category structure, I made the prediction that they can rather easily broaden, since prototype categories have nonrigid boundaries. Next, since prototype categories have multiple instantiations I also predicted that they should be able to undergo narrowing. Finally, since, in addition, prototype category instantiations have nonequal status, such that some are more prototypical than others, speakers may point addressees to only a subset of the lexeme's denotations, either to the prototypical category members or to the nonprototypical members. They may also simultaneously take advantage of both prototype category features, namely (i) the partition into central or contextually salient versus peripheral members and (ii) the flexible boundaries, in order to broaden+narrow, i.e., construct

---

19. A quantitative study of this question is beyond the scope of this paper.

an ad hoc category which includes peripheral members (narrowing) and close-by nonmembers (broadening). Indeed, I briefly showed (§ 1) that nonnumber words are routinely broadened, narrowed, and broadened+narrowed even without the support of overt marking.

My main claim pertained to number words, however. I proposed that number words do not have a prototype category structure. They cannot be narrowed, then, because they do not have multiple denotations (§ 3.1), and they cannot be broadened+narrowed because in addition, they obviously do not distinguish between central versus peripheral denotations (§ 3.2). The only process they can undergo is broadening (approximation), but since number words have rigid boundaries, a more marked strategy, explicit marking, is strongly preferred when the speaker wishes to blur the number word boundaries to achieve an approximate concept. Indeed, a comparison between standard lexemes and number words (§ 2.1) showed that many more number words than standard lexemes are modified by hedging or adapting expressions (more than ten times as much). Finally, the assumption of relatively rigid category boundaries for the number words can explain the well-known fact that number word approximation can be achieved by a variety of linguistic means, many of them either specialized for broadening number words or strongly skewed towards broadening number words (§ 2.2).

We are left with one question, however. Why are number words the way they are? Why do these lexical items lack a prototype category structure, when probably all other lexemes do, and even precisely defined lexemes (e.g., *straight, half*) have a prototype category structure? In Ariel (in preparation) I argue that the uniqueness of number expressions derives from the fact that unlike standard open-class lexemes, they belong to a dense lexical field.[20] This is how I account for the dominant nonbroadened 'precisely N' interpretation of bare number words, for the formal markedness of most number words interpreted as 'approximately N', as well as for the remarkably slow rate of change and renewal for number words (Pagel & Meade 2017).

Funding

---

20. A dense lexical field contains competing lexemes which differ from each other only on a single parameter. This is not the case for most lexical items, which participate in sparse lexical fields, where members differ from each other on multiple parameters.

References

Ariel, Mira. 2006. A 'just that' lexical meaning for *most*. In *Where Semantics Meets Pragmatics* [Current Research in the Semantics/Pragmatics Interface 16], Klaus von Heusinger & Ken Turner (eds), 49–91. Oxford: Elsevier.

Ariel, Mira. 2015. Higher-level category *or* constructions: When many is one. *Studies in Pragmatics* 17: 42–60.

Ariel, Mira. In preparation. *Numerous puzzles*. Tel Aviv University.

Ariel, Mira & Mauri, Caterina. 2018. Why use *or*? *Linguistics* 56: 939–994. https://doi.org/10.1515/ling-2018-0020

Austin, John L. 1963. *Sense and Sensibilia*. Oxford: Clarendon Press.

Barsalou, Lawrence W. 1983. Ad hoc categories. *Memory and Cognition* 11: 211–227. https://doi.org/10.3758/BF03196968

Biber, Douglas. 1995. *Dimensions of Register Variation: A Cross-linguistic Comparison*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511519871

Biber, Douglas, Johansson, Stig, Leech, Geoffrey N., Conrad, Susan & Finegan, Edward. 1999. *Longman Grammar of Spoken and Written English*. Harlow, Essex: Longman.

Breheny, Richard. 2008. A new look at the semantics and pragmatics of numerically quantified noun phrases. *Journal of semantics* 25: 93–139. https://doi.org/10.1093/jos/ffm016

Bultinck, Bert. 2005. *Numerous Meanings. The Meaning of English Cardinals and the Legacy of Paul Grice*. Oxford: Elsevier.

Carston, Robyn. 2002. *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell. https://doi.org/10.1002/9780470754603

Channell, Joanna. 1994. *Vague Language*. Oxford: OUP.

Du Bois, John W., Chafe, Wallace L., Meyer, Charles, Thompson, Sandra A. & Martey, Nii. 2003. Santa Barbara Corpus of Spoken American English, Part 2. Philadelphia PA: Linguistic Data Consortium.

Du Bois, John W., Chafe, Wallace L., Meyer, Charles, Thompson, Sandra A., Englebretson, Robert & Martey, Nii. 2000–2005. Santa Barbara Corpus of Spoken American English, Parts 1–4. Philadelphia PA: Linguistic Data Consortium.

Horn, Laurence R. 2018. The lexical clone: Pragmatics, prototypes, productivity. In *Exact Repetition in Grammar and Discourse*, Rita Finkbeiner & Ulrike Freywald (eds), 233–264. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110592498-010

Kadmon, Nirit. 2001. *Formal Pragmatics*. Oxford: Blackwell.

Lakoff, George. 1972. Hedges: A study in meaning criteria and the logic of fuzzy concepts. In *CLS 8, Papers from the Eighth Regional Meeting of the Chicago Linguistic Society 1972*, Paul M. Peranteau, Judith N. Levi & Gloria C. Phares (eds), 183–228. Chicago IL: CLS.

Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago IL: University of Chicago Press. https://doi.org/10.7208/chicago/9780226471013.001.0001

Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar, Vol. 1: Theoretical Prerequisites*. Stanford CA: Stanford University Press.

Pagel, Mark & Meade, Andrew. 2017. The deep history of the number words. *Philosophical Transactions of the Royal Society of London B 373: 20160517*.

Prince, Ellen F., Frader, Joel & Bosk, Charles. 1982. Hedging in physician-physician discourse. In *Linguistics and the Profession*, Robert J. Di Pietro (ed.), 83–97. Norwood, NJ: Ablex.

Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey N. & Svartvik, Jan. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

Rosch, Eleanor & Mervis, Carolyne B. 1975. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology* 7: 573–605. https://doi.org/10.1016/0010-0285(75)90024-9

Solt, Stephanie, Cummins, Chris & Palmović, Marijan. 2017. The preference for approximation. *International Review of Pragmatics* 9: 248–268. https://doi.org/10.1163/18773109-00901010

Sperber, Dan & Wilson, Deirdre. 2008. A deflationary account of metaphor. In *The Handbook of Metaphor*, Raymond W. Gibbs , Jr. (ed.), 171–203. Cambridge: CUP. https://doi.org/10.1017/CBO9780511816802.007

Wilson, Deirdre & Carston, Robyn. 2007. A unitary approach to lexical pragmatics: Relevance, inference and ad hoc concepts. In *Pragmatics* Noel Burton-Roberts (ed.), 230–259. Houndmills: Palgrave. https://doi.org/10.1057/978-1-349-73908-0_12

Index

This book addresses the topic of linguistic categorization from a novel perspective. While most of the early research has focused on how linguistic systems reflect some pre-existing ways of categorizing experience, the contributions included in this volume seek to understand how linguistic resources of various nature (prosodic cues, affixes, constructions, discourse markers, ...) can be 'put to work' in order to actively build categories in discourse and in interaction, to achieve social goals. This question is addressed in different ways by researchers from different subfields of linguistics, including psycholinguistics, conversation analysis, linguistic typology and discourse pragmatics, and a major point of innovation is represented in fact by the interdisciplinary nature of the volume and in the systematic search for converging evidence.