Generative SLA in the Age of Minimalism

Features, interfaces, and beyond

Language Acquisition & Language Disorders

67

Edited by
Tania Leal
Elena Shimanskaya
Casilde A. Isabelli

John Benjamins Publishing Company

Generative SLA in the Age of Minimalism

Language Acquisition and Language Disorders (LALD)

ISSN 0925-0123

Volumes in this series provide a forum for research contributing to theories of language acquisition (first and second, child and adult), language learnability, language attrition and language disorders.

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

Volume 67

Generative SLA in the Age of Minimalism. Features, interfaces, and beyond.
Selected proceedings of the 15th Generative Approaches
to Second Language Acquisition Conference
Edited by Tania Leal, Elena Shimanskaya and Casilde A. Isabelli

Generative SLA in the Age of Minimalism

Features, interfaces, and beyond

Selected proceedings of the 15th Generative Approaches to Second Language Acquisition Conference

Edited by

Tania Leal
University of Arizona

Elena Shimanskaya
Casilde A. Isabelli
University of Nevada, Reno

John Benjamins Publishing Company
Amsterdam / Philadelphia

The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

DOI 10.1075/lald.67

Cataloging-in-Publication Data available from Library of Congress:
LCCN 2022019210 (PRINT) / 2022019211 (E-BOOK)

ISBN 978 90 272 1144 6 (HB)
ISBN 978 90 272 5756 7 (E-BOOK)

John Benjamins Publishing Company · https://benjamins.com

Table of contents

Acknowledgments

The editors of the volume would like to thank Cristopher Font-Santiago for his help in copy-editing the manuscripts and in compiling the index. We also appreciate the support of the Department of World Languages and Literatures of the University of Nevada, Reno in helping with the copy-editing and indexing efforts.

We owe our gratitude to the reviewers below for their detailed feedback on the chapters. Their help was invaluable in preparing the chapters for publication:

Gonzalo Campos Dintrans, Alejandro Cuza, Maria del Mar Vanrell, Amanda Edmonds, Claire Foley, Heather Goad, Timothy Gupton, Michael Harrington, Julia Herschensohn, Alex Ho-Cheong Leung, Tania Ionin, Tiffany Judy, Oksana Laleko, Christina Lindqvist, Heather Marsden, Utako Minai, Silvina Montrul, Tom Rankin, Jeffrey B. Renaud, Guillermo Rodríguez, Maureen Scheidnes, Neal Snape, Rex A. Sprouse, Mari Umeda.

Finally, we are grateful to the series editors, Roumyana Slabakova and Lydia White, for their careful attention to the chapters at the later stages of the editing process.

Introduction

Elena Shimanskaya
University of Nevada, Reno

GASLA XV took place at the University of Nevada, Reno on March 22–24, 2019, and brought together researchers from around the world to discuss how the latest developments in linguistic theory and psycholinguistic methodologies have influenced the generative study of second language acquisition. The conference included 25 oral presentations, 18 posters, three keynote addresses, a workshop on mixed-effects models, and a panel discussion about the future of generative SLA. The studies selected for this volume showcase how the latest understanding of language architecture can guide and inform research on different aspects of bilingualism/multilingualism: from the linguistic properties under investigation and L1 – L2/Ln language pairings down to the specific research questions in each study. Reflecting the plurality of the field, this volume is divided into four distinct sections: *L2 acquisition of syntax*, *L2 acquisition at interfaces*, *The impact of previously learned languages*, and *Exploring extra-linguistic factors and their impact on L2 acquisition*.

When it comes to syntax, generative linguists have been working on a comprehensive theory that explains cross-linguistic variation while upholding the idea that grammatical knowledge is rule-based (Hornstein, 2009). Accordingly, conceptual models have evolved from Principles-and-Parameters theory in the 1960s, 1970s, and 1980s (Chomsky, 1965, 1986) to the Minimalist program starting in the 1990s (Chomsky, 1995). These developments in linguistic theorizing have had a direct impact on how language acquisition, including second language acquisition, is understood and studied (Rothman & Slabakova, 2018).

Under the Principles-and-Parameters framework, linguistic knowledge was conceptualized as consisting of principles, a set of innate properties containing linguistic information that is common to all human languages, and parameters, which capture grammatical variation among human languages and which were proposed to work like switches. The gist of this notion is that every human being is born with the same "toolbox" of options available to them (Koster, 2013), while the input from a learner's native language shapes their linguistic system and allows them to select the appropriate setting for every parameter. In terms of second language

https://doi.org/10.1075/lald.67.00shi

acquisition, this view of the linguistic faculty led to a series of research questions that were widely studied in the 1990s, including questions such as: Can parameter settings be reset in learning an L2 (e.g., Bley-Vroman, 1990; Clahsen, 1990; Flynn & Martohardjono, 1994)?; do adult L2 learners transfer parameter settings from their L1 into their L2 (e.g., Eubank, 1993; Schwartz & Sprouse, 1994, 1996; Vainikka & Young-Scholten, 1994, 1996; White, 1990/1991)?; and, are the same properties that cluster together in L1 (due to a certain parameter setting) acquired as a cluster in the L2 (e.g., Dekydtspotter, Sprouse, & Anderson, 1998; White, 1992)?

To give an example, Lydia White carried out several studies to investigate the L2 acquisition of the Verb Movement parameter in L2 English and L2 French (White 1990/1991; White, 1991; White 1992). Differences in word order between English and French were attributed to the difference in the setting of the Verb Movement parameter in the two languages: strong in French, weak in English. Without going into detail, the theoretical proposal suggested that verbs move to a higher position in French than in English, resulting in French verbs appearing in front of negation particles, adverbs, and subjects in questions. While the empirical research showed that some native speakers of English learning French accepted the target-like L2 pattern, where adverbs can immediately follow verbs (White, 1991), native speakers of French who were learning English were not successful at rejecting this word order in English (White, 1992). The same Francophone learners of English had target-like judgements of word order in negation and questions, which was taken to indicate that L2 learners did not show target-like performance on the entire cluster of properties that were associated with the same parameter, but only on some of them. This finding was not easy to reconcile with the view of parameters that was accepted at the time. Empirical findings from studies of interlanguage grammars such as the ones described above further shaped the theoretical foundations of the study of second language acquisition in general and generative second language acquisition in particular, resulting in a shift away from the parametric view.

In response to the dramatic increase in proposed parameters, the Minimalist Program (Chomsky, 1995) hypothesized that using an inventory of *features* could provide a more parsimonious way to account for cross-linguistic variation than an inventory of parameters. Some SLA studies, such as White's studies discussed above, had also suggested that a more nuanced approach was needed to account for the phenomena uncovered in L2 acquisition, in this case the lack of clustering effects around a given parameter. Many generative linguists responded to this call by (re)analyzing different linguistic phenomena, relying on an inventory of formal, phonological, semantic, and discourse features including control and raising (Polinsky & Potsdam, 2006), complementizer phrase (Rizzi, 1997), and quantifier scope (Szabolcsi, 2010). In step with this latest thinking in syntactic theory, the field of SLA has also seen the advent of several hypotheses that put features at the forefront: the Feature Reassembly Hypothesis (Lardiere, 2009), the Interpretability Hypothesis (Tsimpli & Dimitrakopoulou, 2007), the Prosodic Transfer Hypothesis (Goad & White, 2006), and the Bottleneck Hypothesis (Slabakova, 2008). These hypotheses were formulated based on the latest understanding of the language faculty, allowing L2 researchers to formulate and test specific predictions about the L2 acquisition of different morpho-syntactic properties and about L1 transfer.

These trends were in evidence at GASLA XV, as discernible from the contributions that make up the first part of this volume: *L2 acquisition of syntax.* In this section, three studies investigate the L2 acquisition of syntactic phenomena such as raising and scope, while a fourth study looks at the L2 acquisition of definiteness. Focusing on raising constructions, Franciotti and Martohardjono's study of control adopts Collins's (2005) *smuggling* approach, where the movement of the subject out of the embedded clause is motivated by a case-checking requirement and is allowed in English because the entire VP undergoes movement, "smuggling" the DP with it. The authors explicitly investigated how such raising constructions are processed by L2 learners of English who are L1 speakers of Italian. Their results show that even though L2 constraints on raising can be acquired, reading time data suggests that NSs of English are more sensitive than learners when processing ungrammatical raising constructions.

The two studies investigating the L2 acquisition of scope in English focus on L1 speakers of scope-rigid languages: Mandarin Chinese, in the case of Wu and Ionin's study, and Japanese in Kimura's study. Both Wu and Ionin and Kimura acknowledge that the current Minimalist proposals do not specify the exact mechanism that would account for the lack of inverse scope interpretations in languages such as Chinese and Japanese. However, in terms of their L2 acquisition, both studies show evidence that speakers of scope-rigid languages failed to provide target-like judgments when it came to inverse scope. The authors of both studies concluded that (negative) L1 transfer was at play. Wu and Ionin surmise that manipulating the quantity and quality of the input, as well as testing different pedagogical approaches, might be necessary to facilitate L2 acquisition of inverse scope.

In their study of the L2 acquisition of definiteness, Stern, Dayal, Martohardjono, Chen, and Madsen looked at whether discourse context and syntactic function play a role in the L2 acquisition of articles. They tested native speakers of a language that does not have articles in its inventory: Mandarin Chinese. Overall, results show that bare subjects in definite contexts turned out to be more acceptable to these L2 learners than those subjects occurring in indefinite contexts. Some of the results, however, were unexpected because bare objects were equally acceptable in both definite and indefinite contexts. The researchers conclude on a methodological note, suggesting that future studies should try to control for the possibility that participants misinterpret singular nouns as plural (for example, by using a picture selection task) and also to test using sentences with an explicit topic to disentangle the effect of syntactic position from the effect of topicality.

Part II, *L2 acquisition at interfaces*, highlights the investigations that focus on the acquisition of structures and properties that involve several linguistic modules, another topic of interest that emerged in generative SLA in the early 2000s. With a heightened interest in linguistic phenomena that fall outside of the scope of narrow syntax, it was hypothesized that linguistic phenomena that require learners to integrate information from external linguistic modules (e.g., the syntax – discourse interface) are more difficult to acquire and more prone to residual optionality than those linguistic structures only requiring the integration of internal modules (e.g., the syntax – semantics interface) or syntax alone. This proposal, termed the 'Interface Hypothesis,' was first introduced by Sorace and Filiaci (2006), and numerous studies have investigated different external interface phenomena in the intervening years. In the current volume, there are three studies that investigate properties that rely on two or more linguistic modules for successful L2 acquisition, although not all of them involve external interfaces.

Fujimori, Yamane, Yoshimura, Nakayama, Teaman, and Yoneyama focus on an external interface in their experimental study, which investigated a linguistic property (sentence focus marked by pitch) that requires input from different linguistic modules – in their case, prosody, discourse, and syntax. Fujimori and colleagues present the results of three tasks: a written task where participants chose the focus of the sentence; a listening task where participants chose the most appropriate pitch pattern for focus questions targeting different constituents in a sentence; and a reading task where participants were asked to read (and, thus, assign a prosodic contour to) question-answer pairs. The study revealed that although lower-proficiency learners could identify words that were the focus in the written task, they chose fewer answers with target-like prosodic contour than NSs and higher proficiency L2ers in the listening task. When it came to production, only NSs of English made a three-way distinction among different types of focus. Higher proficiency learners distinguished between broad/late narrow focus and early narrow focus, while lower proficiency learners made no distinctions based on focus type. Fujimori and colleagues argue that linguistic properties that require integrating information from multiple modules might be harder to acquire.

However, not all interfaces are external, and neither is the difficulty associated with interfaces squarely at the doorstep of external interfaces such as the syntax – discourse interface. In fact, there is evidence that internal interfaces can also present challenges (White, 2011). In his chapter, Archibald investigated a linguistic property at the intersection of morphology and phonology. Even though the participants in his experiment completed a written task, he analyzed the data to see whether they produced plural forms that violated universal phonological constraints. The fact that very few answers fell under this category (only 0.7%) strongly suggests that interlanguage grammars abide by the universal principles constraining all natural language grammars. Ingham's study also examined the interface between phonology and morphology by examining the L2 acquisition of morphological markers. In her case, she examined English tense/aspect inflection. Results of a semi-spontaneous production task indicated that L1 speakers of Bengali transfer some L1 prosodic and phonetic constraints into their L2 English, providing further support to the idea that the task of acquiring morphological markers also relies on the acquisition of L2 prosodic structure, as suggested by the Prosodic Transfer Hypothesis (Goad & White, 2019).

Part III, *The impact of previously learned languages,* is comprised of two studies that concentrate of investigating (the extent of) the role of previously learned languages in L2 and L3 acquisition. Similar to White's studies discussed above, Dahl, Listhaug, and Busterud investigated word-order asymmetries, but in this case focusing on the L3 acquisition of German with Norwegian as L1 and English as L2. Dahl and colleagues found no clear pattern of transfer from either the L1 or the L2; their findings suggest that participants with higher L2 (English) proficiency performed better in their L3 (German) on sentences with adverbials, but not on sentences with topicalization.

While Dahl and colleagues aimed to disentangle L1 and L2 transfer effects on the acquisition of a third language, in her chapter, Smeets set out to test the predictions of the Feature Reassembly Hypothesis (Lardiere, 2009) as it pertains to the L2 acquisition of Italian Clitic Left Dislocation by speakers of English and Romanian. Even though Romanian, like Italian, has Clitic Left Dislocation in its inventory, the feature settings that motivate the appearance of clitics ([±specific], [±anaphor]) are different in the two languages. Using the Feature Reassembly Hypothesis, Smeets made nuanced predictions about the specific discourse contexts that should prove challenging for Romanian speakers, but not for English speakers. Her findings support the predictions of the Feature Reassembly Hypothesis and suggest that overcoming transfer effects proves difficult even for very proficient Romanian speakers of Italian.

Finally, part IV, *Exploring extra-linguistic factors and their impact on L2 acquisition,* includes an experimental study investigating the effect of social bias on L2 acquisition, and two position papers exploring the role of working memory and age. To extend the empirical coverage of generative SLA theories and hypotheses, current generative studies often not only concentrate on linguistic properties, but also systematically incorporate extra-linguistic factors into their experimental design. Even though generative linguistic theory always attributed a vital role to linguistic input (Slabakova, Leal, & Liskin-Gasparro, 2014), the field of generative SLA today makes an explicit goal of investigating how extra-linguistic factors such as working memory limitations, age, and quality and quantity of linguistic input affect how learners use and interpret their L2/Ln. This trend is evident in many chapters discussed above. For example, Franciotti and Martohardjono considered processing, Wu and Ionin considered quantity of input, while Dahl and colleagues examined the effects of L2 proficiency. However, the three chapters that comprise the last section of the volume (contributed by Sokolova and Slabakova, Cunnings, and Montrul) bring to the forefront how language processing and input characteristics influence theorizing in the field of SLA.

Sokolova and Slabakova examined how L2 and L3 speakers resolve relative clause ambiguity when relative clause attachment preferences differ: In this case, the L1 (Russian) and L2 (French/German) evince high attachment while the L3 (English) evinces low attachment. Additionally, the authors investigated social conventions and their impact on relative clause resolution, where social conventions were operationalized as actions that are more likely to be attributed to a certain gender or age group, such as *playing with a ball* which is more likely to be an activity performed by children rather than adults. Sokolova and Slabakova hypothesized that if L3 learners do not use syntactic structure to process relative clauses, then they would be forced to rely on social conventions to resolve them, as predicted by the Shallow Structure Hypothesis (Clahsen & Felser, 2018). This is not, however, what the authors found. Neither group relied on social conventions to resolve ambiguous relative clauses; instead, they adopted the target-like low-attachment preference characteristic of the target language.

In his article, Cunnings also focused on L2 processing and tackled several questions regarding working memory: how we conceptualize working memory, how the different views of working memory emerged throughout the empirical studies of L2 processing, and, finally, how our understanding of working memory can impact our understanding of L2 sentence processing. He argues that interference-based approaches to working memory, focusing on the content of the information in one's memory rather than on the memory capacity, might be a promising approach to the study of L2 acquisition and processing. To conclude the volume, Montrul discusses a particular population of bilinguals – heritage speakers. In recent years, the field of SLA has seen an increased interest in this population, a most welcome trend because studying heritage speakers can shed further light on several fundamental themes in SLA: the difference between adult and child linguistic development, the impact of input on language development, and the difference between bilingual/multilingual and monolingual interlanguage, to name a few. In her article, Montrul gives a thorough review of the empirical research on heritage language development and ultimate attainment. She explains how studying this population contributes to the field of SLA as a whole.

The articles brought together in the four sections of this volume represent some of the latest advances in generative SLA research. From its inception in the mid 1980s to the present day, the field of generative SLA has been trying to determine what roles L1 transfer, age, and proficiency play in learners' interlanguage development while utilizing the latest developments in generative syntactic theory and psycholinguistic methodologies. As we learn even more about L2/Ln acquisition, research questions and theoretical hypotheses have expanded to focus on linguistic properties that rely on integrating information from multiple linguistic modules, on the relationship between underlying grammatical knowledge and language processing, and on the role of input quality and quantity.

References

Bley-Vroman, R. (1990). The logical problem of foreign language learning. *Linguistic Analysis*, 33, 3–49.
Chomsky, N. (1965). *Aspects of the theory of syntax*. The MIT Press.
Chomsky, N. (1986). *Knowledge of language*. Praeger.
Chomsky, N. (1995). *The minimalist program*. The MIT Press.
Clahsen, H. (1990). The comparative study of first and second language development. *Studies in Second Language Acquisition*, 12, 135–153. https://doi.org/10.1017/S0272263100009050
Clahsen, H., & Felser, C. (2018). Notes on the Shallow Structure Hypothesis. *Studies in Second Language Acquisition*, 40(3), 693–706. https://doi.org/10.1017/S0272263117000250
Collins, C. (2005). A smuggling approach to raising in English. *Linguistic Inquiry*, 36(2), 289–298. https://doi.org/10.1162/0024389053710701
Dekydtspotter, L., Sprouse, R. A., & Anderson, B. (1998). Interlanguage A-bar dependencies: binding construals, null prepositions and Universal Grammar. *Second Language Research*, 14(4), 341–358. https://doi.org/10.1191/026765898677792795
Eubank, L. (1993). On the transfer of parametric values in L2 development. *Language Acquisition*, 3(3), 183–208. https://doi.org/10.1207/s15327817la0303_1
Flynn, S., & Martohardjono, G. (1994). Mapping from the initial state to the final state: The separation of universal principles and language-specific principles. In B. Lust, M. Suñer, & J. Whitman (Eds.), *Syntactic theory and first language acquisition: Cross-linguistic perspectives. Vol. 1: Heads, projections and learnability* (pp. 319–335). Lawrence Erlbaum Associates.
Goad, H., & White, L. (2006). Ultimate attainment in interlanguage grammars: A prosodic approach. *Second Language Research*, 22, 243–268. https://doi.org/10.1191/0267658306sr2680a
Goad, H., & White, L. (2019). Prosodic effects on L2 grammars. *Linguistic Approaches to Bilingualism*, 9(6), 769–808. https://doi.org/10.1075/lab.19043.goa
Hornstein, N. (2009). *A theory of syntax: Minimal operations and universal grammar*. Cambridge University Press.
Koster, J. (2013). Theories of language from a critical perspective. In J. Herschensohn & M. Young-Scholten (Eds.), *The Cambridge handbook of second language acquisition* (pp. 9–25). Cambridge University Press. https://doi.org/10.1017/CBO9781139051729.003

Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research*, 25, 173–227. https://doi.org/10.1177/0267658308100283
Polinsky, M., & Potsdam, E. (2006). Expanding the scope of control and raising. *Syntax*, 9(2), 171–192. https://doi.org/10.1111/j.1467-9612.2006.00090.x
Rizzi, L. (1997). The fine structure of the left periphery. In L. Haegeman (Ed.), *Elements of grammar: Handbook of generative syntax* (pp. 281–337). Kluwer. https://doi.org/10.1007/978-94-011-5420-8_7
Rothman, J., & Slabakova, R. (2018). The generative approach to SLA and its place in modern Second Language studies. *Studies in Second Language Acquisition*, 40(2), 417–442. https://doi.org/10.1017/S0272263117000134
Schwartz, B. D., & Sprouse, R. A. (1994). Word order and nominative case in non-native language acquisition: A longitudinal study of (L1 Turkish) German interlanguage. In T. Hoekstra & B. D. Schwartz (Eds.), *Language acquisition studies in generative grammar: Papers in honor of Kenneth Wexler from the 1991 GLOW workshops* (pp. 317–368). John Benjamins. https://doi.org/10.1075/lald.8.14sch
Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the full transfer/full access model. *Second Language Research*, 12(1), 40–72. https://doi.org/10.1177/026765839601200103
Slabakova, R. (2008). *Meaning in the second language*. Walter de Gruyter. https://doi.org/10.1515/9783110211511
Slabakova, R., Leal, T. L., & Liskin-Gasparro, J. (2014). We have moved on: Current concepts and positions in generative SLA. *Applied Linguistics*, 35(5), 601–606. https://doi.org/10.1093/applin/amu027.
Sorace, A., & Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Language Research*, 22, 339–368. https://doi.org/10.1191/0267658306sr271oa
Szabolcsi, A. (2010). *Quantification*. Cambridge University Press. https://doi.org/10.1017/CBO9780511781681
Tsimpli, I. M., & Dimitrakopoulou, M. (2007). The interpretability hypothesis: Evidence from wh-interrogatives in second language acquisition. *Second Language Research*, 23, 215–242. https://doi.org/10.1177/0267658307076546
Vainikka, A., & Young-Scholten, M. (1994). Direct access to X'-theory: Evidence from Korean and Turkish adults learning German. In T. Hoekstra & B. D. Schwartz (Eds.), *Language acquisition studies in generative grammar: Papers in honor of Kenneth Wexler from the 1991 GLOW workshops* (pp. 265–316). John Benjamins. https://doi.org/10.1075/lald.8.13vai
Vainikka, A., & Young-Scholten, M. (1996). Gradual development of L2 phrase structure. *Second Language Research*, 12(1), 7–39. https://doi.org/10.1177/026765839601200102
White, L. (1990/1991). The verb-movement parameter in second language acquisition. *Language Acquisition*, 1(4), 337–360. https://doi.org/10.1207/s15327817la0104_2
White, L. (1991). Argument structure in second language acquisition. *Journal of French Language Studies*, 1(2), 189–207. https://doi.org/10.1017/S0959269500000983
White, L. (1992). Long and short verb movement in second language acquisition. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique*, 37(2), 273–286. https://doi.org/10.1017/S0008413100021988
White, L. (2011). Second language acquisition at the interfaces. *Lingua*, 121(4), 577–590. https://doi.org/10.1016/j.lingua.2010.05.005

L2 acquisition of syntax

Intervention effects in L1 and L2 English raising

Evidence from acceptability judgments and response times

Pamela Franciotti and Gita Martohardjono
The Graduate Center / City University of New York

We investigate L1 Italian-L2 English speakers using three types of subject raising constructions: Raising over lexical DPs, pronominal DPs, and topicalizations. We test locality constraints in L2 English, including how intervention effects affect the L2 processing of A-dependencies and whether exceptionality to certain locality constraints are learnable. Three main findings emerged: (i) L2 speakers are sensitive to intervention, yet exceptions to locality can be learned; (ii) intervening DPs elicited higher processing loads only for native controls; (iii) raising with topicalizations facilitated processing only for native speakers, even though topicalization is grammatical in the L2er's L1. Results indicate that native and non-native grammars eventually converge, exceptionality to universal constraints is learnable, and differences between native and non-native speakers lies primarily in processing.

Keywords: relativized minimality, intervention effects, L1 Italian, L2 English, L2 grammatical representation, L2 processing, subject raising, A-dependencies

1. Introduction

A long-standing issue in second language acquisition (SLA) is evaluating the role that the first language (L1) plays in the development of the second language (L2) grammar. In generative SLA, transfer was a mainstay of research within the Principles and Parameters model (P&P) of Universal Grammar (UG) (Chomsky, 1981). The P&P model offered an ideal framework for L2 studies to investigate transfer phenomena, such as the effects of different parameter settings in the source and target languages. Nativelike resolution of such effects was explained as

https://doi.org/10.1075/lald.67.01fra

successful parameter resetting (see White, 2003 for review). The theoretical shift from the P&P model of grammar to the Minimalist Program (Chomsky, 1995), which redefined grammatical properties with the more fine-grained concept of syntactic *feature*, inevitably led the SLA field to reformulate parameter resetting in terms of feature reassembly (e.g., Lardiere, 2008; see also Rothman & Slabakova, 2018 for a discussion on the effects of the shift in SLA).

In either framework, L1 transfer as source of L2 error has exclusively investigated linguistic properties that are assumed to be fundamentally different in the native and target languages (i.e., derived from different parameter/feature values; see Slabakova, 2016 for review). A less explored question in generative SLA is whether *exceptions* to principles that otherwise hold across the L1 and the L2 are acquirable and, if so, to what degree.

Our study attempts to fill this gap by investigating the instantiation of a syntactic constraint said to hold cross-linguistically: *Relativized Minimality* (RM) (Rizzi, 1990 et seq.). In particular, we investigate the acquisition of exceptions to this constraint. RM, or *locality,* is a structural relation that holds between two syntactic elements when no third element that is similar in nature intervenes in the syntactic dependency. In English, the extraction of a *wh*-element (1a) over another *wh*-element (1b) results in ungrammaticality (examples from Rizzi, 2013, p. 172):

(1) a. **When** do you think [John left ___]?
b. ***When** do you wonder [**who** left ___]?

Intervention effects have been widely explored for *A-bar* dependencies in L1 acquisition (e.g., Belletti et al., 2012; Friedmann et al., 2009) and have been supported by a number of L1 adult sentence-processing studies (e.g., Gibson, 1998; Gordon et al., 2001, 2004; Warren & Gibson, 2002, 2005). While some attention has been paid to the same effects in *A*-dependencies, this research has mostly focused on child L1 acquisition (e.g., Choe & O'Grady, 2017; Hirsch, 2011; Hirsch & Wexler, 2007). In SLA, studies testing syntactic intervention in *A*-dependencies involving raising have been conducted with L1 Korean speakers (Choe, 2016, 2019) and with L1 Japanese speakers learning L2 English (Yoshimura & Nakayama, 2019) (but see also e.g., Xia et al., 2020 for recent findings on intervention effects in the L2 processing of relative clauses).

Our study contributes to this literature by investigating intervention effects with L1 Italian/L2 English speakers and exploring processing effects by using response times (RTs) as a measure. Because locality is assumed to hold cross-linguistically, an interesting question arises when locality is argued to be active for a particular structure in one language, while being exceptionally inactive in the other. One such example is the asymmetry documented in the syntactic literature between subject raising constructions in Italian (and Romance languages, more generally) and English (e.g., Boeckx, 2008; Hartman, 2011; McGinnis, 1998; Rizzi, 1986). This contrast can be observed in (2):

(2) a. *[$_{\text{IP}}$ Gianni sembra **a Maria/a lei** [$_{\text{IP}}$ <Gianni> essere felice]].
John seem.3SG to Mary/to her be.INF happy
"John seems to Mary/to her to be happy."

b. [$_{\text{IP}}$ John seems **to Mary/to her** [$_{\text{IP}}$ <John> to be happy]].

(2a) shows that, in Italian, subject raising across a lexical/pronoun DP experiencer is unacceptable, while the equivalent English sentence is acceptable (2b). The judgment of (2a) is expected given intervention effects: the DP experiencer blocks A-movement of the subject to its designated surface position. The same intervention effects are presumed to render the equivalent construction unacceptable in English as well. However, it has been argued in the literature that sentences like (2b) are "exceptions" to RM (e.g., Chomsky, 1995; Collins 2005, amongst others) and that English allows certain raising constructions to escape the constraint, in contrast to Italian, which behaves as predicted.

Various mechanisms have been proposed to account for these exceptions. We will illustrate the exceptionality to intervention effects in English raising with Collins' (2005) *smuggling* approach (but see also e.g., Boeckx, 2008; Chomsky, 2000; Kitahara, 1997; McGinnis, 1998, for other accounts), yet the empirical validity of the judgments in (2b) has not been tested widely, nor have the analyses advanced in the theoretical literature been supported by experimental data. Thus, one of the indirect, yet necessary aims of our study is to verify that the exception to locality in English raising is indeed reflected in native speaker judgments.[1]

The main goal of our study is to investigate whether the purported asymmetry of the subject raising facts in Italian and English impact L2 competence and performance. We investigate the effect of *Relativized Minimality* in L2 acquisition, including cases where the L2 is said to allow exceptions to the constraint. Our primary research question asks whether L2 English speakers whose L1 is Italian are able to acquire such exceptions to RM. As a second question, we test whether different instantiations of the same constraint in the L1 and L2 result in different processing signatures in native and non-native speakers of English. Finally, we investigate the empirical validity of the raising exception to RM in native speakers of English. To this end, we collected acceptability judgments from a group of L1 Italian-L2 English speakers (L2ers) and contrast them with those of a control group of English native speakers (NSs) (see Ionin & Zyzik, 2014 for a review on the methodology in SLA). We adopted a mixed methodology, collecting acceptability

1. For discussion on how SLA data can contribute to a verification of proposed UG principles in the L1, see White, 1989, p. 60.

judgment RTs in an effort to assess both judgment and processing reaction times with respect to intervention effects in the two target populations.

The paper is organized as follows: In Section 2, we review the notion of syntactic locality by comparing Italian and English facts on raising with experiencers in both intervening and non- intervening positions and briefly describe Collins' (2005) *smuggling* approach to English raising. In Section 3, we review a sample of existing studies in L2 acquisition that explored the phenomenon of intervention in raising structures. Section 4 presents the details of the current study while we discuss the findings and conclusions in Sections 5 and 6.

2. Raising and minimality

2.1 Italian/English asymmetry on raising across DPs

Italian and English raising predicates such as seem/*sembrare* take clausal non-finite complements whose DP embedded subjects move from their original positions, where they are assigned a theta role, to an A-position (Spec, IP). Subject raising is motivated by case checking and by a requirement to establish an *Agree* relation (Baltin, 2001):

(3) a. [$_{\text{IP}}$ Sembra che Gianni [$_{\text{IP}}$ sia felice]]. [Italian]
It seem.3SG that John be.3SG.SBJV happy
"It seems that John is happy."
b. [$_{\text{IP}}$ Gianni sembra [$_{\text{IP}}$ <Gianni> essere felice]].
John seem.3SG be.INF happy
"John seems to be happy."

(4) a. [$_{\text{IP}}$ It seems [$_{\text{IP}}$ that John is happy]]. [English]
b. [$_{\text{IP}}$ John seems [$_{\text{IP}}$<John> to be happy]].

In order for the structural relation to be established, the relation must be local, i.e., no element may intervene between the DP subject and its target position in Spec, IP, as such movement would violate *Relativized Minimality* (Rizzi, 1990 et seq.), which defines syntactic intervention in terms of c-command and typology of positions:

(5) *Relativized Minimality*: In a configuration [X Z Y], a dependency between X and Y cannot hold if there is an intervening Z such that:
i. Z is of the same structural type as X
ii. Z c-commands Y and Z does not c-command X
(adapted from Rizzi, 2004, p. 225 (4), (5), (6))

Since Rizzi's (1990) seminal work on the original RM principle, a *featural* approach to RM (*f*RM) (Rizzi, 2001, 2004; Starke, 2001) further defined syntactic intervention on the basis of the morphosyntactic configurations held by the two elements involved in the structural dependency: When the intervener Z does not fully match the morphosyntactic features responsible for triggering the syntactic movement, intervention effects are predicted to be milder (Belletti et al., 2012; Belletti & Rizzi, 2013; Friedmann et al., 2009).

As in other Romance languages (e.g., French: McGinnis, 1998; Spanish: Torrego, 1996), Italian raising of a subject across a DP lexical or pronominal experiencer leads to the predicted intervention effects (6a), unless the experiencer undergoes cliticization (6b):

(6) a. *[$_{\text{IP}}$ Gianni sembra **a Maria/a lei** [$_{\text{IP}}$ <Gianni> essere felice.]]
John seem.3SG to Mary/to her be.INF happy
"John seems to Mary/to her to be happy."

b. [$_{\text{IP}}$ Gianni **le** sembra [$_{\text{IP}}$ <Gianni> essere felice]].
John to.her.CL.DAT.F.SG seem.3SG be.INF happy
"John seems to her to be happy"

In English, it has been argued that RM does not yield intervention effects in subject raising, and that this is would be considered an "exception to RM" because intervention effects hold elsewhere in the grammar, e.g., in subject-to-object raising (see Hartman, 2011 for discussion). This has been referred to as the "experiencer-paradox" (Boeckx, 1999) whereby DP subjects are apparently permitted to cross over a DP experiencer, either lexical or pronominal [to Mary/to her], (7), without triggering any intervention effects (e.g., Chomsky, 1995; Collins, 2005, among others):

(7) [$_{\text{IP}}$ John seems **to Mary/to her** [$_{\text{IP}}$ <John> to be happy]].

In English, *f*RM would predict intervention effects to be milder with pronominal than with lexical DPs, which should lead to higher acceptability and better performance in comprehension tasks (e.g., Belletti & Rizzi, 2013; Friedmann et al., 2009). This is due to the relevant morphosyntactic features of the intervener ([to her] in (7)) not fully matching those of the target [John].

2.2 Italian/English symmetry on raising with topic DPs

Another subject raising construction that is relevant to our discussion is the topicalized experiencer type. We have seen that Italian and English differ with respect to subjects raising across experiencer DPs. However, both languages can overtly express an experiencer if it is topicalized. This is because the DP does not violate locality constraints in topic position because it is instead dislocated to the left periphery (Boeckx, 2008; McGinnis, 1998).

In terms of information structure, DPs representing old information that is recoverable from the discourse context are commonly called *topics* and are expressed in so-called topic-comment constructions (e.g., Cinque, 1990; Rizzi, 1997). In English and Italian topic-comment constructions, the topic is generally prosodically marked and/or dislocated in the left periphery of the clause, namely the C-domain, and is immediately followed by a comment in the form of a predicate in the T-domain (Rizzi, 1997). Dislocated topic structures have been argued to be pragmatically (but not syntactically) equivalent in English and Romance languages. In English, topicalization has been considered as an instance of *wh*-movement while in Romance languages, including Italian, the operation is referred to as *Clitic Left Dislocation* (CLLD) and, under some accounts, does not involve movement (e.g., Cinque, 1990). According to Cinque's (1990) proposal, CLLD topics base-generate in the left periphery and connect to a resumptive clitic pronoun internal to the clause.

Rizzi's (1997) feature-based account of Italian CLLD and English topicalization instead assumes that topics undergo *A-bar* movement to the left periphery attracted to the specifier of a Topic head by a [+Topic] feature (see also e.g., López (2009) for another movement account). Following Cinque (1990), Rizzi (1997) claims that the main difference between the two languages consists in the type of operator binding the topics: in Italian topics are bound by a clitic pronoun (which is obligatory when the DP is a direct object (see 8) and optional when the topic it is a cliticizable prepositional phrase (PP) (see 9)), while in English they are bound by a null operator (OP) (see 10) (examples adapted from Rizzi 1997, pp. 285–294):

(8)	[$_{CP}$ [$_{TopP}$	Il	tuo	libro,	[$_{IP}$ **lo**	ho	letto]]].
		the.DET.M.SG	your	book,	it.CL.ACC.M.SG	have.1SG	read

“Your book, I have read it.”

(9)	[$_{CP}$ [$_{TopP}$	A	Gianni,	[$_{IP}$ Maria	(**gli**)	ha	parlato
		To	Gianni,	Maria	(to.him.CL.DAT.M.SG)	have.3SG	spoken

recentemente]]].
recently.
“To Gianni, Maria spoke to him recently.”

(10) [$_{CP}$ [$_{TopP}$ Your book, [$_{IP}$ **OP** you should give <your book> to Paul (not to Bill)]]

In CLLD/topicalization constructions of subject raising, a dislocated topic would then be a viable option to express an experiencer in either language because both respect locality principles (example adapted from Boeckx 2008, p. 151):

(11) [$_{CP}$ [$_{TopP}$ A Maria, [$_{IP}$Gianni (le) sembra [$_{IP}$ <Gianni>
To Mary, John (to.her.CL.DAT.F.SG) seem.3SG
essere felice]]]].
Be.INF happy
"To Mary, John seems (to her) to be happy"

(12) [$_{CP}$ [$_{TopP}$ To Mary, [$_{IP}$ John seems [$_{IP}$ <John> to be happy]]]]

The examples in (11) and (12) show that the DP experiencer is dislocated from a clause internal intervening position to an external non-intervening position in the left periphery, avoiding any (apparent) RM effects in either language.

The following table summarizes the expected absence or presence of intervention effects arising in the two languages:

Table 1. Italian and English (non)-intervention effects in raising structures

	Across a lexical/pronoun DP	**w/ a Topic DP**
Italian	* Intervention	✓ No Intervention
English	✓ No Intervention	✓ No Intervention

To sum up: Raising across lexical and pronoun DP experiencers violates RM in both languages, yet triggers intervention effects only in Italian. As predicted by RM, CLLD/topicalization does not raise locality issues in either language, both allowing the option of an overt DP experiencer.

2.3 Smuggling approach to English raising (Collins, 2005)

Within the Minimalist Framework, several theories have been proposed to account for the exceptionality of English when raising across DPs and for its sharp contrast with Romance languages, yet the phenomenon is still under debate. In a recent proposal, Collins (2005) analyzes English raising constructions under a *smuggling* approach. Collins claims that the DP subject *smuggles* over the DP experiencer that c-commands the DP subject prior to raising. In (7) (repeated in (13) below for convenience), the whole VP [John seems] moves to [Spec, v] *smuggling* the DP embedded subject over the intervening experiencer. From its landing position, the subject then makes its way up to [Spec, IP] without crossing any intervening elements:

(13) [$_{IP}$ John seems **to Mary/to her** [$_{IP}$ to be happy]].

Smuggling is assumed to take place after the experiencer is introduced into the Spec of an Applicative Phrase (ApplP) and after an extraposition of the embedded IP [to be happy] to the Spec of an undefined functional head, X. The derivation is represented in Figure 1 (Collins, 2005, p. 295):

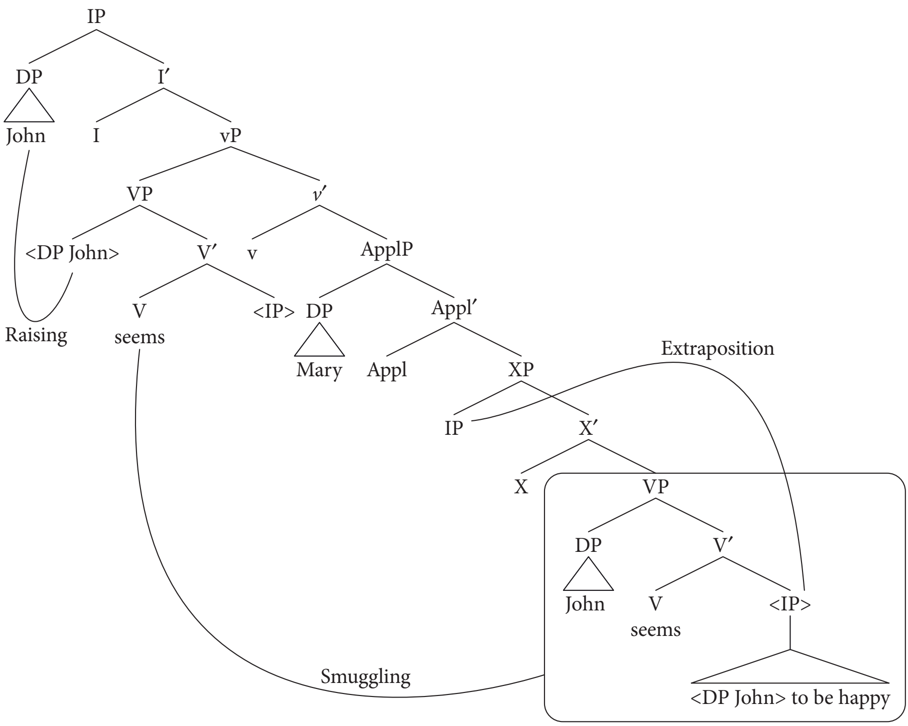


Figure 1. *Smuggling* to English raising (adapted from Collins 2005, p. 295)

Collins leaves open two possibilities for the way the preposition *to* is introduced into the derivation: either as left-adjunction to the DP experiencer during spell-out or as its own phrase merging with ApplP as its complement.

The *smuggling* analysis to raising offered by Collins provides us with a mechanism that explains a possible way a DP subject crossing over a DP experiencer avoids intervention effects. Although we do not further address the exceptionality of English to RM from a theoretical perspective, we will return to the issue later when we discuss our results, which suggest that the acceptability of (13) might not be as ubiquitous as argued in the syntactic literature, putting into question the status of these structures as exceptions to RM. For the L2 learner, this would mean that they would have to apply "smuggling" in the L2 English, even though it does not occur in their L1, Italian.

3. Intervention effects in the L2 acquisition research on raising

In the SLA literature, research on the phenomenon of syntactic intervention in *A*-dependencies has been limited in scope and has focused primarily on L2 English. Studies looking at the acquisition of raising in L2 have examined the effects of different types of experiencers in intervening and non-intervening positions exclusively based on comprehension.

Choe (2016) investigates intervention effects in the L2 comprehension of English raising via L1 transfer from Korean, a language that does not allow raising with either in situ or fronted experiencers. In this study, Choe tested raising across lexical experiencers (14a) and with fronted experiencers (14b) by administering a Truth Value Judgment Task (TVJT) to a group of adult L1 Korean – L2 English learners at different levels of proficiency (examples from Choe, 2016, p. 749):

(14) a. Donald seems to Mickey to be short.
b. To Mickey, Donald seems to be short.

Choe's findings show that L2ers performed poorly in the lexical DP condition (14a) while accuracy rates significantly increased in raising structures with the experiencer in a fronted position (14b) (41.7% vs. 76%). Choe (2016) reports an L2 developmental pattern for the lexical DP condition only: accuracy rates increased significantly with proficiency level, but no difference emerged under the fronted experiencer condition, which improved comprehension regardless of proficiency level. In a follow-up study, Choe (2019) administered a TVJT to the same L2 group, in which the DP type (lexical vs. pronoun) of both the experiencer and the subject was manipulated (15a)–(b) (examples from Choe, 2019, p. 170):

(15) a. Donald seems to him to be short.
b. He seems to Mickey to be short.

This follow-up study shows that while raising of a lexical subject across a pronominal experiencer (15a) improved L2 performance (55.8%), the reversed pattern (15b) elicited lower accuracy rates (26.2%). Additionally, results indicated a developmental pattern only in the comprehension of (15a), i.e., both low- and high-proficiency groups performed equally poorly in the comprehension of (15b) (25% vs. 27.5%). Interestingly, the author reports that the pronoun subject-lexical experiencer pattern (15b) was also problematic in the unraised condition especially for the low proficiency group, an effect that was not observed for (15a). This finding suggests that the source of complexity with the pronominal subject raising might not be intervention. According to Choe (2019), L2 learners might experience difficulties with the interpretation of pronouns preventing them from identifying the correct referent.

Yoshimura and Nakayama (2019) explored the phenomenon in L1 Japanese-L2 English learners. The study investigates the role played by the Japanese grammar, a system in which subject raising constructions do not undergo *A*-movement in the L2 acquisition of English. Accuracy data was collected with a TVJT that included raising structures of the type shown in (14a)–(b) and raising with pronominal experiencers both in-situ (16a) and fronted (16b) positions (examples from Yoshimura & Nakayama, 2019, p. 263):

(16) a. Amy appeared to him to be full.
b. To him, Virginia seemed to be rich.

In line with Choe (2016), the authors found that L2 learners struggled in the comprehension of raising with lexical subjects crossing over an intervening lexical argument as in (14a), but performance improved when the experiencer was fronted as in (14b) (59.1% vs. 90.5% respectively). The authors also report a facilitation effect with a pronominal experiencer (16a) similar to Choe (2019), although the effect was visible in both positions (16a)–(b) (in-situ: 86.3% vs. fronted: 79.4%).

Although the L2 literature is limited, findings seem to pattern with L1 child studies. In a sentence-picture matching comprehension task conducted with 3- to 9-year-old L1 English speakers, Hirsch (2011) found that while *A*-movement is generally delayed, raising with a lexical experiencer in fronted position is much easier for the children to comprehend than raising with an intervening lexical experiencer. In a similar study using a TVJT with 3- to 7-year-old L1 English speakers, Choe and O'Grady (2017) found that comprehension is degraded when a lexical experiencer is in an intervening position, but that performance improved when the experiencer was presented as a pronoun.

4. The present study

The present study aims to investigate intervention effects in the representation of L1 and L2 English raising and to explore whether intervention effects are present in native and non-native processing. As mentioned in the introduction, the primary question we seek to answer is whether L2 English speakers whose L1 is Italian are able to acquire exceptions to RM, thus effectively overriding a locality principle that operates in both the L1 and the L2. We examine grammatical representations by collecting acceptability judgments to the following targeted structures: Subject raising across either a lexical or a pronominal experiencer; subject raising with a topicalized experiencer; and as controls, subject raising with no experiencers and unraised structures. If L2 speakers follow RM constraints, they should reject subject raising across lexical experiencers in English.

On the other hand, if speakers have identified these constructions as exceptions to RM in their L2, they should accept them. Raising across pronominal experiencers should, following *f*RM, elicit higher acceptability ratings than raising across lexical DPs. Finally, we expect relatively high acceptance rates with topicalized constructions because these constructions eschew RM via dislocation. For the control condition, i.e., unraised structures, we also expect relatively high acceptance rates because no movement across interveners occurs in the first place. Under Collins' smuggling account, we expect that raising across lexical DPs will, contra RM, be acceptable to the English NS group. Raising across pronominals should result in higher acceptability, relative to lexical DPs, following *f*RM. Topicalized structures should also be accepted, given that these do not violate locality, via dislocation. If English RM effects do not hold when there is an intervening DP experiencer, NSs should give comparable acceptability ratings to raising over experiencer DPs and those with topicalized DPs, ceteris paribus.

Our second research question is more exploratory and asks whether L1 and L2 processing, as measured in global RTs to acceptability judgments, is affected by syntactic intervention in subject raising across DPs (lexical and pronominal) when compared to topicalized DP raising structures. To our knowledge, no direct evidence on the processing of English raising has yet been presented for either an L1 or L2 population, although there is some related evidence from L2 comprehension studies attesting lower accuracy in the comprehension of raising across lexical DPs and significantly better performances with the DP in topic position (Choe, 2016). Following these findings, which suggest that syntactic intervention triggers higher processing loads, we expect to elicit longer RTs to raising across lexical DPs compared to topicalized DP structures in our groups as well. With regard to type of DP interveners, L2 studies have attested a facilitation effect in the comprehension of raising across pronoun DP experiencers (Choe, 2019; Yoshimura & Nakayama, 2019). This is supported by robust findings in L1 English child acquisition of raising (e.g., Choe & O'Grady, 2017) and in L1 English adult processing of *A-bar* dependencies (e.g., Gordon et al., 2004; Warren & Gibson, 2005). In line with this previous literature, we should observe shorter RTs to raising across pronominal vs. lexical interveners, in both L2 speakers and native speakers.

4.1 Methods

4.1.1 *Participants*

A group of 20 adult L2ers of English (*mean age* = 29.5, *SD* = 5.7) and a control group of 14 English NSs (*mean age* = 27.1, *SD* = 6.09) volunteered to take part in the study. The L2 group consisted of L1 Italian speakers recruited in Pescara, Italy

(n = 15) and in New York, USA (n = 5).[2] Most participants in both groups were highly educated at the time of testing, holding a bachelor and/or a master's degree (n = 28). At minimum participants held a high school diploma (n = 6). The L2ers started learning English between 6 and 11 years old (M = 8.05; SD = 1.95) and had studied English for a period ranging from 7 to 19 school years at the time of the study (M = 13.4; SD = 3.1). Among the L2ers, some reported having learned an L3 (n = 12), while some of the participants from the NS control group reported to have learned an L2 (n = 11).

4.1.2 *Proficiency assessment*

English proficiency was assessed through the Michigan Test of English Language Proficiency (MTELP), administered to both groups of participants. The MTELP tests English grammar and vocabulary in an auditory mode. Participants would listen to a question (17) or statement (18) and select the correct answer from a multiple-choice list. The task included a total of 45 items. Below are two sample items whose grammar focus was on pronoun assignment (17) and correlative conjunction (18):

(17) Is Sally's dinner invitation here yet?
  a. Yes, it has.
  b. Yes, she has.
  c. Yes, it is.

(18) Laura won't sing and Cathy won't either.
  a. Neither of them will sing.
  b. One of them will sing.
  c. Both of them will sing.

The results of the MTELP by language group are summarized in Table 2:

Table 2. MTELP results by language group

Group	N	Accuracy
L2 group	20	38/45 (SD = 7)
NS group	14	44/45 (SD = 1)

Mean results indicate that L2 participants were overall highly proficient in English. However, a two-sample t-test analysis comparing the two groups showed that the NS group scored significantly higher than the L2 group ($t(30) = -3$, $p < .05$).

2. Among the participants recruited in US, 3 out of 5 were based in US for less than 6 months when they took part in the study; 2 out of 5 had lived in US the longest, 1 for a period of 4 years and 1 for less that 1.5 years, both traveling to and spending at least 3 months a year in Italy.

4.1.3 *Materials*

An Acceptability Judgment Task (AJT) was administered to both groups. In the experiment, three variables were manipulated in a 2 × 2 × 4 design: (i) Structure (Raising vs. Unraised), (ii) Grammaticality (Grammatical vs. Ungrammatical) and (iii) Type/Position of DP Experiencer (No Exp. vs. Topic Exp. vs. Pronoun Exp. vs. Lexical Exp.). The task included a total of 160 English sentences (80 grammatical and 80 ungrammatical) including 40 raising structures, 40 unraised, and 80 fillers; each construction type was distributed evenly across grammatical and ungrammatical forms. Both the experimental and the control stimuli included the verb *seem* + *to be* + an Adjective Phrase (AP) complement. The following adjectives were used in the AP complements: *tough, busy, polite, mad, excited, respectful, happy, aware, thankful, proud* and *bored*. Table 3 shows a sample of the grammatical stimuli by condition.[3]

Table 3. Sample of experimental and control grammatical stimuli by condition

Condition	Raising	Unraised
No Exp.	The journalists seem to be busy during the presidential campaign	It seems that the social workers are proud of the children's progress
Topic Exp.	**To the artist,** the curators seem to be polite with the guests	**To the diplomat,** it seems that the ambassadors are bored during the meeting
Pronoun Exp.	The doctors seem **to him** to be respectful of the protocol	It seems **to him** that the passengers are aware of the flight's delay
Lexical Exp.	The customers seem **to the manager to be happy about the sales**	It seems **to the landlord** that the tenants are respectful of the laundry room

In order to compare sentences with experiencers (which would be ungrammatical in the L1, Italian) to other types of ungrammatical sentences, items were created by switching the order of V (to be) and its AP complement (e.g., busy) resulting in sentences such as **John seems busy to be*. The variable was included to ensure that participants were sensitive to ungrammaticality in the L2. The 80 filler items consisted of 40 grammatical and 40 ungrammatical sentences: subject and object relative clauses and raising structures with the predicate *tend to.*

3. The reader is reminded that the relevant sentences in this study are raising structures considered to be grammatical in the theoretical literature, rather than ungrammatical sentences, which are the more common types tested in L2 studies.

4.1.4 *Procedure*

The task was built in E-Prime and administered on a tablet with a keyboard. Each sentence was visually and auditorily presented one at a time while global RTs were recorded. Participants were tested individually and were asked to judge the acceptability of each sentence on a 6-point Likert scale, with 1 and 2 labeled *totally unacceptable*, intermediate ratings of 3 and 4 labeled *not completely acceptable* and 5 and 6 labeled *totally acceptable*. Participants were instructed not to judge the sentences according to school grammar nor to the plausibility of the meaning of the sentence. During the task, participants were asked to use only their dominant hand on the keyboard in order to avoid the use of the two ends of the scale as defaults.

Participants were not allowed to give their responses until the end of the audio, at which point RT measurement began. The experiment was preceded by a warm-up session during which participants were shown three stimuli, none of which included a raising-type structure. The AJT included two sections divided by a one-minute break with the entire task lasting approximately 20 minutes. At the end of the task, all participants took part in the proficiency test and completed a language background questionnaire.

4.1.5 *Data analyses*

The data was analyzed using linear mixed effects models (Bates et al., 2015; Kuznetsova et al., 2017) in RStudio (RStudio Team, 2016). In order to verify whether participants were sensitive to ungrammaticality, mean ratings were first analyzed in a 2 × 2 model (Model 1) with Language Group (L2ers vs. NSs) and Grammaticality (Grammatical vs. Ungrammatical) serving as fixed variables and random intercepts for subjects and items. RTs were first log transformed to normalize the data; extreme outliers by subjects were removed, resulting in 2.05% of data loss. Log-transformed RTs were fitted in the same model described above for the mean ratings, following the same procedure. A summary of the fixed effects and significant interactions from Model 1 on mean ratings and log RTs respectively can be found in Section 4.2.

To compare structures with a DP experiencer in an intervening position versus structures with a topicalized DP in a non-intervening position, mean ratings and log RTs for grammatical raising and unraised structures were separately analyzed in a 2 × 2 × 4 model (Model 2) with Language Group (L2ers vs. NSs), Structure (Raising vs. Unraised) and DP Experiencer type/position (No Exp. vs. Topic Exp. vs. Pronoun Exp. vs. Lexical Exp.) as fixed variables and random intercepts for subjects and items. Log-transformed RTs were fitted in the same model described above for the mean ratings, following the same procedure. A first model on log RTs was compared to a second model from which non-significant three-way interactions were excluded. The two models did not significantly differ when compared using a Likelihood Ratio Test, which did not show a significant reduction in deviance, χ^2 (3) = 2.29, p = .5. For this reason, the reduced model without non-significant three-way interactions was kept. A summary of the main effects and significant interactions from Model 2 for mean ratings and from the reduced Model 2 for log RTs is provided in Section 4.2.

4.2 Results

4.2.1 *Acceptability judgments*

Table 4 reports descriptive statistics of mean acceptability scores and SDs by condition and language group. As can be seen in this table, both groups judged ungrammatical structures across conditions as unacceptable, with means tending towards the lower end of the scale.

Table 4. Acceptability mean scores and SDs of structures by condition and language group

	L2 speakers				**Native speakers**			
	Raising		**Unraised**		**Raising**		**Unraised**	
	Gram.	**Ungram.**	**Gram.**	**Ungram.**	**Gram.**	**Ungram.**	**Gram.**	**Ungram.**
No Exp.								
Mean	5.55	2.97	5.42	3.28	5.81	2.58	5.68	3.10
SD	0.82	1.33	0.85	1.62	0.42	1.23	0.55	1.49
Topic Exp.								
Mean	4.76	2.62	4.86	3.38	5.52	2.61	5.44	3.42
SD	1.17	1.31	1.16	1.4	0.65	1.19	0.71	1.47
Pronoun Exp.								
Mean	4.38	2.35	4.94	3.62	4.32	2.20	5.57	3.35
SD	1.44	1.23	1.28	1.50	1.44	1.18	0.69	1.45
Lexical Exp.								
Mean	4.20	2.49	5.20	3.54	4.28	2.11	5.34	3.41
SD	1.44	1.30	0.97	1.51	1.32	1.11	0.84	1.61

Table 5 presents a summary of main effects and interactions from Model 1 on mean acceptability ratings.[4] In the model, factors were sum coded and the intercept is the

4. A reviewer raised the concern that our participant numbers are low for an experimental study. However, Brysbaert and Stevens (2018) have argued that, in mixed effects analyses, power depends on the total number of data points used in the analysis. The current study counts a total of 2720 data points for the acceptability judgments and a total of 2664 data points for RTs, excluding outliers. Brysbaert (2019) argued that power is also related to reliability of the dependent variables. The reliabilities of the response variables, as measured by Intraclass Correlations (ICC2), show high reliability overall (ICC2's > 0.8) by participants, and high reliability by participants for the within conditions (ICC2's > 0.8). Reliability by participants for the between conditions measured by ICC2 was greater than 0.8 for the log RTs dependent variable and lower than 0.8 for the acceptability responses dependent variable.

grand mean of the acceptability ratings. The model shows a non-significant main effect of Language group (*est.* = −0.03, *SE* = 0.07, *t* = 31.98, *p* = .6), a significant main effect of Grammaticality (*est.* = 1.05, SE 0.02, *t* = 42.12, *p* < .001) and a significant interaction of Language group x Grammaticality (*est.* = −0.12, *SE* = 0.02, *t* = −5.30, *p* < .001).

Table 5. Summary of main effects and interactions from MODEL 1 on mean ratings (Response ~ Language * Grammaticality + (1|Subject) + (1|Item))

Fixed effects	Estimate	Std. error	df	t-value	*p*-value
Intercept	4.01	0.10	26.73	39.64	*p*<.001 *
Language Group	−0.03	0.07	31.98	−0.51	*p* = .6
Grammaticality	1.05	0.02	2672.06	42.12	*p*<.001 *
Language x Grammaticality	−0.12	0.02	2675.00	−5.30	*p*<.001 *

* = Statistically significant; † = marginal

To explore the significant interaction, pairwise comparisons were performed using the FDR correction. Pairwise comparisons revealed that L2 speakers clearly differentiated structures by grammaticality, as their judgments towards ungrammatical structures did not significantly differ from those provided by the NS group (*p* = .2); however, the L2 speakers provided significantly lower acceptability ratings towards the grammatical structures when compared to those given by the NSs (*p*<.05). This result can also be seen in Figure 2 below:

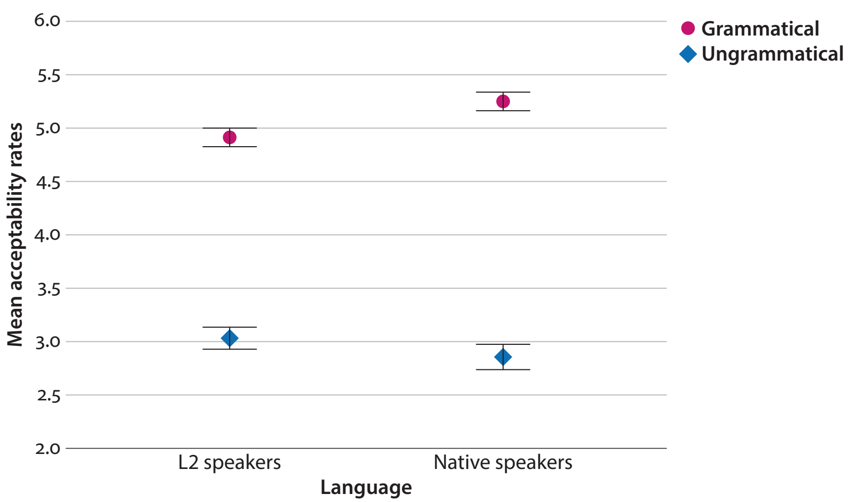


Figure 2. Mean acceptability ratings and 95% CI of raising and unraised structures by grammaticality and language group

A closer look at the critical sentences by DP experiencer condition (Figure 3) reveals that structures with raising over experiencer DPs were not highly rated, with both groups giving lexical and pronoun DPs ratings averaging only slightly above 4 (*not completely acceptable*). Compare this to mean acceptability of unraised structures, averaging around 5 and above (*completely acceptable*) in both L2 and NS groups. Acceptability also increased in the Topic condition (i.e., with a dislocated, non-intervening experiencer), especially for the NS group, with mean ratings overlapping across structure types.

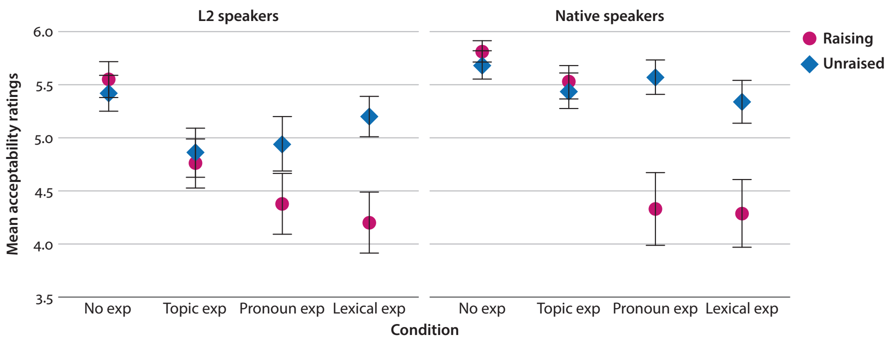


Figure 3. Mean acceptability ratings and 95% CI for grammatical structures by language group and condition

Results anticipated by the means in Figure 3 were statistically confirmed. In Table 6 we report main effects and significant interactions from Model 2, that is, on acceptability judgments to grammatical sentences only, under the four DP experiencer conditions across raising and unraised structures. As in the previous model, the factors were sum coded and the intercept is the grand mean of the acceptability ratings.

The model reports a main effect of Language group reaching significance (*est.* = −0.16, *SE* = 0.08, *t* = −1.89, *p* = .06) and a significant main effect of Structure (*est.* = −0.22, *SE* = 0.02, *t* = −8.08, *p* < .001). The model also shows all main effects of DP experiencer type/position conditions to be significant (*p* < .001) except for the main effect of the Topic Exp. condition (*p* = .1). The model reports several significant two-way interactions of Language x Topic Exp. (est. = −0.16, *SE* = 0.04, *t* = −3.73, *p* < .001), Language x Lexical Exp. (*est.* = 0.11, *SE* = 0.04, *t* = 2.43, *p* < .05) and Structure interacting significantly with all DP experiencer type/position conditions (Table 6). Finally, the model shows that the three-way interaction Language x Structure x Topic reached significance (*est.* = −0.08, *SE* = 0.04, *t* = −1.78, *p* = .07) while the three-way interaction of Language x Structure x Pronoun Exp. was significant (*est.* = 0.13, *SE* = 0.04, *t* = 2.98, *p* < .01).

Table 6. Summary of main effects and significant interactions from Model 2 on mean ratings (Response ~ Language*Structure*DP experiencer + (1|Subject) + (1|Item))

Fixed effects	Estimate	Std. error	df	t-value	*p*-value	
Intercept	5.08	0.08	31.71	56.74	*p*<.001	*
Language Group	−0.16	0.08	31.97	−1.89	*p*=,06	†
Structure	−0.22	0.02	21.38	−8.08	*p*<.001	*
DP experiencer type/position						
No Exp.	0.53	0.04	1235.84	11.74	*p*<.001	*
Topic Exp.	0.06	0.04	1235.84	1.44	*p* = .l	
Pronoun Exp.	−0.27	0.04	1235.84	−6.07	*p*<.001	*
Lexical Exp.	−0.32	0.04	152.52	−6.98	*p*<.00l	*
Interactions						
Language × Topic Exp.	−0.16	0.04	1303.83	−3.73	*p*<.001	*
Language × Lexical Exp.	0.11	0.04	1303.83	2.43	*p*<.05	*
Structure × No Exp.	0.29	0.04	1235.84	6.37	*p*<.00l	*
Structure × Topic Exp.	0.22	0.04	1235.84	4.87	*p*<.001	*
Structure × Pronoun Exp.	−0.22	0.04	1235.84	−4.92	*p*<.001	*
Structure × Lexical Exp.	−0.28	0.04	152.52	−6.20	*p*<.001	*
Language × Structure × Topic Exp.	−0.08	0.04	1303.83	−1.78	*p* = .07	†
Language × Structure × Pronoun Exp.	0.13	0.04	1303.83	2.98	*p*<.01	*

* = Statistically significant; † = marginal

We performed pairwise comparisons of the significant interactions using the FDR correction. Results show that L2 speakers accepted topicalized structures at a significant lower rate than NSs across raising and unraised structures ($p < .01$), while no significant difference in mean ratings emerged across groups for raising and unraised structures with a Lexical Exp. ($p = .6$). Pairwise comparisons following significant interaction of Structure x each DP experiencer type/position condition show that acceptance of items without an experiencer was not significantly different across raising/unraised conditions ($p = .1$); the same result emerged for topicalized structures ($p = .8$). By contrast, raising across lexical DPs obtained significantly lower mean ratings in both language groups when compared to unraised items under the same condition ($p < .001$) and the same result emerged with raising across a pronoun experiencer ($p < .001$). From pairwise comparisons following three-way significant interactions it emerged that L2 speakers accepted topicalized structures at a significantly lower rate than the NS group in both raising ($p < .001$) and unraised ($p < .05$) structures, while no significant difference emerged across raising and unraised structures with a Topic Exp. within each language group. By contrast, raising across a pronoun experiencer received significantly lower acceptance rates than unraised structures with a pronoun experiencer by both L2 speakers ($p < .01$) and NSs ($p < .001$), while no significant difference in mean ratings emerged within raising and unraised structures with a pronoun experiencer across language groups.

Planned comparisons were further performed to verify our predictions on the expected higher acceptability of raising across pronoun experiencers when compared to raising across a lexical experiencer and on the expected higher acceptability of raising with a Topic Exp. when compared to raising across a lexical/pronoun experiencer. Results show that, contra our predictions, no significant difference in mean acceptance emerged between raising across a pronoun experiencer and across a lexical experiencer either in L2 speakers ($p = .2$) or NSs ($p = .8$). However, as predicted, planned comparisons confirmed that acceptance of raising across lexical experiencers was significantly lower than when experiencers were topicalized, a result that emerged in both L2 speakers ($p < .001$) and NSs ($p < .0001$). The same results emerged when comparing raising across a pronoun experiencer vs. raising with a topicalized experiencer: the acceptance of the former was significantly lower than the acceptance of the latter by both L2 speakers ($p < .001$) and NSs ($p < .0001$).

Overall, analyses showed a significant sensitivity towards ungrammaticality. In addition, both groups showed higher acceptability of unraised structures with experiencers when compared to their raised counterparts, showing sensitivity to RM. Both groups showed only marginal acceptability of raising across a lexical and a pronoun experiencer, and no significant difference in acceptance of lexical vs. pronoun experiencers. In the topic condition, both L2 and NS groups judged constructions as acceptable and provided higher acceptability rates than raising across DP experiencers. However, there was a significant difference between NS and L2 groups, with higher acceptability for the former.

4.2.2 *Response times (RTs)*

Table 7 shows descriptive statistics of log RTs by condition and language group. A cursory look at the means might suggest that ungrammaticality is not a predictor of longer RTs, or higher processing load for either group. Descriptive data further seems to indicate that the L2 group did not show shorter RTs in judging unraised structures, i.e., suggesting no difference in processing load across structure types. This result seems to contrast with that observed in the NS group. Means also suggest a difference between groups in RTs to raising with a DP experiencer. In the L2 group, raising structures with an intervening DP elicited comparable RTs to raising in Topic construction, indicating comparable processing loads. This result is not replicated in the native controls, who showed a processing advantage in response to the topicalized structures, as evidenced by shorter RTs.

Table 7. Log mean RTs and SDs by condition and language group

	L2 speakers				Native speakers			
	Raising		Unraised		Raising		Unraised	
	Gram.	Ungram.	Gram.	Ungram.	Gram.	Ungram.	Gram.	Ungram.
No Exp.								
Mean	6.95	7.31	7.16	7.28	5.99	6.97	6.03	6.60
SD	1.23	1.23	1.08	1.15	0.96	1.06	1.17	1.27
Topic Exp.								
Mean	7.51	7.26	7.31	7.65	6.58	6.62	6.48	7.10
SD	1.17	1.12	1.26	1.04	1.17	1.26	1.17	1.08
Pronoun Exp.								
Mean	7.42	7.18	7.21	7.29	6.96	6.64	6.37	6.79
SD	1.18	1.14	1.30	1.21	1.27	1.20	1.18	1.31
Lexical Exp.								
Mean	7.46	7.42	7.26	7.61	6.99	6.64	6.61	6.83
SD	1.16	1.29	1.17	1.07	1.25	1.17	1.10	1.41

In Table 8 we report a summary of the main effects and interactions from Model 1 on log RTs to judgments.[5] In the model, factors were sum coded and the intercept is the grand mean of the log RTs. The observations gleaned from Table 7 are partially confirmed. The model shows a significant main effect of Language group (*est.* = 0.34, $SE = 0.11$, $t = 3.00$, $p < .01$), a significant main effect of Grammaticality (*est.* = −0.09, $SE = 0.02$, $t = -4.66$, $p < .001$) and a significant interaction of Language group x Grammaticality (*est.* = 0.04, $SE = 0.02$, $t = 2.28$, $p < .05$).

Table 8. Summary of main effects and interactions from Model 1 on log RTs (logRTs ~ Language * Grammaticality + (1|Subject) + (1|Item))

Fixed effects	Estimate	Std. error	df	t-value	*p*-value	
Intercept	6.99	0.11	36.15	58.32	*p*<.001	*
Language Group	0.34	0.11	31.99	3.00	*p*<.0l	*
Grammaticality	−0.09	0.02	2509.83	−4.66	*p*<.001	*
Language x Grammaticality	0.04	0.02	2619.15	2.28	*p*<.05	*

* = Statistically significant; † = marginal

Pairwise comparisons were performed using the FDR correction to explore the significant interaction. Results from the comparisons show that L2 speakers were sensitive to ungrammaticality and that they took significantly more time to judge ungrammatical sentences than grammatical ones ($p = .05$), in line with the result

5. See footnote 4.

that also emerged with the NS controls ($p < .001$). However, results from pairwise comparisons also show that the L2 speakers were generally significantly slower than the NSs when judging grammatical structures as well ($p < .01$).

In Table 9 we report the main effects and significant interactions from reduced Model 2 on log RTs to grammatical sentences under the four DP experiencer conditions,across raising and unraised structures. Factors were sum coded and the intercept is the grand mean of the log RTs. The model reports a significant main effect of Language group (*est.* = 0.39, *SE* = 0.12, $t = 3.23$, $p < .01$) and a significant main effect of Structure (*est.* = 0.08, *SE* = 0.02, $t = 3.15$, $p < .01$). The model also shows main effects of No Exp. and Lexical Exp. to be significant ($p < .001$) and main effects of Topic Exp. and Pronoun Exp. to reach significance (Table 9). The model reports several significant two-way interactions: Language x No Exp. (*est.* = 0.12, *SE* = 0.04, $t = 2.61$, $p < .01$), Language x Lexical Exp. (*est.* = −0.10, *SE* = 0.04, $t = -2.32$, $p < .05$) and Structure interacting significantly with No Exp. and Pronoun Exp. conditions (Table 9).

Table 9. Summary of main effects and significant interactions from reduced Model 2 on log RTs (logRTs ~ Language * Structure*DP Experiencer + (1|Subject) + (1|Item))

Fixed effects	Estimate	Std.error	df	t-value	*p*-value	
Intercept	6.90	0.12	34.01	56.87	*p*<.001	*
Language Group	0.39	0.12	34.01	3.23	*p*<.01	*
Structure	0.08	0.02	1291.05	3.15	*p*<.01	*
DP experiencer type/position						
No Exp.	−0.35	0.04	1292.23	−7.53	*p*<.001	*
Topic Exp.	0.07	0.04	1292.07	1.68	*p* = .09	†
Pronoun Exp.	0.09	0.04	1292.07	1.95	*p* = .05	†
Lexical Exp.	0.18	0.04	1292.05	4.03	*p*<.001	*
Interactions						
Language x No Exp.	0.12	0.04	1292.23	2.61	*p*<.01	*
Language x Lexical Exp.	−0.10	0.04	1292.05	−2.32	*p*<.05	*
Structure x No Exp.	−0.15	0.04	1292.14	−3.38	*p*<.001	*
Structure x Pronoun Exp.	0.10	0.04	1292.04	2.25	*p*<.05	*

* = Statistically significant; † = marginal

To explore the significant interactions, pairwise comparisons were performed using the FDR correction. Results showed that the L2 speakers took significantly longer RTs than NSs to judge items without an experiencer ($p < .01$) and with a lexical experiencer ($p = .05$) across raising and unraised conditions. From the pairwise comparisons following the significant interactions of Structure x No Exp. and Structure x Pronoun Exp., it emerged that mean logRTs were not significantly different when judging raising and unraised structures without an experiencer ($p = .2$).

On the other hand, raising structures across a pronoun experiencer elicited significantly longer RTs than unraised structures under the same condition ($p < .01$).

Planned comparisons were also performed to verify our predictions on (i) the expected longer RTs when judging raising across a lexical/pronominal experiencer compared to the unraised counterpart under the same conditions; (ii) the expected shorter RTs when judging raising across pronoun vs. lexical experiencers; and (iii) on the expected shorter RTs when judging raising with a Topic experiencer. vs. *across* a DP experiencer. For (i), results show that, for native speakers, raising across a lexical/pronominal experiencer elicited significantly longer RTs than the unraised counterpart under the same condition ($p < .05$). For the L2 group, this difference was not significant ($p < .2$). For (ii), planned comparisons showed that, in line with the mean ratings and contra our predictions, an intervening pronominal experiencer did not trigger shorter RTs than an intervening lexical experiencer in either group ($p > .05$). For (iii), planned comparisons show that raising across a lexical experiencer elicited shorter logRTs than with a topicalized experiencer, but only in the NSs group ($p < .05$). This difference proved non-significant for the L2 speakers ($p = .9$).

In sum, statistical analyses of global RTs indicate that overall ungrammaticality was a predictor of longer RTs in both language groups, although the L2 speakers were generally slower than the NS controls regardless of grammaticality. Our results further indicated that subject raising across DP experiencers (both lexical and pronominal) compared to unraised structures under the same conditions elicited higher mean RTs only in native speakers. This is contrary to the findings in the acceptability judgment task where the difference was significant for both groups. Finally, results indicated that the L2 group did not show any reduction in RTs with an experiencer dislocated in topic position, while the NS group showed a robust processing advantage in this respect.

5. Discussion

The present study aimed to investigate the effects of *Relativized Minimality* (Rizzi, 1990 et seq.) in English raising structures in L2 speakers and native speakers of English. Of particular interest to us were the exceptions to RM found in subject raising over experiencer DPs. We contrasted English raising structures with intervening and non- intervening (topicalized) DP experiencers, and with lexical and pronominal DPs. Specifically, we asked whether: (i) exceptions to RM are acquirable i.e., intervention effects can be overridden in the L2, (ii) processing of such exceptions manifests differently in L2 speakers and native speakers of English, and (iii) we also tested the long-held assumption that English is exceptionally insensitive to intervention in subject raising constructions is empirically valid. We discuss our results to these questions below.

5.1 Intervention effects in the representation of L1 and L2 English raising

Results from the AJT revealed that *Relativized Minimality* is indeed operative in the representation of the L2. This is evidenced by significantly lower acceptance ratings for raised vs. unraised structures involving a DP experiencer, showing the expected intervention effect when subjects were *A*-moved in raising, in contrast to unraised structures where no syntactic movement occurs.

While these findings show that L2 speakers are sensitive to intervention effects, something that might be expected given the universality of RM, our results also show that *exceptions* to RM are learnable in L2, even if they do not hold in the L1. In fact, these exceptions yield comparable acceptability ratings in the L2 group and the NS group. In the raised condition, the L2 group rated subject raising across lexical DP experiencers as *not completely acceptable*, with mean ratings that were significantly lower than those for DPs in topic position. This result echoes previous L2 studies testing comprehension (Choe, 2016, 2019; Yoshimura & Nakayama, 2019); L2 responses indicate that advanced speakers of English can indeed acquire exceptions to a principle that is operative in both the L1 and the L2. Turning now to the NSs, responses show only marginal acceptance of these structures; we can say that while intervention effects are perhaps mitigated, they are certainly not absent. Thus, it seems that crossing over DP experiencers in English still evokes *mild* intervention effects. The fact that acceptability judgment ratings were not significantly different across the two groups for these structures indicates high comparability in the L1 English system of the NSs and the L2 English system of the L2ers. Recall that the equivalent structures in Italian are ungrammatical. Therefore, if direct transfer from the L1 had played a role, the L2 group should have rejected these structures to a higher degree.

As predicted, a DP experiencer in topic position increased acceptability ratings in the L2 group as well as in the NS group, although the L2 group did not show the same degree of preference for the topicalized structure as did the NS. This latter result suggests that, at least for some of this particular group of speakers, English-style topicalization may not be fully acquired yet as indicated by a significant three-way interaction of Language × Structure × Topic Exp. and follow-up pairwise comparisons.[6] Recall that according to Cinque (1990), English topicalization differs from Italian CLLD in that the former derives directly via *wh*-movement and involves a gap, and although both require the topic to be connected to the comment, the bindee is of a different nature in the two languages, i.e., a resumptive clitic in Italian and a null OP in English (Rizzi, 1997).

6. The majority of the L2 speakers (17 out of 20) did not fully accept at least one out of five topicalization structures, and four of them rated at least three or more of these structures as either completely unacceptable or not completely acceptable.

Although the group tested in our study was highly proficient in English, this result is not totally surprising and is, in fact, in line with the existing literature on the L2 acquisition of structures operating at the syntax-discourse interface. Lowered acceptability of topicalization structures is also reported by Slabakova (2015), who found that while L1 English NS of L2 Spanish showed mastery of Spanish CLLD, highly proficient L1-Spanish learners of English did not successfully acquire English topicalization yet (see Smeets, 2019 for similar results in L2 Italian). Slabakova (2015) interpreted the result as both a L1 transfer effect from Spanish and as an effect of the low frequency of the structure in the L2 input. Our results strongly support Slabakova (2015) hypothesis, given the equivalency between Spanish/Italian CLLD. Further research is needed in order to confirm whether the same holds for CLLD raising structures in the opposite learning direction, i.e., L1 English/L2 Italian as expected according to the existing literature (Smeets, 2019).

The L2 group showed sensitivity to intervention effects regardless of DP type, a result that also emerged in the NSs, contra our prediction which followed the *f*RM approach. Note that this result also contrasts with existing L2 literature on the comprehension of English raising. Choe (2019) did find a facilitation effect with intervening pronominals, i.e., L2 accuracy improved with non-lexical experiencer, in line with L1 child English (Choe & O'Grady, 2017) and L1 results on the comprehension of *A-bar* dependencies (e.g., Friedmann et al., 2009). Considering that Choe (2019) tested for comprehension rather than acceptability, we might argue that a manipulation of the lexical category helps L2 speakers' comprehension, but that it might not have an effect in the grammatical representation of the structure.

Because the NS group did not show a difference in acceptability between pronominal and lexical DPs, the L2 results cannot be ascribed to developmental factors. Recall that the *f*RM correctly predicts differential treatment of lexical vs. pronominal interveners in the case of *A-bar* movement (see Belletti & Rizzi, 2013 for discussion). A potential explanation of our results could be that in contrast to *A-bar* movement phenomena, lexical category is not a relevant morphosyntactic feature in *A*-movement, leading to comparable acceptability for pronominal and lexical interveners. This hypothesis could be further investigated in English, e.g., by collecting judgments on different *A*-movement constructions that are not said to be exceptions to RM and manipulating the lexical category of the intervener DP.

With regard to the empirical validity of structures that constitute exceptions to intervention effects, our study provided mixed evidence. Only *mild* intervention effects were observed, as our control NSs judged the critical structures as *not completely acceptable* regardless of whether the DP experiencer was lexical or pronominal. Thus, if our results prove to be generalizable, the exceptionality of *smuggling* cases (Collins, 2005) may not be as strong as previously claimed. Alternatively, it may be the case that smuggling comes at a cost, lowering intervention effects.

In support of the former hypothesis, Hartman (2011) claims that English, just like Romance languages, shows syntactic intervention effects in subject raising with experiencers, as it does in other *A*-dependencies, a claim that would need to be tested by comparing judgments across different *A*-dependencies with intervening DPs.

Further research is necessary to trace the development of L2 acquisition of exceptions to RM. Our study did not include lower proficiency level L2 speakers and can therefore not verify whether at earlier stages of development L2 speakers would significantly diverge from the native speakers. However, this would be expected, given that previous studies have shown negative correlations between proficiency levels and intervention effects in L2 English (e.g., Choe, 2016, 2019).

5.2 Intervention effects in L1 and L2 processing of English raising

While similar patterns across the two language groups were found in the acceptability judgments taks, the processing results show some differences. For example, overall the L2 group showed generally longer RTs than NSs, an outcome that is expected. Difference in sensitivity between the two groups could be seen in the comparison of grammatical and ungrammatical sentence processing. While ungrammaticality was indeed a predictor of longer RTs in both groups, results indicated that the differential between grammatical and ungrammatical was smaller for the L2 group.

Intervening DPs did not have a comparable effect in L1 and L2 processing of English raising. The NS group but not the L2 group showed significantly longer RTs in raised structures across experiencers when compared to the unraised counterparts. Our NS results align with some adult processing data collected by Hirsch and Wexler (2007) in a L1 child English study on raising. Hirsch and Wexler reported that the adult response time to a comprehension question during a picture matching task was 1079 ms slower in the critical raising structures across an experiencer than in the equivalent unraised structure. The result also patterns with the adult processing literature on English *A-bar* dependencies (Gibson, 1998; Gordon et al., 2001; Warren & Gibson, 2002), whose findings agree on the negative effects that intervening DPs have in the processing of long-distance dependencies. The fact that L2 speakers did not show native-like processing suggests that, while at the representational level native and non-native grammars eventually converge, L2 speakers do not have the same processing resources as the NSs, as proposed by Hopp (2009).

In contrast to what we found in the AJT, topicalized structures lacking an intervening DP, facilitated processing only in the NS group, who showed significantly shorter RTs viz-a-viz structures with intervening DPs. This facilitation was not found in the experimental group, as L2ers showed comparable RTs to raising structures involving a DP experiencer in either an intervening or non- intervening position. The difference between the two groups can be ascribed to greater processing facility for these structures in the NS group compared to the L2 group, possibly owing to the latter not yet fully mastering L2 topicalization.

Another possible explanation for this difference comes from the Interface Hypothesis (IH, Sorace, 2011), which posits that structures occurring at the syntax/discourse interface, such as topics, are harder to master even at higher levels of L2 proficiency, when aspects of L2 core grammar are already fully acquired. The increased difficulty for interface structures is potentially due to processing factors. Sorace (2011) argues that bilinguals might be less "efficient" than monolinguals at integrating information from different domains of language when processing structures operate at the interface: processing factors could be responsible for the different outcomes across native and non-native groups, along with any L1 to L2 grammar influences. Sorace argues that the difference between native and non-native access to processing resources can be traced to (i) a less fine-grained knowledge of L2 grammatical properties and constraints and/or (ii) a less "automatic" access to L2 knowledge during language processing. Our findings point in the direction of both possibilities, given the contrast that emerged between the acceptability ratings and the RTs to topicalization for the L2 group, a contrast that did not emerge for the NS group.

Raising across pronominal DPs did not produce shorter RTs when compared to the lexical condition in either group. This aligns with the result found in the AJT which showed comparable ratings for lexical and pronominal interveners contra what would be expected under the *f*RM . The absence of a RT differential between pronominal and lexical DPs found in the NSs contradicts findings in much of the processing literature on *A-bar* dependencies (e.g., Gordon et al., 2004; Warren & Gibson, 2005). This result might be taken as further indication that different types of interveners have different effects on the processing of *A* vs. *A-bar* movement constructions, a hypothesis that would need to be investigated further in future inquiries on native and non-native processing.

6. Conclusions

Our study showed that while a universal locality principle, such as *Relativized Minimality,* governs L2 grammars, exceptions to such principles found only in the L2 and not the L1 can indeed be acquired to native-like levels. By using raising constructions, we have provided evidence that intervention effects are also found in the L2 representation of *A*-dependencies. On the other hand, this was not reflected in the L2 processing of these structures, as measured in global RTs. It is possible that a more sensitive measure such as self-paced reading or eye-tracking would be able to better capture this effect in L2 processing. Interestingly, the predicted difference between lexical and pronominal interveners was not supported by either the AJT or RT results in either group, leading us to speculate that this difference may only hold in *A-bar* constructions or only in comprehension tasks. Finally, the predicted difference between topicalized vs. non-topicalized (i.e. intervening) experiencers was found for both groups in the AJT task; however, a processing advantage of this phenomenon was only found for the NSs. We have speculated that this difference in L1/L2 processing may be explained the IH which claims that increased L2 processing loads for structures at the interface of discourse and syntax occur even at higher levels of proficiency. Finally, our study sheds some light on the purported exceptionality of English raising across intervening DPs and provided evidence that native speakers of English still experience *mild* intervention effects with these structures.

Overall, our study suggests comparable grammars for L1 and (advanced) L2 speakers, respecting RM and generally showing similar patterns of intervention effects. However, our study showed significant differences in the RTs of English raising between L2 and L1 groups, confirming that the two groups differ primarily in the processing of these structures. Further research using online measures may lead to different outcomes and conclusions.

References

Baltin, M. (2001). A-movements. In M. Baltin & C. Collins (Eds.), *The handbook of contemporary syntactic theory* (pp. 226–254). Blackwell. https://doi.org/10.1002/9780470756416.ch8

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. https://doi.org/10.18637/jss.v067.i01

Boeckx, C. (1999). Conflicting C-command requirements. *Studia Linguistica*. 53, 227–250. https://doi.org/10.1111/1467-9582.00046

Boeckx, C. (2008). *Aspects of the syntax of agreement*. Routledge. https://doi.org/10.4324/9780203930335

Belletti, A., & Rizzi, L. (2013). Intervention in grammar and processing. In I. Caponigro & C. Cecchetto (Eds.), *From grammar to meaning* (pp. 294–311). Cambridge University Press. https://doi.org/10.1017/CBO9781139519328.015

Belletti, A., Friedmann, N., Brunato, D., & Rizzi, L. (2012). Does gender make a difference? Comparing the effect of gender on children's comprehension of relative clauses in Hebrew and Italian. *Lingua*, 122, 1053–1069. https://doi.org/10.1016/j.lingua.2012.02.007

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition*, 2, 16. https://doi.org/10.5334/joc.72

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, 1, 9. https://doi.org/10.5334/joc.10

Choe, J. (2016). Intervention effects in the comprehension of English raising constructions by Korean learners. *Korean Journal of Linguistics*, 41, 741–759. https://doi.org/10.18855/lisoko.2016.41.4.009
Choe, J. (2019). Effect of NP type on L2 Raising acquisition. In M. M. Brown & B. Dailey (Eds.), *Proceedings of the 43rd annual Boston University Conference on Language Development* (pp. 165–174). Cascadilla Press.
Choe, J., O'Grady, W. (2017). Asymmetry in children's comprehension of raising. *Journal of Child Language*, 44, 752–765. . https://doi.org/10.1017/S0305000916000179
Chomsky, N. (1981). *Lectures on government and binding*. Foris.
Chomsky, N. (1995). *The minimalist program*. The MIT Press.
Chomsky, N. (2000). Minimalist inquiries: The framework. In R. Martin, D. Michaels, & J. Uriagereka (Eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik* (pp. 89–155). The MIT Press.
Cinque, G. (1990). *Types of A-bar dependencies*. The MIT Press.
Collins, C. (2005). A smuggling approach to raising in English. *Linguistic Inquiry*, 36, 289–298. https://doi.org/10.1162/0024389053710701
Friedmann, N., Belletti, A., & Rizzi, L. (2009). Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua*, 119, 67–88. https://doi.org/10.1016/j.lingua.2008.09.002
Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1–76. https://doi.org/10.1016/S0010-0277(98)00034-1
Gordon, P. C., Randall H., & Johnson, M. (2001). Memory Interference during language processing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 27, 1411–1423.
Gordon, P., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language*, 51, 97–114. https://doi.org/10.1016/j.jml.2004.02.003
Hartman, J. (2011). Non-intervention in A-movement, some cross-constructional and cross-linguistic considerations. *Linguistic Variation*, 11, 121–148. https://doi.org/10.1075/lv.11.2.01har
Hirsch, C. (2011). The acquisition of raising (Unpublished doctoral dissertation). MIT.
Hirsch, C., & Wexler, K. (2007). The late acquisition of raising: What children seem to think about seem. In W. D. Davies & S. Dubinsky (Eds.), *New horizons in the analysis of control and raising* (pp. 35–70). Springer. https://doi.org/10.1007/978-1-4020-6176-9_3
Hopp, H. (2009). The syntax – discourse interface in near-native L2 acquisition: Off-line and on-line performance. *Bilingualism: Language and Cognition*, 12(4), 463–483. https://doi.org/10.1017/S1366728909990253
Ionin, T., & Zyzik, E. (2014). Judgment and interpretation tasks in second language research. *Annual Review of Applied Linguistics*, 34, 1–28. https://doi.org/10.1017/S0267190514000026
Kitahara, H. (1997). *Elementary operations and optimal derivations*. The MIT Press.
Kuznetsova A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82, 1–26. https://doi.org/10.18637/jss.v082.i13
Lardiere, D. (2008). Feature-assembly in second language acquisition. In J. Liceras, H. Zobl, & H. Goodluck (Eds.), *The role of formal features in second language acquisition* (pp. 106–140). Lawrence Erlbaum Associates.
López, L. (2009). *A derivational syntax for information structure*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199557400.001.0001
McGinnis, M. (1998). Locality in A-movement (Unpublished doctoral dissertation). MIT.

Rizzi, L. (1986). On chain formation. In H. Borer (Ed.), *The grammar of pronominal clitics* (pp. 65–95). Academic Press.

Rizzi, L. (1990). *Relativized minimality*. The MIT Press.

Rizzi, L. (1997). The fine structure of the left periphery. In L. Haegeman (Ed.), *Elements of grammar* (pp. 281–337). Kluwer. https://doi.org/10.1007/978-94-011-5420-8_7

Rizzi, L. (2001). Relativized minimality effects. In M. Baltin & C. Collins (Eds.), *The handbook of contemporary syntactic theory* (pp. 89–110). Blackwell. https://doi.org/10.1002/9780470756416.ch4

Rizzi, L. (2004). Locality and left periphery. In A. Belletti (Ed.), *Structures and beyond: The cartography of syntactic structures* (pp. 223–251). Oxford University Press.

Rizzi, L. (2013). Locality. *Lingua*, 130, 169–186. https://doi.org/10.1016/j.lingua.2012.12.002

Rothman, J., & Slabakova, R. (2018). The generative approach to SLA and its place in modern second language studies. *Studies in Second Language Acquisition*, 40, 417–442. https://doi.org/10.1017/S0272263117000134

RStudio Team. (2016). RStudio: Integrated development for R. RStudio. Retrieved on 18 February 2022 from http://www.rstudio.com/

Slabakova, R. (2015). The effect of construction frequency and native transfer on second language knowledge of the syntax – discourse interface. *Applied Psycholinguistics*, 36, 671–699. https://doi.org/10.1017/S0142716413000386

Slabakova, R. (2016). *Second language acquisition*. Oxford University Press.

Smeets, L. (2019). Conditions on L1 transfer in L2 discourse-syntax mappings: The case of clitic left dislocation in Italian and Romanian (Unpublished doctoral dissertation). McGill University.

Sorace, A. (2011). Pinning down the concept of "interface" in bilinguals. *Linguistic Approaches to Bilingualism*, 1, 1–33. https://doi.org/10.1075/lab.1.1.01sor

Starke, M. (2001). Move dissolves into merge: A theory of locality (Unpublished doctoral dissertation). University of Geneva.

Torrego, E. (1996). Experiencers and raising verbs. In R. Freidin (Ed.), *Current issues in comparative grammar* (pp. 101–120). Kluwer. https://doi.org/10.1007/978-94-009-0135-3_5

Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition*, 85, 79–112. https://doi.org/10.1016/S0010-0277(02)00087-2

Warren, T., & Gibson, E. (2005). Effects of NP-type on reading English clefts. *Language and Cognitive Processes*, 20, 751–767. https://doi.org/10.1080/01690960500051055

White, L. (1989). *Universal grammar and second language acquisition*. John Benjamins. https://doi.org/10.1075/lald.1

White, L. (2003). *Second language acquisition and universal grammar*. Cambridge University Press. https://doi.org/10.1017/CBO9780511815065

Xia, V. Y., White, L., & Guzzo, N. B. (2020). Intervention in relative clauses: Effects of relativized minimality on L2 representation and processing. *Second Language Research*. https://doi.org/10.1177/0267658320958742

Yoshimura, N., & Nakayama, M. (2019). Intervention meets transfer in raising constructions. In P. Guijarro-Fuentes & C. Suárez-Gómez (Eds.), *Proceedings of GALA 2017: Language Acquisition and development* (pp. 255–270). Cambridge Scholars.

Explaining the difficulty with L2 acquisition of scope interpretation by speakers of a scope-rigid language

Takayuki Kimura
The University of Tokyo

This study explores the L2 acquisition of quantifier scope in English, a notably difficult property to acquire by speakers of a scope-rigid language like Japanese. This study examines the knowledge of universal quantifiers in English that Japanese Learners of English (JLEs) have, focusing on the distributivity and collectivity of the quantifiers. Results of a picture-based acceptability judgment task showed that JLEs had problems with scope judgments and interpreted *every NP* as a distributive/collective quantifier akin to *all* NP_{PL} in the grammar of English native speakers. Thus, this study implies that issues with the L2 acquisition of scope ambiguity are rooted in a problem with the reassembly of quantificational features of universal quantifiers, in accord with Lardiere (2008, 2009).

Keywords: scope, quantifier, distributivity, collectivity, feature reassembly, uninterpretable feature

1. Introduction

Quantifier scope has been one of the most important topics in theoretical linguistics, especially in the generative approach (e.g., Aoun & Li, 1993; May, 1985; Szabolcsi, 2010, among many others). A sentence with multiple quantifiers such as (1) is ambiguous with respect to scope interpretation. In (1a), the scope interpretation matches the linear order (i.e., *a NP* precedes *every NP*), whereas it is possible to obtain another reading for (1), where the scope interpretation is in an inverse relation to the linear order, as in (1b). The former is called *surface scope* and the latter is referred to as *inverse scope*.

(1) A boy loves every girl.
  a. $\exists>\forall$: There is a boy who loves every girl.
  b. $\forall>\exists$: For each girl *x*, there is a boy who loves *x*.

https://doi.org/10.1075/lald.67.02kim

This sort of sound – meaning mismatch is important not only for linguistic theory but also in the context of L2 theories. Not all languages permit the inverse scope reading (∀>∃) for a sentence. Japanese, for instance, is a representative example of a scope-rigid language (e.g., Takahashi, 2002; Yatsushiro, 2001), such that the inverse scope is not permitted.[1] In Japanese, *dono-NP-mo* (*wh*-NP-conj(unctive))[2] is considered as a universal quantifier, and when it is in the object position, it cannot take scope over a subject quantifier, as in (2).

(2)	Aru	syonen-ga	dono-syozyo-mo	aishiteiru.	(∃>∀, *∀>∃)
	a	boy-NOM	wh-girl-conj	love	
	'A boy loves every girl.'				

Although it is well known that the L2 acquisition of the inverse scope reading is very hard, the source of the difficulty is still highly controversial (for earlier attempts, see e.g., Chu et al., 2013; Chung, 2013; Čulinović, 2013; Kimura, 2019; Marsden, 2004, 2009; Wu & Ionin, 2018). In this paper, I explore the possibility that the feature of the universal quantifier that syntactically triggers the inverse scope interpretation is not acquired.

This paper is organized as follows. The next section explains how the inverse scope reading is derived in English and why it is absent in Japanese. Section 3 reviews previous studies of the L2 acquisition of scope interpretation. Section 4 describes the experiment while Section 5 reports on the behavioral results and statistical analysis. In Section 6, I will discuss the results and their implications, and Section 7 concludes the paper.

2. Syntax of scope

2.1 English

As mentioned earlier, in addition to the linear scope reading, English also permits the inverse scope reading (∀>∃).

A standard approach to inverse scope relies on May's (1985) *Quantifier Raising*, according to which a lower quantifier undergoes covert movement over the subject quantifier, as in (3). Importantly, scope relations are determined on the basis

1. I assume, with Bobaljik and Wurmbrand (2012) among others, that there does not exist a parameter like [+/−inverse scope]. Instead, as we will see later, the lack of the inverse-scope reading results from lexical – syntactic factors.

2. Conjunctives refer to multiple members in a given set (e.g., John ∧ Bill ∧ Mary…).

of structural, or more specifically, c-command, relations (e.g., Aoun & Li, 1993; Kitahara, 1996; May, 1985) such that X has scope over Y if X c-commands Y. Therefore, the (covert) movement of an object Quantifier Phrase (QP) over a subject QP results in the inversion of the original scope relation.

(3) A boy loves every girl.
→ [$_{TP}$ every girl [$_{TP}$ a boy loves *t*]][3]

However, this theory over-generates inverse scope for a sentence like (4); an *all NP*$_{PL}$ object is expected to permit inverse scope, contrary to fact (Ioup, 1975; Vender, 1962).

(4) A boy loves *all* girls. (∃>∀, *∀>∃)

In order to allow inverse scope of *every* and rule out it out for *all*, Beghelli and Stowell (1997) propose a cartographic structure for quantifier scope. The landing sites in the proposed structure are determined on the basis of the features that quantifiers bear. Now, let us consider the featural properties of the relevant quantifiers. Both *every* and *all* are universal quantifiers whose quantificational force ranges over the entire members of the relevant set. A key difference between the two quantifiers lies in *distributivity/collectivity*. As shown in (5a), *every NP* only permits the distributive interpretation (i.e., the boys in the context lifted up a table *individually*) while disallowing the collective interpretation (i.e., the boys in the context lifted up one table *together*).[4] This suggests that *every* NP has a distributive meaning (or a distributive feature), and the distributivity of *every* is incompatible with the collective context. In contrast to *every NP*, *all NP*$_{PL}$ allows both readings, as shown in (5b) (e.g., Lakoff, 1972).

(5) a. *Every* boy lifted up a table. (✓distributive/*collective)
b. *All* the boys lifted up a table. (✓distributive/✓collective)

The basic interpretation of *all NP*$_{PL}$ is collective, and its distributive interpretation is obtained in a sentence with *each* in (6).

(6) *All* the boys **each** lifted up a table.

For these reasons, it has been believed that the distributivity of *all NP*$_{PL}$ in (5b) is merely apparent (pseudo-distributivity), made available via phonologically silent

3. Grey characters are used for covert or phonologically null elements.

4. Note that *everyone* is a well-known counterexample. This issue is a long-standing problem, but we do not yet have a standard explanation for it (Szabolcsi, 2010, p. 121).

each (Beghelli & Stowell, 1997; see also Vendler, 1962), so that we can see *all* NP_{PL} as a collective universal quantifier in nature (I come back to this issue later). On the other hand, *every NP* is a distributive universal quantifier. Recall that *every NP* permits inverse scope, whereas *all* NP_{PL} does not. Beghelli and Stowell (1997) relate these distributivity/collectivity facts to the availability of inverse scope. *All* NP_{PL} inherently specifies a [collective] value for its quantificational feature, as shown in (7a). By contrast, *every NP* lacks such an inherent quantificational value (7b), as opposed to *all* NP_{PL}.[5]

(7) a. $all_{[\text{Q: collective}]}$
    b. $every_{[\text{Q:\_\_\_\_\_\_\_\_}]}$

These featural properties explain the presence/absence of inverse scope for each quantifier if we assume the mechanism of *Agree* in Minimalist syntax (Chomsky, 2001). Under the Minimalist system, an unvalued feature must pick up a value from its valued counterpart via Agree, as in subject movement or *wh*-movement (Chomsky, 2001). For a feature to be valued, movement must take place in order to create a proper c-command relation for the execution of Agree. To illustrate this, see (8). In (8a), Y lacks a value, so that it must be valued via Agree with a corresponding valued feature X. Y's unvalued feature searches down in the c-command domain in the tree, but fails to find the valued goal in the domain. Thus, it moves upward to re-search the goal, as in (8b) (Bošković, 2007). Here, it successfully locates the goal feature and gets valued.

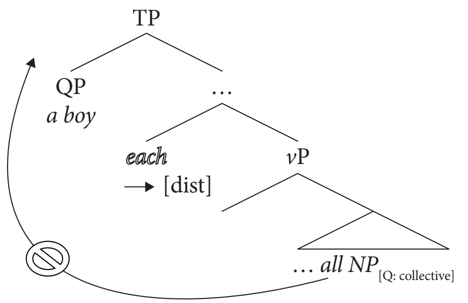
(8) a. $X_{[\text{F: val}]}$ … $Y_{[\text{F: \_\_\_}]}$

    b. → $Y_{[\text{F: \underline{val}}]}$ … $X_{[\text{F: val}]}$ … *t*

Keeping this background in mind, let us return to the featural specifications of the quantifiers. As shown in (7a), *all* NP_{PL} has a specified value, meaning that movement for valuation is unnecessary. The lack of movement of *all* NP_{PL} means that *all* NP_{PL} cannot have scope over the subject quantifier and that inverse scope is unavailable such that distributivity with *all* NP_{PL} is a consequence of the presence of a silent binominal *each*, an operator that distributes either an entity or an event (Beghelli & Stowell, 1997), as schematized below:

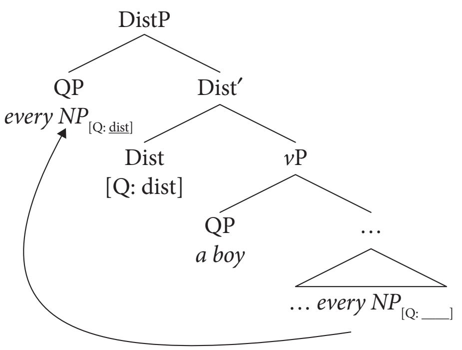
5. *Each* NP has an inherent [distributive] value (see Beghelli & Stowell, 1997).

(9)



Next, as illustrated in (7b), because *every NP* lacks an inherent Q-value, it must undergo movement to get valued. Beghelli and Stowell (1997) propose a functional category just above *v*P, *Dist(ributive)P*, which gives a [dist(ributive]] value to a distributive quantifier via Agree.[6] As represented below, *every NP*, which lacks a Q-value, moves to Spec-DistP to pick up a [dist] value via feature-checking against Dist. As a consequence, *every NP* lands at a position higher than the subject quantifier, resulting in the ∀>∃ reading.

(10)



In summary, the ∀>∃ reading is only possible when an object quantifier moves to a position higher than the subject quantifier in Spec-*v*P due to the lack of a Q-feature value. *Every NP* moves to Spec-DistP over the subject quantifier because of the missing value. By contrast, *all NP$_{PL}$* cannot move upward because it inherently has a value specification (i.e., [Q: collective]).

6. There is also a functional category below DistP called ShareP, whose Spec is filled by a distributed event or individual (Beghelli & Stowell, 1997). We will not go into the discussion of this category. See also Marsden (2004) and Szabolcsi (2010) for comprehensive reviews and discussions.

2.2 Japanese

As already mentioned, Japanese generally does not permit the ∀>∃ reading, as in the example below:

(11) Aru syonen-ga dono-syozyo-mo aishiteiru.[7]
a boy-NOM wh-girl-conj love
'A boy loves every girl.'

It has been acknowledged that the universal quantifier *mo* has a quantificational force (cf. Kobuchi-Phillip, 2009; Kuroda, 1965). As shown in (12), an NP with *mo* has a distributive meaning but not a collective one.[8]

(12) Dare-ga kaita hon-**mo** omoshiroi. (Takahashi, 2002, p. 577)
who-NOM wrote book-conj interesting
Lit: "Every book that a person wrote is interesting."

---

7. The reader may wonder whether *aru* (*a*) is the cause of the lack of the ∀>∃ reading, but the use of other QPs disconfirms the possibility. *Dare-ka* (who-disj(unctive): *someone*) does not permit the reading, either, as shown in (i) (Shimoyama, 2006; Yatsushiro, 2009). Neither do counting QPs, as in (ii).

(i) Dare-ka-ga dono-syozyo-mo aisiteiru. (∃>∀,*∀>∃)
who-disj-NOM wh-girl-conj love
'Someone loves every girl.

(ii) Futari-no syonen-ga dono-syozyo-mo aisiteiru. (two>∀,*∀>two)
2-LINKER boy-NOM wh-girl-conj love
'Two boys love every girl.'

Furthermore, it might be the case that *-ga*, which renders the NP it attaches to definite blocks scope interaction (I would like to thank John Matthews for pointing out this problem). However, if we apply scrambling of the object QP to (11), the ∀>∃ reading becomes available, as shown in (iii).

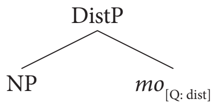
(iii) Dono-syozyo-mo aru syonen-ga *t* aishiteiru. (∃>∀,∀>∃)
wh-girl-conj a boy-NOM love
Lit: 'Every girl, a boy loves.'

This means that if (overt) movement occurs, the ∀>∃ reading is possible even though the subject is marked with *-ga*. Then, the absence of the (covert) movement of *dono-NP-mo* is what matters, which is consistent with my assumption.

8. Japanese has another universal quantifier, *subeteno* (*all*). As discussed by Marsden (2009), however, it supports a collective reading, similarly to *all* in English. By contrast, *dono-NP-mo* is compatible only with a distributive reading, similarly to *every* in English. For this reason, we focus more on the properties of *dono-NP-mo* than on *subeteno*.

Thus, following the featural description above, I assume that NP-*mo* has the following structure.[9]

(13)



By having an inherent feature value, NP-*mo* does not need move for Agree, and hence the ∀>∃ reading is not available.

In summary, the syntax of scope is not different from other movement phenomena, such as subject movement and *wh*-movement, in that it relies on the Agree-based system. *Every NP*, just like *wh*-words, lacks a Q-feature value at the onset of derivation, so that it must move to Spec-DistP to pick up a value from Dist. As a result of this process, *every NP* moves over the other quantifier, resulting in the inverse scope reading. In cases where inverse scope is not available, the object QP (e.g., *all*, *mo*) is complete from the beginning in terms of a feature valuation (e.g., by having [Q: collective] or [Q: distributive]), and hence there is no need to undergo any further movement.

3. Scope interpretation in L2 acquisition

3.1 Previous studies

A number of studies have reported that the L2 acquisition of scope interpretation is extremely hard for learners whose L1 is a scope-rigid language such as Japanese or Chinese (e.g., Chu et al., 2013; Čulinović, 2013; Ito, 2009; Kimura, 2019; Wu & Ionin, 2018, 2019, among many others). None of these studies found that L2 learners convincingly showed target-like performance in a picture- or context-based acceptability judgment task.

Chu et al. (2013) considered two possibilities that constrain L2 scope interpretation. First, they assume with Anderson (2004) that the derivation for inverse scope is more complex than the one for surface scope. As we saw earlier, in order

9. A theoretical explanation for the lack of ∀>∃ reading is still controversial (e.g., Shimoyama, 2006; Szabolcsi, 2010; Takahashi, 2002; Yatsushiro, 2009). The theoretical assumption adopted here has not been explicitly proposed, but rather, it is a collection of some previous works. It is worth noting that particles in general have the role that Agree plays in a language like English (e.g., Hagstrom, 1998; Saito, 2012, 2016; Takahashi, 2002). The assumption adopted here aligns well with this property.

to obtain the inverse scope reading, the object QP needs to undergo movement to Spec-DistP. Chu et al. suggest that this higher processing cost prevents L2 learners (as well as native speakers) from deriving inverse scope (see also Wakabayashi (2021) for relevant discussion). Second, they consider the possibility that the L2 grammar is constrained by the L1 grammar. Indeed, they showed that native speakers of Chinese accepted the surface scope reading at 100%, while accepting the inverse scope reading at 0% for Chinese sentences. These speakers showed a similar behavior in their L2 English (intermediate: 94.38% for surface scope and 7.5% for inverse scope, advanced: 98.33% for surface scope and 5% for inverse scope).

Chu et al's first suggestion involves a hidden assumption: that the syntax for inverse scope is accessible to learners. However, we have no guarantee for this underlying assumption.[10] The second possibility is likely, but the question then is "what" in the L1 grammar constrains the L2 grammar.

By contrast, Ito (2009) showed, by using a truth-value judgment task with intermediate Japanese learners of English, that inverse scope is available for learners in a context like (14).

---

10. One way to test this hypothesis is to see learners' behavior for Antecedent-Contained Deletion (ACD) like in (ia) (May, 1985). If we try to reconstruct the elided part, we will recover (ib). The elided VP is *read every book that Mary did*, meaning that the copied part again contains the same material involving VP ellipsis and this regression continues infinitely (the 'infinite regress' problem).

(i) a. Tom read every book that Mary did [$_{VP}\Delta$].
 b. Tom read every book that Mary did [$_{VP}$ *read every book that Mary did* [$_{VP}$ *read every book that Mary did*…

An ingenious solution that May (1985) proposes is to apply Quantifier Raising first, as in (ii), and then, copy VP, *read t*, as in (iii). Here, the copied VP (*read t*) no longer contains [$_{VP}\Delta$], and the infinite regress problem is solved.

(ii) [[$_{QP}$ every book that Mary did [$_{VP}\Delta$]] Tom read *t*].
(iii) [[$_{QP}$ every book that Mary did [$_{VP}\Delta\rightarrow$ <u>read *t*</u>]] Tom read *t*].

Therefore, the acceptability of ACD provides strong support for the existence of Quantifier Raising in the grammar. Some L1 studies capitalize on this logic to reveal the availability of the mechanism of Quantifier Raising despite the lack of inverse scope in child grammar (Sugawara et al., 2013; Syrett & Lidz, 2009).

In my ongoing small-scale survey, JLEs had no difficulty in accepting ACD, just like English-acquiring children. However, caution is needed to interpret this result because it is highly likely that learners might not elide VP as native speakers of English do, as Kimura and Hirokawa (2019) indeed discussed. For this reason, I cannot draw a solid conclusion for now, but this is worth investigating further to test the presence of Quantifier Raising and the economy-based account for the difficulty with scope interpretation.

(14) Every girl loves a boy.

However, we cannot conclude that L2 learners can syntactically derive inverse scope based on this observation. Some indefinites are known to take scope beyond an island, in contrast to other QPs (Fodor & Sag, 1981). When a non-indefinite QP, *every NP*, is located inside an island, as in (15), it cannot take scope over an existential QP external to the island by being trapped within an island.

(15) Josh knows *someone* [who speaks *every Germanic language*].
($\exists > \forall$, $^*\forall > \exists$) (Sabbagh, 2007, p. 367)

On the other hand, an existential QP can take scope across an island (Fodor & Sag, 1981; Szabolcsi, 2010, p. 92).

(16) Each colleague overheard [$_{\text{ISLAND}}$ the rumor that a paper of mine contains an error].
$^{\text{OK}}$ 'There is a paper of mine such that each colleague overheard the rumor that it contains an error.' ($^{\checkmark}\exists > \forall$)

This contrast tells us that while *every NP* takes inverse scope by movement, *a NP* does not take scope via movement. Specifically, the scope of indefinites is determined by other semantic means such as an existential closure or Skolemized choice function (e.g., Reinhart, 1997). Therefore, although L2 learners can judge scope in a target-like manner for a sentence like (14), this fact cannot lead us to the conclusion that syntax of scope in L2 can be acquired.

Moreover, Ito (2009) argues that the problem with the scope interpretation may lie in the non-targetlike knowledge of the indefinite, *a NP*. Indeed, it has been shown that the acquisition of articles is hard for learners whose L1 lacks them (e.g., Hawkins et al., 2006; Ionin, Ko, & Wexler, 2004; Snape, 2006, 2019). Moreover, it has also been shown that JLEs' interpretation of *a NP* is not target-like (e.g., Kimura, 2013; Kimura & Wakabayashi, 2019). Given these related findings, whether JLEs accept the inverse scope reading for a sentence involving a *two* NP_{PL} subject, as in (17), is worth testing.

(17) Two boy loves every girl. (2 > every, every > 2)

Additionally, it may be possible that L2 learners' scope interpretation is based solely on linear, rather than structural, relations (cf. Clahsen & Felser, 2006). Recall that scope relations are determined by structural positions (i.e., c-command relations), as in (18a). Thus, if learners ignore syntactic information when determining scope relations such that their scope judgments rely on the surface linear order, learners would never arrive at the inverse scope interpretation (18b).

(18) A boy loves every girl.
  a. Syntax: [*every* girl [a boy loves *every* girl]].
     Interpretation: ✓every > a / ✓a > every
  b. Linear Order: A boy loves every girl.
     Interpretation: a > every / * every > a

This account also predicts that in passives in (19), where both surface and inverse scope readings are available, learners will only accept the scope reading that matches the linear order.

(19) Every girl is loved by a boy.
  a. Syntax: [*every* girl [a boy loves *every* girl]].
     Interpretation: ✓a > every / ✓every > a
  b. Linear Order: Every girl is loved by a boy.
     Interpretation: every > a / *a > every

It has been argued on independent grounds that interlanguage grammar generally hinges on structural, rather than linear, relations (e.g., O'Grady, 2002). That said, this possibility cannot be entirely discarded as far as scope interpretation is concerned (cf. Clahsen & Felser, 2006).

3.2 Hypotheses

As briefly reviewed above, most of the previous studies on the L2 acquisition of scope have reported great difficulty with the acquisition of inverse scope. However, researchers have not reached consensus regarding the potential sources of difficulty. This study attempts to identify one of the difficulties by shedding light on lexicon – syntax properties pertaining to quantifier scope. Recall the earlier discussion on the syntax of scope here. I have argued that *every NP* can take scope over the subject due to its unvalued Q-feature, whereas *all* NP_{PL} and *mo* cannot because they do not have an unvalued Q-feature (or putting it differently, because they have an inherent feature specification). Thus, what seems to interfere with the acquisition of quantifier scope is a non-target-like feature specification or the need for *feature reassembly* (cf. Lardiere, 2008; Marsden, 2009).[11]

11. Marsden (2009), following the assumption that the quantifiers that can take distributive scope must be singular in number (Beghelli & Stowell, 1997), argues that L2ers must specify the correct [number] feature value of the quantifiers. *Every NP* carries a [singular] feature, whereas *dono NP-mo* is unspecified with number. Therefore, JLEs must know that *every NP* is a singular QP in order to have inverse scope. It might be possible that JLEs consider *every NP* to be underlyingly plural despite the surface form, prohibiting *every NP* to take inverse scope. My assumption that the distributivity is the key feature in having inverse scope is not fundamentally different from

Based on the review presented so far, in order for *every NP* to not take inverse scope, *every NP* must not move to seek the [Q: dist] value. There are two syntactic possibilities at this juncture: that JLEs recognize *every NP* as a collective quantifier, just like *all* NP_{PL}, or as an inherently distributive quantifier, just like *dono*-NP-*mo*.

Hypothesis 1: JLEs will treat *every* as a collective quantifier
Hypothesis 2: JLEs will treat *every* as an inherently distributive quantifier

As discussed in Section 2, the difference between *every NP* and *all* NP_{PL} is subtle in the sense that both NPs can refer to all members in the set (i.e., the denotation for universal quantifiers is $\forall$x [x verb y]). Thus, to distinguish between the two the semantic notion of distributivity/collectivity must be introduced. Besides, as mentioned earlier, a highly frequent expression, *everyone*, exceptionally permits the collective reading, which might confuse learners. If Hypothesis 1 is correct, JLEs will wrongly accept both distributive and collective readings for *every NP* and reject the inverse scope reading.

Hypothesis 2 is also highly likely for the following reasons. The learners' L1, Japanese, has an inherently distributive universal quantifier, namely, *dono*-NP-*mo*. If the property of this lexical item is mapped to *every*, the learners should treat *every* as an inherently distributive universal quantifier. Alternatively, they may reach the conclusion that *every* is distributive independently of L1 transfer, possibly through the input. Specifically, if Hypothesis 2 is correct, JLEs will correctly accept the distributive reading and reject the collective reading for *every NP*. The important difference from native speakers' grammar is that JLEs would reject the inverse scope reading because the distributivity comes from inherent feature specification, rather than movement to DistP.

Other hypotheses can be made on the bases of the aforementioned proposals. First, what is problematic for JLEs' interpretation of quantifier scope could be the acquisition of the indefinite singular *a NP*, as pointed out by Ito (2009):

Hypothesis 3: JLEs will have problems with existential QP subjects (i.e., *a NP*)

It is assumed in Hypothesis 3 that the problem with scope interpretation does not lie in *every NP* in the object position. Moreover, other kind of QPs such as counting QPs (e.g., *two boys*) is expected to reduce the difficulty with scope interpretation.

Furthermore, it may be possible that learners' scope interpretation is based on linear ordering:

Marsden's. There might exist, however, subtle differences between these two assumptions, and one might be more accurate than the other in predicting learners' grammar. I leave the issue to future research.

Hypothesis 4: JLEs will make scope judgments on the basis of linear ordering.

Thus, I will test these four hypotheses, which attempt to account for the difficulty with the L2 acquisition of scope interpretations.

4. Experiment

4.1 Participants

Twenty-two native speakers of English (NSEs) residing in the U.S. and 19 JLEs, who were second- to fourth-year university students majoring in English or English literature at Chuo University, participated in the experiment.[12] The proficiency of L2 participants was measured by the Minimal English Test developed by Maki et al. (2010) (maximum score is 64). The mean score was 35.2 (*max* = 57, *min* = 17, *SD* = 11.2), and most of the L2 participants should roughly be grouped as intermediate learners, based on Hasebe's (2014) criterion for grouping participants by using the Minimal English Test. Although some participants could be labeled as 'advanced' learners, I decided not to divide them into proficiency groups, as I found that proficiency did not impact the performance in the task.[13]

4.2 Task

I administered a picture-based acceptability judgment task wherein a test sentence and a picture providing a context were presented together and participants were asked to judge the acceptability of the test sentence in the given context, using binary judgment options with the addition of an 'I don't know' option.

The judgment task had two conditions, a scope test and a distributivity/collectivity condition. The former included three types, all of which were given two scope contexts (∃(2)>∀&∀>∃(2)).

(20) Type 1: Active, existential QP subject (∃>∀&∀>∃)
e.g., *A boy loves every girl.*
Type 2: Active, counting QP subject (2 >∀&∀> 2)
e.g., *Two boys read every book.*
Type 3: Passive (∃>∀&∀>∃)
e.g., *Every man is hated by a woman.*

12. I would like to thank Takuya Anahara and Yuki Takahashi for helping me collect data.

13. No correlation was found between proficiency test scores and task performance (Type 1, the ∀>∃ context) (*r* = .18).

Figure 1 provides a sample of the ∃>∀ scope reading context while Figure 2 provides a sample of the ∀>∃ scope reading context.

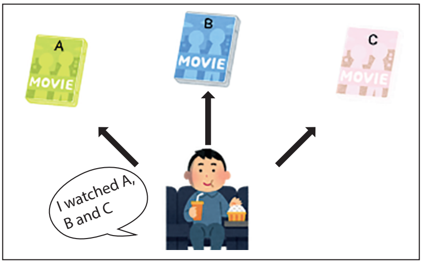


Figure 1. Sample of the ∃>∀ context
'A boy watched every movie.'

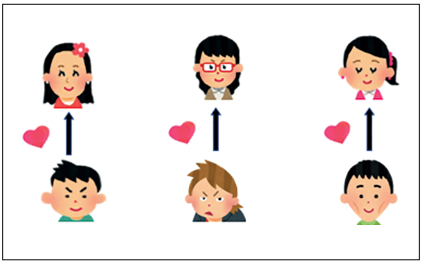


Figure 2. Sample of the ∀>∃ context
'A boy loves every girl.'

The distributivity/collectivity test targeted JLEs' knowledge about the quantifier itself. The following types were investigated. Two contexts, distributive and collective, were presented as a picture, as shown in Figures 3 and 4, respectively.[14] The collective activity in Figure 4 was intended to elicit a "reject" response in the NSE group.

(21) Type 4: *Every NP* subject (distributive & collective)
e.g., *Every student pulled a rope.*

14. The task originally included a type containing *all*. However, due to a problem with the materials, I will only report on the results for *every*. As for the experimental design, the purpose is to compare the two contexts for *every*, the distributive (correct) and collective (incorrect) context, within Type 4, so that the data concerning *all* do not matter very much for the main purposes. I would like to thank the anonymous reviewer for pointing out the problem.

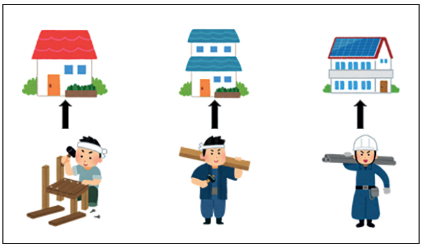


Figure 3. Sample of the distributive context
'Every man built a house.'

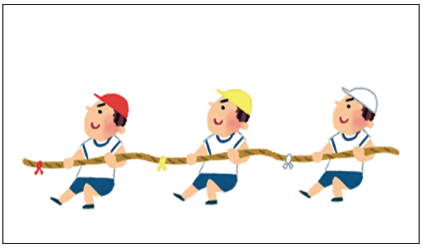


Figure 4. Sample of the collective context
'Every student pulled a rope.'

Four tokens were prepared for each type (total n = 24 for scope tests and n = 16 for distributivity/collectivity tests) and 20 fillers were also included.[15]

4.3 Predictions

I presented four hypotheses in Section 3.2. Let us now consider the specific predictions of each. First, Hypothesis 1 assumes lexical confusion between *every* and *all*.

Hypothesis 1: JLEs will treat *every* as a collective quantifier.

This hypothesis claims that *every NP* will be treated as a collective quantifier just like *all* NP_{PL} and predicts that JLEs will accept the collective reading for *every NP* (i.e., non-targetlike performance on Type 4). Therefore, *every NP* would resist inverse scope in the grammar in the same way *all* NP_{PL} does in English native speakers' grammar.

15. As pointed out by the editor, the number of tokens was very small, which is potentially problematic for inferential statistics.

Next, under Hypothesis 2, by transferring the property of *mo* to *every*, JLEs should be able to correctly understand that *every* is a distributive universal quantifier (i.e., predicting target-like performance on Type 4).

Hypothesis 2: JLEs will treat *every* as an inherently distributive quantifier.

However, since *every* is inherently but not derivationally distributive in this hypothetical grammar, the inverse scope reading would be unavailable for both Types 1 and 2.

In contrast to the above two hypotheses, Hypothesis 3 claims that the problem lies in the existential QP subject in the inverse scope interpretation.

Hypothesis 3: JLEs will have problems with existential QP subjects (i.e., *a NP*)

This hypothesis, then, expects no difficulty to arise once the subject is replaced with a counting QP such as *two* NP_{PL}, which permits both surface and inverse scope readings (i.e., expecting more target-like performance for Type 2 than for Type 1).

Lastly, Hypothesis 4 assumes that nonlinear scope interpretations are unavailable because L2 grammar may be based on linear relations.

Hypothesis 4: JLEs will make scope judgments on the basis of linear ordering.

This hypothesis predicts that nonlinear (inverse) scope is unavailable even if a lower QP is overtly moved via passivization, as shown below.

(22) Every girl$_i$ is loved t_i by a boy.
  a. Syntax: [Every girl … [by a boy [loved *every* girl]].
     Interpretation: every > a / a > every
  b. Linear Order: Every girl is loved by a boy.
     Interpretation: every > a

This leads us to predict that acceptance rates for nonlinear (inverse) scope readings would be low, resulting in no difference in the inverse scope condition between active (Types 1 and 2) and passive sentences (Type 3). Table 1 below summarizes the predictions made by these four hypotheses.

Table 1. Summary of the predictions

	Type 1 *A*-**subj (IS)**	**Type 2** *Two*-**subj (IS)**	**Type 3** **Passive (IS)**	**Type 4** *every*-**dist/coll**
Hypothesis 1	*	*	N/A	*
Hypothesis 2	*	*	N/A	✓
Hypothesis 3	*	✓	N/A	N/A
Hypothesis 4	*	*	*	N/A

Note. IS means inverse scope, and * and ✓ indicate non-targetlike and target-like performance, respectively.

5. Results

For the statistical analyses, I calculated acceptance rates (x/4 tokens) for each type and participant. Non-parametric tests (Friedman's test and Wilcoxon's Signed Rank test with p-values corrected) were run because of the relatively low number of participants and non-normal distributions (Shapiro-Wilk tests, $p < .05$ for all conditions).

To begin with, let us examine results on scope tests (Types 1–3) obtained from NSEs. As shown in Figure 5, NSEs generally accepted both scope readings for all types, as is expected by the literature. Wilcoxon's Signed Rank test with p-values corrected by Holm showed no significant differences between the two readings within each type ($Z(n = 22) = -2.67, p > .05$ for Singular Q – SS vs. Singular Q – IS; $Z(n = 22) = -2.52, p > .05$ for Counting Q – SS vs. Counting Q – IS; $Z(n = 22) = -2.43, p > .05$ for Passive – SS vs. Passive – IS) and between types on the inverse scope reading ($Z(n = 22) = -1.89, p > .05$ for Singular Q – IS vs. Counting Q – IS; $Z(n = 22) = -.30, p > .05$ for Singular Q – IS vs. Passive – IS; $Z(n = 22) = -1.87, p > .05$ for Counting Q – IS vs. Passive Q – IS).[16]

Next, as for the results from JLEs represented in Figure 6, the numerical acceptance rates for inverse scope readings for active sentences ($\forall>\exists$ for Types 1 and 2) were rather low in comparison to the passive sentences ($\exists>\forall$ for Type 3). Friedman's test showed that there was a significant difference ($\chi^2(n = 19) = 55.86, p < .001$). Wilcoxon's Signed Rank test with the p-value corrected by Holm indicated that the difference between surface and inverse scope readings for Type 1 was significant ($Z(n = 19) = -3.60, p < .01$). Similarly, the tests for Type 2 ($Z(n = 19) = -3.60, p < .01$) and Type 3 were also significant ($Z(n = 19) = -2.97, p < .05$). These results suggest that the learners strongly prefer the surface scope to the inverse scope. Furthermore, the acceptance rates for the inverse scope reading for Types 1 ($\forall>\exists$) and 3 ($\exists>\forall$) ($Z(n = 19) = -3.12, p < .01$) and Types 2 ($\forall>\exists$) and 3 ($\exists>\forall$) ($Z(n = 19) = -3.07, p < .01$) were significantly different. However, there was no significant difference in the acceptance for inverse scope readings between Type 1 and 2 ($Z(n = 19) = -.68, p = .49$). Taken together, the inverse scope for actives (Types 1 and 2) and passives (Type 3) are treated differently in JLEs' grammar.

In addition to the group analysis, I conducted an individual analysis, where I counted the number of participants who accepted three or more tokens (out of four) for each type. Table 2 presents the results. It clearly shows that while the

16. For p-value adjustment due to multiple tests, the Holm method was used instead of the better-known Bonferroni, because Holm avoids the problem of considerably decreasing statistical power (Holm, 1979).

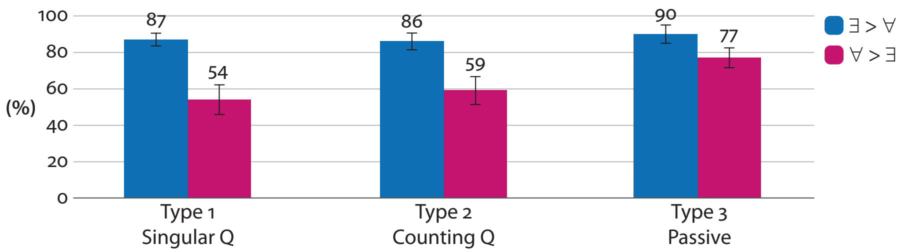


Figure 5. Acceptance rates for surface and inverse scope readings by NSEs
Note. Error bars in the figures represent standard errors of the mean.

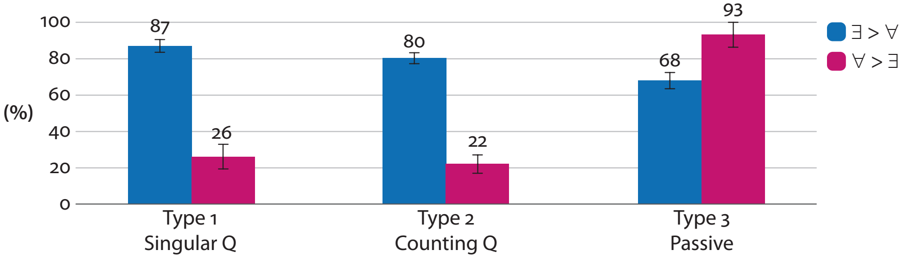


Figure 6. Acceptance rates for surface and inverse scope readings by JLEs

inverse reading for active sentences (the right columns) were consistently rejected by almost all of the JLEs, that for passive sentences were much more acceptable.

Table 2. Number of participants who consistently (≧ 75%) accepted types 1–3

	Singular Q		Counting Q		Passive	
	a > every	every > a	2 > every	every > 2	every > a	a > every
NSEs (n = 22)	20 (91%)	11 (55%)	16 (73%)	9 (56%)	19 (86%)	15 (79%)
JLEs (n = 19)	17 (89%)	2 (11%)	17 (89%)	1 (6%)	18 (95%)	11 (61%)

Note. The participants who consistently rejected the surface readings were excluded from the results for the inverse scope readings.

Let us now turn to the results of the distributivity/collectivity test. As given in Figure 7, NSEs strongly preferred the distributive reading to the collective one for *every*, as confirmed by inferential statistics (Wilcoxon's Signed Rank test, $Z(n = 22) = -3.10, p < .01$).[17]

17. The relatively high acceptance for the collective reading by the NSE group might be because the task could not eliminate the potential availability of the activity reading, in which each event participant participated in the effort directed at the event (Szabolcsi, 2010, p. 121). Moreover, the use of the indefinite QP might trigger an inverse scope reading. I leave these issues to further research.

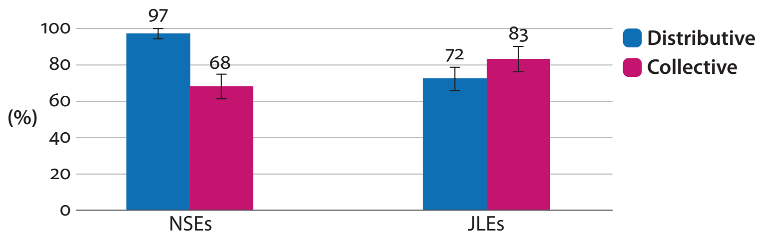


Figure 7. Acceptance rates for distributive and collective readings

By contrast, JLEs showed a slightly stronger descriptive preference for the collective reading than to the distributive one for *every*, as shown in Figure 7. However, no significant difference was found between them (Wilcoxon's Signed Rank test, $Z(n = 19) = -.96, p = .34$). The results of the individual analysis are given in Table 3. Note that almost all of the learners accepted the collective reading.

Table 3. Number of participants who consistently (≧ 75%) accepted type 4

	Distributive context	Collective context
NSEs ($n = 22$)	21 (95%)	12 (55%)
JLEs ($n = 19$)	14 (74%)	16 (84%)

6. Discussion

6.1 JLEs' knowledge on quantifiers and their scope

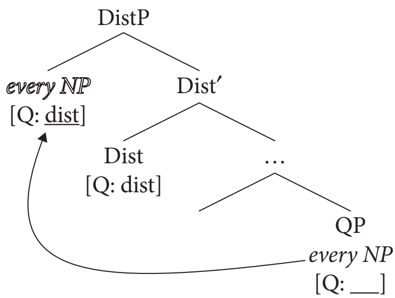
Let us discuss the results of scope and distributivity/collectivity tests. In NSEs' grammar, every is recognized as a distributive universal quantifier.[18] By contrast, in JLEs' grammar, *every* is seen as a distributive/collective universal quantifier just like *all* in NSEs' grammar. Thus, JLEs' knowledge of properties of quantifiers is divergent from NSEs'.

Every NP in JLEs' grammar appears to carry the collectivity feature as its inherent feature specification, blocking movement to DistP, which is the essential step for yielding inverse scope. Then, how does the distributivity of *every NP* arise in JLEs' grammar without movement to DistP? One possibility is that it results from the Merge of an operator, silent binominal *each.* (Recall how *all* NP_{PL} can be distributive.) Due to the presence of this silent operator, *every NP* acquires the distributive

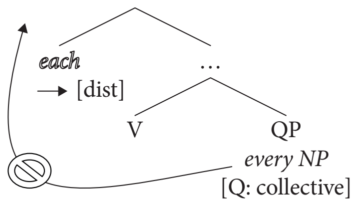
18. As noted in footnote 17, I might need to revise the task so as to totally eliminate the activity reading in a future study.

property without moving to DistP. Thus, the syntax of quantification with *every* in JLEs' grammar looks as (24) (see for (23) for the target-like representation).

(23)



(24)



6.2 Evaluating the hypotheses

Let us now turn to evaluation of the four hypotheses made previously in Section 3.2. As represented in Table 5, the actual results generally showed (i) that inverse scope was difficult for JLEs to obtain with *a NP* subject (Type 1), (ii) that the difficulty was not reduced by replacing *a NP* with *two* NP_{PL} (Type 2), but (iii) that the inverse scope was much easier to yield in passives (Type 3) than the active types (Types 1 and 2). Additionally, the distributivity/collectivity judgments for *every NP* was not target-like (Type 4). Thus, Hypothesis 1, repeated below, is totally consistent with the overall results.

Hypothesis 1: JLEs will treat *every* as a collective quantifier.

In contrast, it is less clear whether Hypothesis 2 reiterated below was strongly supported by the results.

Hypothesis 2: JLEs will treat *every* as an inherently distributive quantifier.

As shown in Table 3, three participants consistently accepted the distributive reading and rejected the collective reading for *every NP*. Furthermore, they rejected the inverse scope readings for Types 1 and 2, which is consistent with Hypothesis 2. These participants might transfer *mo* in their L1 to *every* or acquire this knowledge through (misleading) input. It is very hard to clarify where the knowledge comes from, and I leave this issue to further research. Moreover, the number of participants who displayed this pattern is low, possibly due to the small data set (total $n = 19$), and this pattern is clearly less general than the one described above. Further investigation with a refined experimental design and more participants is required to draw a more solid conclusion.

Finally, the predictions made by Hypotheses 3 and 4 did not match the results.

Hypothesis 3: JLEs will have problems with existential QP subjects (i.e., *a NP*).
Hypothesis 4: JLEs will make scope judgments on the basis of linear ordering.

Hypothesis 3 is wrong because it incorrectly predicts that the use of *two* NP_{PL} will reduce the difficulty with inverse scope. As for Hypothesis 4, it explains the difficulty with inverse scope with active sentences correctly. However, it fails to explain the fact about the passive (Type 3): If scope relations were interpreted solely on the basis of linear relations, the inverse scope reading in passives would also be difficult to obtain.

In summary, out of the four hypotheses, Hypothesis 1 made correct predictions while Hypothesis 2 was only partially supported. What was made clear in this study is that the problem lies in the acquisition of the properties of the object universal quantifier (*every*), but not in the properties of the subject existential quantifier, contra Hypothesis 3, or in the interpretive system, pace Hypothesis 4. Table 5 summarizes the evaluation of the hypotheses.

Table 5. Evaluation of the hypotheses

	Type 1 *A*-subj (IS)	Type 2 *Two*-subj (IS)	Type 3 Passive (IS)	Type 4 *every*-dist/coll
Hypothesis 1	*	*	N/A	*
Hypothesis 2	*	*	N/A	✓
Hypothesis 3	*	✓	N/A	N/A
Hypothesis 4	*	*	*	N/A
Actual Results	*	*	✓	*

Note. ✓ = target-like performance; * = non-target-like performance; shaded cells = wrong predictions.

6.3 Implications

These findings have implications for general difficulties in L2 acquisition. It has been proposed that a sentence-level deficit in L2 acquisition (e.g., *wh*-movement) roots in an unavailability of an uninterpretable feature that should be present on the attractor, namely C-head in the case of *wh*-movement (e.g., Hawkins & Hattori, 2006; Tsimpli & Dimitrakopoulou, 2007) and Dist-head in the case of quantifier movement (e.g., Kimura, 2019). Nevertheless, what this study implies is that an apparent sentence-level deficit (i.e., a deficit in the attractor) in L2 acquisition could be a noun-phrase-level deficit (i.e., a deficit in the attractee) in disguise, and what makes the learning task truly hard for L2 learners is the acquisition of a target-like property at a noun-phrase level. That is, the failure of the *feature mapping/feature reassembly* of quantifiers causes a problem in moving them in syntax and taking inverse scope, in accord with Lardiere (2008, 2009), Gil and Marsden (2013), and Kimura (2022) among others. Moreover, it seems extremely hard to find unambiguous evidence that *every NP* has the [Q:_____] feature, rather than [Q: coll] or [Q: dist]. In order to dissociate the collective character from *every* and associate the distributive one, learners need to encounter evidence showing that *every NP* in collective contexts are unacceptable, as in (25).

(25) *Every boy lifted up the table together.

Some of the participants seem to have reached this stage, possibly owing to their L1 or to the input they received. However, even if a learner arrives at the conclusion that *every NP* is distributive in some way, s/he must also know that the distributive feature comes from feature checking. A piece of evidence that tells that *every* does not carry an inherent distributive feature value is (26): When *every NP* is trapped within negation, it cannot be distributive (Beghelli & Stowell, 1997).

(26) A boy didn't read every book. (∀>∃>¬,∃>¬>∀(non-distributive))

Data like (26) are not at all easily detectable in the input, if at all, and hence it is extremely hard or even impossible to instantiate the feature reassembly process. This appears to explain why few learners could accept the inverse scope reading with *every*.[19]

7. Conclusion

This paper examined a well-known problem for scope-rigid language-speaking learners of English, namely scope interpretation. It has been acknowledged that the L2 acquisition of scope interpretation is extremely hard, but we have little knowledge about the source of the difficulty. I explored this question by investigating JLEs' knowledge of the properties of quantifiers (i.e., distributivity/collectivity). The results of the experiment showed that L2 learners' knowledge of universal quantifiers

19. I would like to thank anonymous reviewers for suggesting adding these possibilities.

was non-target-like in that JLEs believe *every* to be a distributive/collective universal quantifier just like *all*. Based on the results obtained, I argued that the obstacle with scope interpretation lies in the problem with the acquisition of target-like properties of the quantifiers. This line of discussion can be extended to a general difficulty with L2 syntax: what has been believed to be a problem in the attractor is a problem in the attractee (i.e., feature reassembly at the noun-phrase level).

Acknowledgments

I would like to thank the audience at GASLA 15, two anonymous reviewers, and the editors for their helpful comments and discussion. Special thanks go to Yuki Takahashi and Takuya Anahara for their help with data collection. I am also grateful to Makiko Hirakawa, John Matthews, and Shigenori Wakabayashi for their feedback and discussion. All remaining errors are of course my own.

References

Anderson, C. (2004). The structure and real-time comprehension of quantifier scope ambiguity (Unpublished doctoral dissertation). Northwestern University, Illinois.

Aoun, J., & Li, Y.-H. A. (1993). *Syntax of scope*. The MIT Press.

Beghelli, F., & Stowell, T. (1997). Distributivity and negation: The syntax of each and every. In A. Szabolcsi (Ed.), *Ways of scope taking* (pp. 71–107). Kluwer. https://doi.org/10.1007/978-94-011-5814-5_3

Bobaljik, J., & Wurmbrand, S. (2012). Word order and scope: Transparent interfaces and the 3/4 signature. *Linguistic Inquiry*, 43, 371–421. https://doi.org/10.1162/LING_a_00094

Bošković, Ž. (2007). On the locality and motivation of Move and Agree: An even more minimal theory. *Linguistic Inquiry*, 38, 589–644. https://doi.org/10.1162/ling.2007.38.4.589

Chomsky, N. (2001). Derivation by phase. In M. J. Kenstowicz (Ed.), *Ken Hale: A life in language* (pp. 1–52). The MIT Press.

Chu, C-Y., Gabriele, A., & Minai, U. (2013). Acquisition of quantifier scope interpretation by Chinese-speaking learners of English. In C-Y. Chu, C. E. Coughlin, B. L. Prego, U. Minai, & A. Tremblay (Eds.), *Selected proceedings of the 5th Conference on Generative Approaches to Language Acquisition North America (GALANA 2012)* (pp. 157–168). Cascadilla Proceedings Project.

Chung, E.-S. (2013). Sources of difficulty in L2 scope judgments. *Second Language Research*, 29, 285–310. https://doi.org/10.1177/0267658312464969

Clahsen, H., & Felser, C. (2006). Grammatical processing in first and second language learners. *Applied Psycholinguistics*, 27, 3–42. https://doi.org/10.1017/S0142716406060024

Čulinović, D. (2013). Development of scopal ambiguities in L1-Japanese interlanguage English. In J. C. Amaro, T. Judy, & D. P. Cabo (Eds.), *Proceedings of the 12th Generative Approaches to Second Language Acquisition Conference (GASLA 2013)* (pp. 22–31). Cascadilla Proceedings Project.

Fodor, J., & Sag, I. (1981). Referential and quantificational indefinites. *Linguistics and Philosophy*, 5, 355–398. https://doi.org/10.1007/BF00351459

Gil, K-H., & Marsden, H. (2013). Existential quantifiers in second language acquisition: A feature reassembly account. *Linguistic Approaches to Bilingualism*, 3, 117–149. https://doi.org/10.1075/lab.3.2.01gil

Hagstrom, P. (1998). Decomposing questions (Unpublished doctoral dissertation). MIT.

Hasebe, M. (2014). The acquisition of the wh-movement operation in English by Japanese EFL learners (Unpublished doctoral dissertation). Yokohama National University.

Hawkins, R., & Hattori, H. (2006). Interpretation of English multiple *wh*-questions by Japanese speakers: A missing uninterpretable feature account. *Second Language Research*, 22, 269–301. https://doi.org/10.1191/0267658306sr2690a

Hawkins, R., Al-Eid, S., Almahboob, I., Athanasopoulos, P., Chaengchenkit, R., Hu, J., Rezai, M., Jaensch, C., Jeon, Y., Jang, A., Leung, Y-KI., Matsunaga, K., Ortega, M., Sarko, G., Snape, N., & Velasco Zárate, K. (2006). Accounting for English article interpretation by L2 speakers. In S. Foster-Cohen, M. Medved Krajnovic, & J. Mihaljević Djigunović (Eds.), *EUROSLA Yearbook* 6 (pp. 7–25). John Benjamins. https://doi.org/10.1075/eurosla.6.04haw

Holm S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65–70.

Ionin, T., Ko, H., & Wexler, K. (2004). Article semantics in L2 Acquisition: The role of specificity. *Language Acquisition*, 12, 3–69. https://doi.org/10.1207/s15327817la1201_2

Ioup, G. (1975). Some universals for quantifier scope. In J. Kimball (Ed.), *Syntax and semantics* (Vol. 4, pp. 37–58). Academic Press.

Ito, A. (2009). Nihonjin Eigogakushusha niyoru suuryoushi no shutoku [The acquisition of English quantifiers by Japanese speakers]. *Eibeibungaku Kenkyu*, 27, 145–161.

Kimura, T. (2013). *The development of noun phrase structure in L2 acquisition* (Unpublished BA thesis). Chuo University.

Kimura, T. (2019). *Quantifier raising in Japanese-English interlanguage* (Unpublished manuscript). Chuo University.

Kimura, T. (2022). *Feature selection, feature reassembly, and the role of Universal Grammar: The acquisition of wh-questions by Japanese and Chinese learners of English (Unpublished PhD thesis)*. Chuo University.

Kimura, T., & Hirokawa, T. (2019). Exploring the syntactic structure of silence in Japanese-English interlanguage. Poster presented at *Generative Approaches to Language Acquisition (GALA 14)*, Milan, Italy.

Kimura, T., & Wakabayashi, S. (2019). Articles and telicity: The grammar of early and intermediate L2 learners of English. *Second Language*, 18, 89–103.

Kitahara, H. (1996). Raising quantifiers without quantifier raising. In W. Abraham, S. D. Epstein, H. Thrainsson, & C. J-W. Zwart (Eds.), *Minimal ideas: Syntactic studies in the minimalist framework* (pp. 189–198). John Benjamins. https://doi.org/10.1075/la.12.10kit

Kobuchi-Philip, M. (2009). Japanese MO: Universal, additive and NPI. *Journal of Cognitive Science*, 10, 172–194. https://doi.org/10.17791/jcs.2009.10.2.173

Kuroda, S.-Y. (1965). Generative grammatical studies in the Japanese language (Unpublished doctoral dissertation). MIT.

Lakoff, G. (1972). Linguistics and natural logic. *Synthese*, 22, 151–271. https://doi.org/10.1007/BF00413602

Lardiere, D. (2008). Feature-assembly in second language acquisition. In J. Liceras, H. Zobl, & H. Goodluck (Eds.), *The role of formal features in second language acquisition* (pp. 106–140). Lawrence Erlbaum Associates.
Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research*, 25, 173–227. https://doi.org/10.1177/0267658308100283
Maki, H., Sarenqimuge, Yoshimura, J., Makino, Y., Hasebe, M., Goto, K., Ito, T., Yumoto, Y., Oku, S., Hamasaki, M., Ueda, Y., Nagasue, K., Kasai, H., Munakata, T., & Dunton, J. (2010). The Minimal English Test: A revised version. In Korea Tesol (Ed.), *Kotesol Proceedings 2009*, (pp. 225–232). Korea Tesol.
Marsden, H. (2004). Quantifier scope in non-native Japanese: A comparative interlanguage study of Chinese, English, and Korean-speaking learners (Unpublished doctoral dissertation). University of Durham.
Marsden, H. (2009). Distributive quantifier scope in English-Japanese and Korean-Japanese interlanguage. *Language Acquisition*, 16, 135–177. https://doi.org/10.1080/10489220902967135
May, R. (1985). *Logical form: Its structure and derivation*. The MIT Press.
O'Grady, W. (2002). A linguistic approach to the study of language acquisition. *Second Language*, 1, 3–19.
Reinhart, T. (1997). Quantifier scope: How labor is divided between QR and choice functions. *Linguistics and Philosophy*, 20, 335–397. https://doi.org/10.1023/A:1005349801431
Sabbagh, J. (2007). Ordering and linearizing rightward movement. *Natural Language & Linguistic Theory*, 25, 349–401. https://doi.org/10.1007/s11049-006-9011-8
Saito, M. (2012). Case checking/valuation in Japanese: Move, Agree or Merge? *Nanzan Linguistics*, 8, 109–127.
Saito, M. (2016). (A) Case for labeling: Labeling in languages without ϕ-feature agreement. *The Linguistic Review*, 33, 129–175. https://doi.org/10.1515/tlr-2015-0017
Shimoyama, J. (2006). Indeterminate phrase quantification in Japanese. *Natural Language Semantics*, 14, 139–173. https://doi.org/10.1007/s11050-006-0001-5
Snape, N. (2006). The acquisition of the English determiner phrase by Japanese and Spanish learners of English (Unpublished doctoral dissertation). University of Essex.
Snape, N. (2019). The acquisition of articles: The story so far. *Second Language*, 18, 7–24.
Sugawara, A., Kotek, H., Hackl, M., & Wexler, W. (2013). Long vs. short QR: Evidence from the acquisition of ACD. In S. Baiz, N. Goldman, & R. Hawkes (Eds.), *Proceedings of the annual 37th Boston University Conference on Language Development* (pp. 410–422). Cascadilla Press.
Syrett, K., & Lidz, J. (2009). QR in child grammar: Evidence from antecedent-contained deletion. *Language Acquisition*, 16, 67–81. https://doi.org/10.1080/10489220902769226
Szabolcsi, A. (2010). *Quantification*. Cambridge University Press. https://doi.org/10.1017/CBO9780511781681
Takahashi, D. (2002). Determiner raising and scope shift. *Linguistic Inquiry*, 33, 575–615. https://doi.org/10.1162/002438902762731772
Tsimpli, I-M., & Dimitrakopoulou, M. (2007). The Interpretability Hypothesis: Evidence from wh-interrogatives in second language acquisition. *Second Language Research*, 23, 215–242. https://doi.org/10.1177/0267658307076546
Vendler, Z. (1962). Each and every, any and all. *Mind*, 71, 145–160. https://doi.org/10.1093/mind/LXXI.282.145

Wakabayashi, S. (2021). A principle of economy in derivation in L2 grammar: Do everything in narrow syntax. *Second Language Research*, 37, 521–545. https://doi.org/10.1177/0267658319879969

Wu, M. J., & Ionin, T. (2018). L1-Mandarin L2-English speakers' acquisition of English universal quantifier-negation scope. In M. M. Brown & B. Dailey (Eds.), *Proceedings of the annual 43rd Boston University Conference on Language Development* (pp. 716–729). Cascadilla Press.

Wu, M. J., & Ionin, T. (2019). L1-Mandarin L2-English learners' acquisition of English inverse scope. Paper presented at Generative Approaches to Second Language Acquisition (GASLA 15), Reno, Nevada, USA.

Yatsushiro, K. (2001). The distribution of *mo* and *ka* and its implications. In M. C. Cuervo., D. Harbour., K. Hiraiwa., & S. Ishihara (Eds.), *MIT Working Papers in Linguistics 41: Formal Approaches to Japanese Linguistics* 3 (pp. 181–198). The MIT Press.

Yatsushiro, K. (2009). The distribution of quantificational suffixes in Japanese. *Natural Language Semantics*, 17, 141–173. https://doi.org/10.1007/s11050-009-9044-8

Definiteness and argument position in the interpretation of bare nouns

From L1 Mandarin to L2 English

Michael C. Stern[1], Veneeta Dayal[1], Gita Martohardjono[2], Ming Chen[2] and Christen N. Madsen II[2]
[1]Yale University / [2]The Graduate Center, City University of New York

We investigate the hypothesis that nouns in article-less languages are unambiguous with respect to definiteness, an unambiguity that is evident in the interpretation of ungrammatical "bare" or article-less singular nouns in L2 English. Specifically, we tested the hypothesis that bare nouns in L2 English are interpreted as definite, administering an acceptability judgment task to intermediate L2 learners of English whose native language is Mandarin. We used sentences containing bare singular nouns in different syntactic positions and discourse contexts. We found bare subjects to be most acceptable in contexts that required definites, while bare objects were acceptable across contexts. Bare nouns in L2 English are argued to be unambiguous, following a systematic pattern determined by argument position and discourse context.

Keywords: articles, definiteness, L2 acquisition, bare nouns, subject-object asymmetries

1. Introduction

It is well known that L1 speakers of languages without definiteness-marking articles have difficulty acquiring the English article system. This difficulty is particularly apparent in production, where it manifests as misuse of English articles – such as substituting definite *the* for indefinite *a*, as in (1) (see also Ionin, Ko, & Wexler, 2004; Ionin & Montrul, 2010; Zdorenko & Paradis, 2011) – and article omission, or using ungrammatical 'bare' singular noun phrases, as in (2) (see also Goad & White, 2006, 2009; Robertson, 2000; Trenkic, 2009).

https://doi.org/10.1075/lald.67.03ste

(1) *She take the bath*. (White, 2008)

(2) *So Ø brain is already shaped*. (White, 2003)

In this chapter, we investigate the interpretation of ungrammatical bare nouns in L2 English, thus focusing our analysis on the *omission* of definite and indefinite articles, rather than misuse or substitution. We begin with a brief introduction to the issues that form the theoretical backdrop of our investigation into the L2 acquisition of determiners, taking the status of bare noun phrases in the article-less L1 as the starting point.

1.1 A look at (in)definiteness in article-less languages

In order to understand why and how bare noun phrases (NPs) are used in L2 English, it is useful to ask how bare NPs behave in the article-less L1. A number of semanticists have attempted to answer this question. Löbner (1985) speculated that bare NPs in article-less languages are simply ambiguous between definite and indefinite interpretations. A more nuanced view was proposed by Heim (2011): in article-less languages, bare NPs are semantically indefinite (i.e., they are equivalent to English indefinites), but since they do not compete with definite nouns, they do not carry the implicatures carried by English indefinites, and therefore have a wider range of felicitous uses. Both of these proposals agree that bare NPs in article-less languages cover the logical space of both the English indefinite and the English definite, and that they are, therefore, ambiguous between definite and indefinite interpretations. We group both of these views under the name *ambiguity hypothesis* (AH).

However, some studies probing the semantics of bare NPs in article-less languages have found that, contrary to the AH, the use of bare NPs is constrained by factors like discourse context and syntactic position (e.g., Dayal, 1999, 2004, 2017, 2018 in Mandarin, Russian and Hindi; Jiang, 2012 and Yang, 2001 in Mandarin). We term the notion that bare NPs in article-less languages are not ambiguous with respect to definiteness the *unambiguity hypothesis* (UH).

To demonstrate the restrictions on the interpretation of bare NPs in article-less languages, we use three contexts designed to diagnose the (in)definiteness of NPs, drawing on the semantic literature on these topics: (i) anaphoricity (Heim, 1982; Kamp, 1981), (ii) partitive specificity (Enç, 1991), and (iii) discourse referent introduction (Heim, 1982; Kamp, 1981). First, we demonstrate how these contexts restrict NP interpretation in English, before setting up these contexts in Mandarin, which is the first language of the participants in the current study.

1.2 Diagnosing (in)definiteness in English

In a context of anaphoricity (3a), the cardinality of the set of contextually salient possible referents of a singular NP is restricted to only one. As a follow-up to (3a), (3b) must have a definite NP; otherwise, the sentence is infelicitous. In other words, when the antecedent denotes a unique entity, a definite is needed to refer back to it.

(3) a. *A student and a teacher are talking.*
b. *The/#A student is holding a pen.*

(4a) sets up a context of partitive specificity, where the cardinality of the set of contextually salient possible referents of an NP is greater than one. As seen in (4b), in order to refer back to one of the previously introduced cats, an NP must be indefinite, or the sentence will be infelicitous. The definite is ruled out because there is no single unique cat.

(4) a. *There are many cats in the house.*
b. *A/#The cat is scratching a couch.*

Finally, in a context of discourse referent introduction, no contextually salient possible referents have been introduced yet. As seen in (5), in this context, indefinite NPs are preferred over definite NPs:

(5) *Many years ago, an/#the old woman used to live in this house.*

We see, based on (3–5) above, that in (i) a context of anaphoricity, definite NPs are preferred, while in contexts of (ii) partitive specificity and (iii) discourse referent introduction, indefinite NPs are preferred.

1.3 The (in)definiteness of Mandarin bare nouns

We can now create these same contexts in Mandarin in order to probe the interpretations of Mandarin NPs without articles (i.e., 'bare' NPs). In (6) we have an anaphoric context, where the bare NP *xuesheng* 'student' in (6b) refers back to the student introduced by the indefinite *yi ge xhuesheng* 'one CLF student' in (6a).[1] This suggests that bare NPs are interpreted as definite in this context.[2] But this example,

1. The judgments reported in this section are from our own informants however, see Jiang (2012) for confirmation.

2. In a recent paper, Jenks (2018) claims that Mandarin bare NPs are not acceptable in anaphoric contexts, except in subject position. Dayal and Jiang (2021) challenge this claim on the basis of a broader set of facts, while accepting the core data in Jenks' (2018) study. Since we are restricting ourselves to subject position in this section, these complications do not affect the point of the discussion.

in itself, would not be surprising if bare NPs in Mandarin were simply ambiguous with respect to definiteness, because they would then be acceptable in all contexts.[3]

(6) a. *Yi ge xuesheng he yi ge laoshi zai shuohua.*
one CLF student and one CLF teacher PROG talk
'A student and a teacher are talking.'
b. *Xuesheng wo zhe yi zhi gangbi.*
student hold PROG one CLF pen
'The student (lit: student) is holding a pen'.

In (7), we have a context of partitive specificity, which we have seen privileges indefinites over definites in English. The introductory sentence in (7a) introduces a plurality of cats. The use of the bare NP to refer back to one of these cats in (7b) is infelicitous. This is the expected behavior of NPs that presuppose uniqueness. This suggests that bare NPs are still interpreted as definite, even in this context which requires indefinite NPs. In fact, the numeral 'one' *yi* and a classifier would need to be added before 'cat' *mao* in order to convey an indefinite interpretation and make the sentence felicitous.

(7) a. *Wuzi li you hen duo mao.*
house inside have very many cat
'There are many cats in the house.'
b. *#(Yi zhi) mao zai zhua yi ge shafa.*
(one CLF) cat PROG scratch one CLF couch
'A/#The cat (lit: cat) is scratching a couch.'

Finally, we see in (8) that in a context of discourse referent introduction, bare NPs are dispreferred. Although this diagnostic is not as strong as the other two, it adds to the view that Mandarin bare NPs are interpreted as definite and do not comfortably take on the role of indefinites.[4] The numeral 'one' *yi* and a classifier

3. By using an antecedent sentence with two distinct indefinites, the possibility of referring back with a pronoun instead of a full NP is ruled out.

4. We can get some sense of the intuitions at play by comparing bare plurals with overt indefinites in English: *Many years ago, old women/some old women used to live in this house. They worked on a nearby farm*. Dayal (1999, 2004) notes that in languages like Hindi that allow both bare singulars and bare plurals, bare plurals are better at introducing discourse referents than bare singulars, but not as good as indefinites formed with the numeral 'one'. Mandarin, of course, does not make a distinction between singular and plural nouns, but Law & Syrett (2017) provide experimental evidence showing a preference in Mandarin for overt indefinites over bare NPs in such contexts. See footnote 6.

would need to be added in order to convey an indefinite interpretation and render the NP clearly felicitous.[5]

(8) *Hen duo nian qian, #(yi wei) nianlao de nvren ceng zhu zai*
very many year before (one CLF) old DE woman used.to live LOC
zhe dong fangzi-li
this CLF house-in
'Many years ago, a/#the old woman (lit: old woman) used to live in this house.'

In summary, the three diagnostics presented in (6–8) demonstrate that Mandarin bare NPs are preferred in a context in which English definites are preferred (6), and dispreferred in two contexts in which English indefinites are preferred (7–8). Based on these three diagnostics, we propose that Mandarin bare NPs are not ambiguous with respect to (in)definiteness, but are in fact semantically definite (following Yang, 2001; Dayal, 2004; Jiang, 2012, 2018).[6] Note, however, that the NPs of interest in (3–8) were all in subject position. Therefore, our proposal, so far, is restricted to NPs in subject position.

1.4 The role of syntactic position

Dayal (2011) has pointed out that when a bare NP is in direct object position in languages like Hindi, the definite interpretation generated by the semantics is overridden by another interpretation, i.e., 'pseudo-incorporation' (see also Huang, 2018 for Mandarin). When a bare NP in direct object position is pseudo-incorporated into its verb, the entire verb phrase (VP) is interpreted as an activity, analogous to English *bar-tending* or *rock-throwing*. For example, as seen in (9), a Mandarin sentence that translates literally to English 'Anu is looking after child' gets interpreted as 'Anu is babysitting' or 'Anu is doing childcare'. In this way, a bare NP that would otherwise be definite can be rendered felicitous even in a context that prefers indefinites.

5. It is possible that a sentence like *Many years ago, king ruled in this country* could be accepted under a contrastive reading juxtaposing monarchy with democracy. By using entities that may not be standard or expected, we force a choice between a bare and an indefinite and avoid this confound.

6. These studies establish that Mandarin bare NPs are kind denoting terms. Kind terms are typically classified with definites. Plural kind terms can have a type of indefinite reading, but that does not make them bona fide indefinites. See Carlson (1977), Chierchia (1998) and Dayal (2004) for important diagnostics that distinguish between the indefinite readings of NPs with overt indefinite determiners or numerals and the apparent indefinite readings of plural kind terms.

(9) *Anu zheng-zai* *zhaokan* *xiaohai*.
Anu right.now-PROG look.after child
'Anu is babysitting/doing childcare.'

Luckily, the effects of pseudo-incorporation on the interpretation of bare objects can be teased apart by examining a context where the pseudo-incorporated interpretation is blocked. This occurs when the referent of a bare NP is already salient in the discourse context, which then renders the pseudo-incorporated interpretation infelicitous (Dayal, 2011). This is illustrated in (10), where the same sentence as in (9) is preceded by a sentence that introduces five salient children into the context (10a). Here, (10b) must mean 'Anu is looking after the child'. Consequently, the sentence is rendered infelicitous, because the bare NP as a definite requires that there be a unique child, which is contradicted by (10a).

(10) a. *Zheli you wu ge xiaohai.*
here exist five CLF child
'There are five children here.'

b. *#Anu zheng zai zhaokan xiaohai*.
Anu right.now PROG look.after child
'#Anu is looking after the child.'

In summary, we have seen that bare NPs in Mandarin do not behave ambiguously with respect to definiteness; rather, they generally behave like definite NPs (see Yang, 2001; and Jiang, 2012 for further evidence), with the caveat that bare objects can be pseudo-incorporated into the verb, though only when the referent is not already salient in the context. This collection of observations is the basis of what we are calling the *unambiguity hypothesis*, or UH (see Dayal, 2017, 2018 for further details).

2. Articles in second language acquisition

In this section, we consider the different implications for the L2 acquisition of articles that the AH and UH make and relate them to the current literature on the topic.

2.1 L2 article acquisition from the perspectives of AH and UH

There is, by now, substantial literature experimentally supporting the view that the presence or absence of determiners in the L1 is a crucial factor in the L2 acquisition of determiners, especially from the viewpoint of production. In general, learners whose L1 has articles do significantly better than those whose L1 does not (e.g., Schönenberger, 2014; Zdorenko & Paradis, 2011).[7] We review some of these studies in Section 2.2, but it would be fair to say that most of the studies do not question the premise that bare NPs in the article-less L1 are ambiguous with respect to (in)definiteness. In order to contextualize the findings of these studies, we first lay out the different expectations for the L2 acquisition of determiners from the perspectives of the AH and the UH. If the AH is correct in its claim that bare nouns in article-less languages are ambiguous, and under the assumption that learners transfer[8] this ambiguity in the process of L2 acquisition, the challenge in acquiring definite and indefinite L2 articles would be to learn a distinction that is not morpho-syntactically encoded in their L1. From the perspective of the AH, it is therefore reasonable to expect that the acquisition of the definite and the indefinite article should be equally difficult. From the perspective of the UH, on the other hand, expectations are more nuanced.

Assuming that the UH is correct with respect to the facts in article-less languages, an L2 learner transferring the characteristics of bare nouns in the L1 onto the L2 might exhibit the following error pattern: article omission errors in the L2 would be predicted to be more frequent in definite than in indefinite contexts. Taking this a step further, we have seen in Section 1.4 that bare nouns in Mandarin can be pseudo-incorporated in object position. Therefore, instead of drawing on their understanding of bare nouns as definites more generally, the learner's L2 grammar may also be syntactically sensitive. That is, it would not be surprising to find that the differential interpretations of bare nouns in different syntactic positions in the L1 grammar play a role in L2 acquisition. There would, then, likely be asymmetries in the errors that occur in subject and object position, for example. From the perspective of the UH, we expect there to be a difference between the L2 acquisition of definites and indefinites, restricted by syntactic position.

---

7. A reviewer suggests that this finding may be explained via the Full Transfer/Full Access Hypothesis (FTFA, Schwartz & Sprouse, 1996). While we agree that this finding is likely the result of L1 influence, we do not want to commit to the FT/FA for the following reason: the FT/FA is a highly specific formulation of transfer based on the Principles and Parameters model of Universal Grammar. It explicitly states that "the entirety of the L1 grammar…is the L2 initial state" and that "all the principles and parameters as instantiated in the L1 grammar immediately carry over as the initial state of a new grammatical system on first exposure to input from the target language" (Schwartz & Sprouse, 1996, p. 41). It is not clear to us that the UH and its consequences for either the L1 or the L2 can be articulated in terms of a UG principle or parameter. Therefore, we prefer to rely on the broader concept of L1 influence or L1 transfer.

8. We use the term here (and in subsequent mentions) in its broader sense, i.e., not in the strict sense of transfer of grammatical features.

Let us remind ourselves that, as far as the adult grammar of Mandarin is concerned, the AH has been shown to be empirically flawed (Section 1.3). Thus, regardless of the particular theoretical explanations one may favor, the UH is supported by empirical data from the L1. The question we are interested in probing is whether and how the UH is reflected in the L2 acquisition of determiners. To contextualize the concrete predictions that we set out to test (described in Section 3.2), we briefly review relevant L2 literature. In doing so, we restrict ourselves to issues related to omission rather than misuse of articles since that is the focus of the study that we report on in this chapter.

2.2 Article omission in L2 development: Systematicity vs. randomness

The acquisition of determiners has been widely studied in both L1 and L2 development, albeit from diverse theoretical approaches and with varying results (e.g., Huebner, 1983, 1985; Master, 1987; Parrish, 1987; Thomas, 1989; among many others). We focus on those studies that are particularly relevant to the issue of how *bare nouns* are (mis)used in L2 English, thus highlighting those results that bear on article *omission*, even though this may not have been the main focus of these studies. In particular, we discuss how these studies reveal systematicity in article omission arising from pragmatic, semantic, syntactic, and processing constraints.

In an overview study, Trenkic (2009) writes that article omission is rife in L2 production, but that it shows a systematicity that generally follows discourse and pragmatic principles, rather than occurring randomly (as would be predicted by the AH). One of these principles is linked to the saliency of the NP, whereby the higher its saliency in the discourse, the more likely the article is to be dropped. Trenkic (2009) links this to the Information Load Hypothesis (Almor, 1999, 2005; Almor & Nair, 2007) and ascribes it to processing factors, rather than grammatical representation. Trenkic's (2007, 2009) own proposal, based on analyses of Serbian, Thai and Chinese learners of L2 English, is that, in addition to discourse and pragmatic principles, L2 articles are misanalyzed as adjectives at earlier stages of acquisition, thus also addressing issues of representation. Other contexts where Trenkic (2009) notes higher degrees of article omission are: topic vs. non-topic position, for definite articles; subsequent vs. first mention; and when speakers are referring to objects present in the immediate environment. Trenkic's (2009) observations and explanations are specific to production, which is highly vulnerable to working memory constraints and processing factors. It should be noted that Trenkic (2009) does not directly address an asymmetry in the omission of articles in definite vs. indefinite environments. However, what clearly emerges in her discussion is that article omission does not occur randomly but is systematic, and, in her view, governed by a combination of discourse, processing and grammatical principles.

In a study that, similar to ours, looks more directly at article omission, Robertson (2000) tests a group of L2 speakers of English whose L1 is Mandarin. He uses a production task that was designed to elicit nouns in definite and indefinite contexts (Brown, 1995). Robertson reports his results separately for what he calls "echo contexts" and "non-echo contexts". Echo contexts function as comprehension checks and contain reformulations of a previous NP. These, he claims, are more representative of the grammatical representation of the learner than if the utterances were simply a repetition of a previously encountered NP. The utterances produced in the echo contexts are more revealing, as they show a higher percentage of article omission overall (both definite and indefinite) than in non-echo contexts. But of particular interest to us is that, in the echo contexts, Robertson finds a higher percentage of article omission (thus a higher percentage of bare nouns) when a *definite* article is required than when an *indefinite* is required. This result aligns with the predictions of the UH.

Similar to Trenkic (2009), Robertson (2000) argues for systematicity of omission rather than optionality (i.e., randomness as would be predicted by the AH). He writes that this systematicity is reflective of a shift from discourse-oriented language (Mandarin) to syntax-oriented language (English), a claim that is similar to our hypothesis that speakers of article-less languages rely in part on the semantic function of bare nouns in the L1 as they develop target-like use of articles in the L2.

More evidence that article omission follows a distinct pattern, rather than a random one, is found in Schönenberger (2014) who sets out to test Ionin, Ko, & Wexler's Fluctuation Hypothesis (FH) (2004). The FH posits an Article-Choice Parameter set to either definiteness or specificity in languages with articles. L2 learners whose L1 does not have articles are then predicted to fluctuate between the two settings for a certain period of time until they converge on the correct one for the L2. Schönenberger tests the FH by investigating the L2 English of three groups: an L1 German group which shares the parameter setting of the L2 English; and two groups of L1 Russian speakers, one with more advanced L2 English proficiency than the other. Under the FH, the Russian groups would not have the Article Choice Parameter set yet would still have access to it. Schönenberger found evidence for transfer of the L1 parameter setting in the German group, but reports that her results show only partial evidence for the FH.

The contrast between definiteness and specificity central to the FH is orthogonal to the claims of the UH, and we will not address it further here. Instead, we note one particular result in Schönenberger's study pertaining to the differential omission of definite vs. indefinite articles, a contrast that *is* relevant to the UH. She reports that article omission (i.e., selecting "no article" in the forced-choice article selection task) in the Russian groups was systematic, "occurring more often in contexts where nouns are perceived as clearly definite" (2014, p. 102). This in turn leads her to raise the question: *are learners of an L1 without articles in general more likely to omit articles with definites?*[9]

In an early study investigating the acquisition of articles in L2 Spanish by L1 Quechua speakers, Sanchez and Gimenez (1998) also found an asymmetry in the omission of articles. In particular, the definite article (*el/la*) tended to be omitted at a higher rate when the noun was in a prepositional phrase, compared to when it was in subject position. Only in object position did the omission of the definite article not appear to be systematic. This led them to conclude that omission of the definite article is influenced by the particular argument position occupied by the noun. This claim is similar to that of the UH, that syntactic position will play a role in the appearance of L2 bare nouns.

Systematicity and asymmetries in definite vs. indefinite article omission were also present in data collected by Zdorenko and Paradis (2011). This study tested child L2 learners coming from a variety of L1 backgrounds, some with articles (Spanish, Arabic), some without (Mandarin, Cantonese, Hindi, Urdu and Punjabi). While, overall, the children were more accurate in their production of *the* than of *a*, the authors found, similar to the adult data in Schönenberger (2014) and in Robertson (2000), that article omission was higher in definite contexts than in indefinite contexts.

What is of note in all of these studies is that they point to systematicity of L2 article omission, guided by a variety of pragmatic, semantic, syntactic, and processing factors. Of particular note for the present study is the tendency for L2 articles to be omitted more in definite environments than indefinite ones. This aligns with the predictions of the UH for L2 acquisition, to which we turn next.

3. The study

We propose that (at least some of the) systematicity observed in article omission can be derived from the UH claim that bare NPs in the article-less L1 are definite, and that L2 learners generally follow the L1 pattern of using bare nouns in definite contexts. This would also entail that bare nouns in the L2 are (erroneously) interpreted as definite during comprehension. The studies we cited in Section 2.2 all focused on L2 *production*, but to our knowledge, no study has directly examined the interpretation of bare NPs in L2 English *comprehension*, which is the aim of our study.

9. This asymmetry in definite vs. indefinite article omission was also present in Ionin, Ko, & Wexler (2004), though it might have been restricted to particular nouns such as “mother” and “father”, which can be used as names.

We test the proposal that restrictions on the interpretation of bare NPs in Mandarin (as described by the UH) guide the interpretation of ungrammatical bare singular NPs in L2 English. This proposal contradicts the AH, which would predict an even distribution in the acceptability of bare NPs in L2 English across contexts. The study described in this chapter was the first of a set of four experiments probing the interpretation of bare nouns in L2 English. It consisted of an acceptability judgment task (AJT) designed to establish whether and in what specific contexts and positions intermediate learners of L2 English whose L1 is Mandarin would accept bare nouns.

Our research questions are:

RQ1. Will Mandarin-speaking learners of L2 English show differential acceptability of ungrammatical bare nouns in English, based on the patterns in the L1, whereby bare nouns are preferred in definite over indefinite contexts?

RQ2. Will there be differential treatment of L2 English bare nouns in subject position compared to object position, due to mitigating effects in the L1 such as pseudo-incorporation?

3.1 Methodology

Participants. 52 native speakers of Mandarin Chinese who were L2 learners of English (age 18–54, $M = 26.62$, $SD = 8.43$) were given financial compensation for participating in the experiment. All participants were residents of New York City and had arrived in the US from China between the ages of 13 and 49 ($M = 23.02$, $SD = 8.05$). Participants' length of residence in the US at the time of testing ranged from less than one year to 17 years ($M = 3.60$, $SD = 3.54$). Note that most participants lived in predominantly Chinese communities in Queens, New York, where the dominant language is Mandarin, so residence in these communities does not necessarily entail English immersion to the same degree as would be entailed by living in a more dominantly anglophone community.

At the time of testing, all participants were enrolled in an intermediate-level English language class at one of two campuses of Queensborough Community College (QCC), City University of New York (CUNY), in Queens, New York. Participants had been placed in this class level based on the results of a battery of proficiency tests: a multiple-choice test focusing on grammar, vocabulary, sentence structure, reading, and listening (at one campus, this test was designed by CUNY faculty; the other campus uses the CaMLA English Placement Test [Walter & Hentschel, 2013]); an oral test of verbal ability administered by CUNY faculty; and an essay test designed by CUNY faculty. Participants were further screened with a questionnaire developed in our lab, which collected information pertaining to language exposure and self-reported proficiency. Only those respondents who reported no significant exposure to languages with articles (other than English), and who reported their English proficiency as intermediate, were included in the study. This proficiency level was selected because we needed to ensure that the L2 participants had not fully acquired the English article system, yet could comprehend the stimulus sentences without too much difficulty. In addition to the L2 experimental group, a control group of 20 native English speakers was tested on a pilot version of the experiment consisting of a subset of the stimuli. All of the native English speakers self-reported that they had no greater than beginner proficiency in any language other than English.

Stimuli. Each stimulus for the AJT consisted of two English sentences: a context sentence followed by the target sentence. Stimuli were presented simultaneously in written and auditory form. The stimuli were grouped into 6 conditions: type of context (anaphoricity, partitive specificity, discourse referent introduction) by syntactic position of the bare NP in the target sentence (subject, object). Example stimuli from each condition can be seen in (11–13), with each bare NP underlined for clarity. 'ANA-D' is shorthand for the anaphoricity context, with 'D' signifying that this is a context in which definites are preferred; 'PAS-I' is shorthand for the partitive specificity context, with 'I' signifying that this is a context in which indefinites are preferred; and 'DRI-I' is shorthand for the discourse referent introduction context, another context in which indefinites are preferred.

(11) *ANA-D*
  a. *Subject:* A student and a teacher are talking. <u>Student</u> is holding a pen.
  b. *Object:* A necklace and a bracelet are on sale. A woman is buying <u>necklace</u>.

(12) *PAS-I*
  a. *Subject:* There are many cats in the house. <u>Cat</u> is scratching a couch.
  b. *Object:* There are many shirts in the store. A girl is buying <u>shirt</u>.

(13) *DRI-I*
  a. *Subject:* The beach is sunny today. <u>Child</u> is wearing a hat.
  b. *Object:* The lawn needs more plants. A worker is planting <u>tree</u>.

Five stimuli were presented from each condition, for a total of 30 experimental stimuli. Also included were 30 grammatical control stimuli, of the same form as the experimental stimuli, but with no bare NPs (e.g., 'A boy and a girl are playing. <u>The boy</u> is throwing a ball'), in order to ensure that the L2 learners generally accepted bare NPs to an equivalent degree as grammatical NPs, and therefore, had not fully acquired the English article system. Finally, we included 60 filler stimuli, which also contained two sentences but were of different structures than the experimental stimuli. 30 filler stimuli were grammatical, and 30 filler stimuli were ungrammatical. In total, each experimental participant gave acceptability judgments on 120 stimuli.[10] Auditory stimuli were recorded in a sound-attenuating booth by a native English speaker from the northeastern United States using neutral prosody.

Procedure. The experiment was conducted in a quiet room either at the Graduate Center, CUNY, or in an empty classroom at QCC. In order to ensure comprehension of the task, all communication between the participant and the experimenter was conducted in Mandarin. Stimuli were presented in pseudo-random order on a tablet computer running E-Prime 2.0 (Schneider, Eschman, & Zuccolotto, 2002). During each trial, participants were simultaneously presented with a written English stimulus on the tablet screen and an auditory recording of the same stimulus through headphones at a comfortable volume. Then, they selected their acceptability rating on a 6-point Likert scale, which appeared on the tablet screen with endpoints labeled in Mandarin (非常好 'very good' and 非常差 'very bad'), as is common practice (see, for example, Liu & Keusch, 2017). At the beginning of the experiment, they were given the following instructions, again in Mandarin: '请对每一对英语句子进行评分，从 '非常好' 到 '非常差。' 'Please rate each pair of English sentences on the scale from "very good" to "very bad"'. Then the participant performed six practice trials with sentences unrelated to the experimental stimuli, after which they were prompted to ask the experimenter any questions that they had, in order to ensure comprehension of the instructions. Participants were given a short break after each third of the experiment (after 40 trials and after 80 trials). An entire session lasted approximately 30–40 minutes.

3.2 Predictions

The predictions of the UH are summarized in Table 1. Each row summarizes the predictions by context within each syntactic position. For bare subjects, the UH predicts that ratings should be higher in ANA-D, the definite context, than in PAS-I and DRI-I, in which indefinites are preferred. For bare objects, the UH also predicts that ratings should be higher in ANA-D than in PAS-I. However, DRI-I should also have relatively high ratings because, as seen in Section 1.4, this context allows

10. A reviewer points out that using multiple experimental lists, rather than a single list for all participants, would have allowed us to better control the confounding effects of the lexical variation of stimuli between conditions (Cowart, 1997). We agree, and we will design our follow-up experiments with this consideration in mind. In our analysis of this first experiment, however, we mitigate these confounding effects by using mixed effects models (see Section 3.3), which allow us to model the stochastic variability of participant responses caused by the idiosyncratic features of each stimulus using random intercepts (Singmann & Kellen, 2018).

pseudo-incorporation. Each column summarizes the predictions by syntactic position within each context. The UH predicts that bare subjects and bare objects should have equivalent ratings in the ANA-D and PAS-I contexts, but that bare objects should have higher ratings than bare subjects in the DRI-I context, because bare objects will be rendered acceptable via pseudo-incorporation, while bare subjects will remain definite.

Table 1. Predictions from the UH for relative acceptability ratings in the six experimental conditions

	ANA-D	PAS-I	DRI-I
Bare Subject	✓	X	X
Bare Object	✓	X	✓

Note. Check marks indicate relatively higher ratings caused by felicity, and Xs indicate relatively lower ratings caused by infelicity.

3.3 Analyses

First, in order to test whether the L2 learners had acquired the English article system, rating data from both ungrammatical (including bare nouns) and grammatical control (no bare nouns) conditions were fit with a cumulative link mixed effects model using the ordinal::clmm() function in R (Christensen, 2019; R Core Team, 2019) with condition (grammatical, ungrammatical), group (L2 learner, native speaker), and their interaction as fixed effects, and random intercepts by participant and by item. The ungrammatical condition and the L2 learner group were coded as the baseline level of each factor. We compared the fit of this model to that of the null, intercept-only, model using the likelihood-ratio test. The Wald test was used to test for the significance of fixed effects, and the Tukey HSD test was used to test for significant differences between conditions using the emmeans::emmeans() function in R (Lenth, 2020) at an alpha level of .05.

Next, to test our main predictions (described in Section 3.2), rating data from L2 learners on ungrammatical conditions only were fit with a cumulative link mixed effects model using the same function in R and the same random intercepts described above, but with fixed effects of syntactic position (subject, object), discourse context (ANA-D, PAS-I, DRI-I), and their interactions. Subject position and the ANA-D context were coded as the baseline level of each factor. Again, the fit of this model was compared to that of the null, intercept-only, model using the likelihood-ratio test, and the significance of fixed effects and differences between conditions were tested using the same methods described above.

4. Results

4.1 Overall acceptability of bare nouns

As seen in Figure 1, although the L2 learners did rate the ungrammatical stimuli ($M = 4.78$, $SD = 1.41$) as less acceptable than the grammatical stimuli ($M = 4.92$, $SD = 1.46$), this difference was much smaller than that seen in the native speakers ($M_{ungram} = 2.15$, $SD_{ungram} = 1.41$; $M_{gram} = 5.14$, $SD_{gram} = 1.13$). Moreover, unlike the native speakers, the L2 learners rated ungrammatical stimuli far above the midpoint of the scale (3.5).

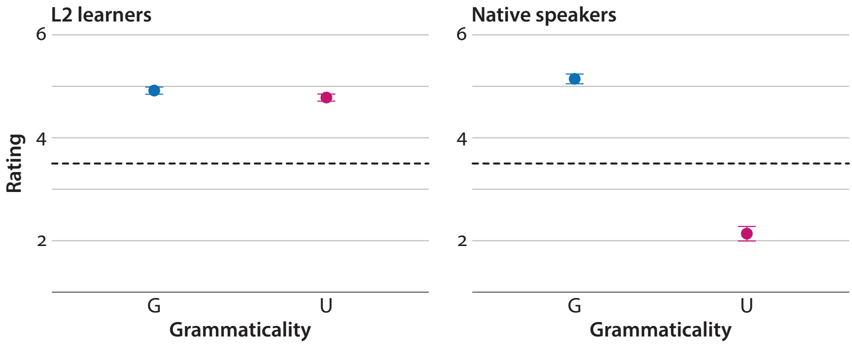


Figure 1. Mean acceptability ratings by grammaticality among L2 learners and native speakers

(Note: 95% confidence interval error bars. The dotted black line represents the midpoint of the acceptability scale.)

The cumulative link mixed effects model described in Section 3.3 was a significantly better fit to the data than the null, intercept-only model ($\chi^2(3) = 210.96$, $p < .001$). As seen in Table 2, the main effect of grammaticality was not significant, but the main effect of group and the interaction between grammaticality and group were significant.

Table 2. Cumulative link mixed effects model fit to rating data by L2 learners and native speakers on ungrammatical and grammatical stimuli

	Estimate	Std. Error	z-value	p-value
Condition (Gram)	0.24	0.13	1.75	.08
Group (Native)	−3.48	0.32	−11.03	< .001
Condition:Group	3.59	0.23	15.62	< .001

Post hoc tests revealed that native speakers' ratings of ungrammatical stimuli were significantly lower than their ratings of grammatical stimuli ($p < .001$), and significantly lower than L2 learners' ratings of both ungrammatical stimuli ($p < .001$) and grammatical stimuli ($p < .001$). L2 learners' ratings of ungrammatical stimuli did not significantly differ from their ratings of grammatical stimuli ($p = .24$). Based on these results, we conclude that the L2 learners generally accepted bare NPs as grammatical in English, unlike the native speakers.

4.2 Acceptability of bare nouns by discourse context and syntactic position

Next, we test our main predictions by examining the acceptability of bare NPs by context and syntactic position. As seen in Figure 2, bare subjects in the ANA-D context ($M = 5.08$, $SD = 1.31$) were rated higher than those in the PAS-I context ($M = 4.23$, $SD = 1.67$) and the DRI-I context ($M = 4.57$, $SD = 1.50$). Moreover, bare subjects in the DRI-I context were rated higher than those in the PAS-I context. For bare objects, no clear differences were found between the three contexts ($M_{ANA\text{-}D} = 4.90$, $SD_{ANA\text{-}D} = 1.38$; $M_{PAS\text{-}I} = 5.02$, $SD_{PAS\text{-}I} = 1.34$; $M_{DRI\text{-}I} = 4.85$, $SD_{DRI\text{-}I} = 1.38$).

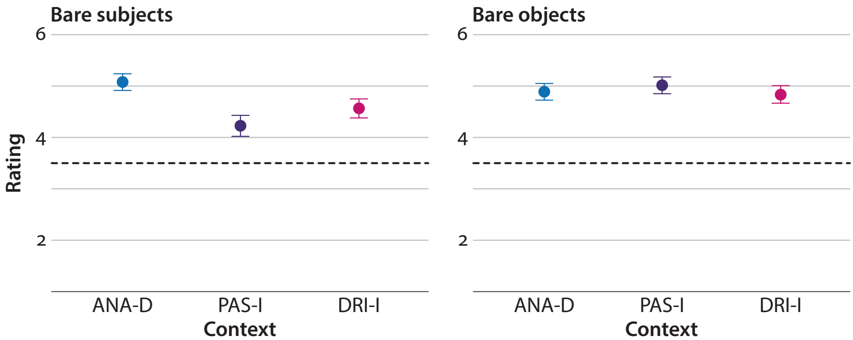


Figure 2. Mean acceptability ratings by discourse context and syntactic position among L2 learners

(Note: 95% confidence interval error bars. The dotted black line represents the midpoint of the acceptability scale.)

The cumulative link mixed effects model described in Section 3.3 was a significantly better fit to the data than the null, intercept-only model ($\chi^2(5) = 18.46$, $p < .01$). As seen in Table 3, there were significant main effects of context (PAS-I and DRI-I, compared to ANA-D), and a significant interaction between position and context (PAS-I). There was no significant main effect of position or significant interaction between position and context (DRI-I).

Table 3. Cumulative link mixed effects model fit to rating data by L2 learners on ungrammatical stimuli

	Estimate	Std. Error	z-value	p-value
Position (Object)	−0.34	0.30	−1.13	.26
Context (PAS-I)	−1.28	0.30	−4.23	< .001
Context (DRI-)	−0.88	0.30	−2.92	< .01
Position (Object):Context (PAS-I)	1.47	0.42	3.45	< .001
Position (Object):Context (DRI-I)	0.77	0.43	1.80	.07

Post hoc tests revealed that bare subjects were rated significantly higher in the ANA-D context than the PAS-I context ($p < .001$) and the DRI-I context ($p = .04$). In the PAS-I context, bare subjects were rated significantly lower than bare objects ($p < .01$). Bare subjects in the PAS-I context were also rated significantly lower than bare objects in the ANA-D context ($p = .02$). No other pairwise comparisons revealed significant differences.

5. Discussion

Comparing the L2 learner group to the native speaker group:

i. L2 learners rated the sentences with bare nouns as equivalently acceptable to those without bare nouns, while native speakers rated the sentences with bare nouns as far less acceptable than the sentences without bare nouns.

Among the L2 learner group:

ii. Bare subjects were significantly more acceptable in the definite context (ANA-D) than in the two indefinite contexts (PAS-I and DRI-I);
iii. Bare objects were equivalently acceptable across all three contexts;
iv. Bare subjects and bare objects were equivalently acceptable in the definite context (ANA-D);
v. Bare objects were significantly more acceptable than bare subjects in the PAS-I context; and
vi. There was no significant difference between bare subjects and bare objects in the DRI-I context, although the interaction between position and context (DRI-I) seen in Table 3 suggests a tendency for the negative effect of the DRI-I context to be mitigated for bare objects. However, this interaction was not significant ($p = .07$).

Table 4 summarizes the main findings (ii-vi) in relation to our predictions. The greater acceptability of bare subjects in the definite context compared to the two indefinite contexts (the first row in Table 4) is consistent with our prediction that bare subjects would be interpreted as definite. Moreover, the equivalent acceptability of bare subjects and bare objects in the definite context (the first column in Table 4) is consistent with our prediction that bare objects would be interpreted as definite in anaphoric contexts.

The X with the single asterisk in Table 4 signifies that the difference in the acceptability of bare subjects between the ANA-D and DRI-I contexts was relatively small, and that the difference between bare subjects and bare objects in the DRI-I context was not significant. The weakness of these two effects arises from the relatively higher acceptability of bare subjects in the DRI-I context, which might be explained by the fact that the DRI-I context is a relatively weaker diagnostic of (in)definiteness than the other two contexts (see Section 1.3 and the references cited there).

Table 4. Summary of relative acceptability ratings in the six experimental conditions

	ANA-D	**PAS-I**	**DRI-I**
Bare Subject	✓	X	X*
Bare Object	✓	✓**	✓

Note. Check marks and Xs without asterisks indicate strong significant effects consistent with our predictions; one asterisk indicates a weaker effect consistent with our predictions, and two asterisks indicate a strong significant effect in the *opposite* direction of our predictions.

The equivalent acceptability of bare objects across all three contexts, and the significantly greater acceptability of bare objects compared to bare subjects in the PAS-I context, were inconsistent with our predictions. We see from the check mark with two asterisks in Table 4 that both of these unexpected findings can be understood as arising from the relatively high acceptability of bare objects in the PAS-I context. Recall that the UH predicted bare objects to be infelicitous in the PAS-I context, because pseudo-incorporation should have been blocked and bare objects should have been interpreted as definite in this indefinite context. In the following two subsections, we explore two possible explanations of this unexpected result and propose two follow-up experiments – a picture-point task and an AJT – each designed to test one possibility.

5.1 Definiteness and the singular-plural distinction

The first possible explanation relates to the role of number distinction in the noun phrase. In introducing the partitive specificity diagnostic in Section 1.2, we focused on singular noun phrases. However, moving to the plural form affects this diagnostic in significant ways. The plural version of (4) admits both indefinite *and definite* noun phrases:

(14) a. *There are many cats in the house.*
b. *Some/The cats are scratching a couch.*

While both the indefinite and the definite are felicitous, there is a crucial difference in interpretation. The indefinite has an implicature that only a subset of the cats in the house are scratching the couch, while the definite conveys the information that all the cats are. That is, the plural definite has a maximality implicature.[11]

Crucially, Mandarin is known not to make number distinctions in the noun phrase (Yang, 2001; Jiang, 2012, 2018; among many others). As such, our characterization of (7) (repeated below) as infelicitous holds only for the intended singular interpretation. It would be perfectly acceptable under the plural interpretation 'The cats are scratching a couch', referring to the totality of cats. It is possible, therefore, that bare objects in the PAS-I context were generally accepted because participants misanalyzed them as plural.

(7) a. *Wuzi li you hen duo mao.*
house inside have very many cat
'There are many cats in the house.'
b. *#(Yi zhi) mao zai zhua yi ge shafa.*
(one CLF) cat PROG scratch one CLF couch
'A/#The cat (lit: cat) is scratching a couch.'

A forthcoming picture-point experiment will test the possibility that the felicity of bare objects in the PAS-I context was maintained through the availability of a plural, maximal interpretation. In this experiment, participants will be given sentences with a bare NP in either subject position or object position and asked to choose between three pictures which represent three alternative interpretations of the bare NP. One picture will represent an indefinite singular interpretation, another will represent an indefinite plural interpretation, and the third will represent a definite plural, or maximal, interpretation. An example picture array for a stimulus with

11. The maximality in the interpretation of plural definites is characterized as an implicature, since it is not strictly required that every single cat be engaged in the activity, as would be required if the universal *every cat* were used instead. This phenomenon has been studied under the label of pragmatic slack (Brisson, 1997, 2003; Lasersohn, 1999; Schwarz, 2013; among others).

a bare object is displayed in Figure 3. If the UH is correct that bare objects tend to be interpreted as definite in this context, but felicity is maintained through a plural interpretation, the rightmost (definite plural) picture will be the most likely to be selected.[12]

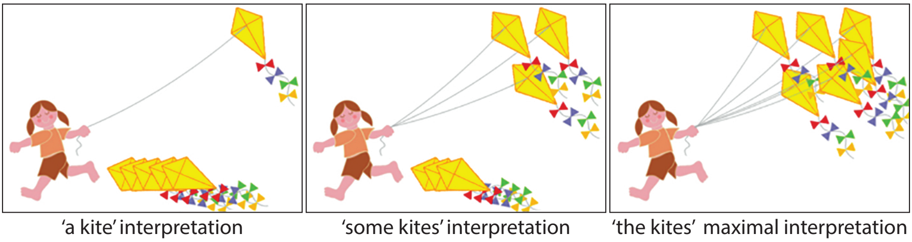


Figure 3. Example picture array from the forthcoming picture-point experiment for the stimulus sentence 'The girl flew kite.'

If this hypothesis is supported, the question would remain why bare objects lend themselves to plural interpretations, while bare subjects do not. One possibility is that in our stimuli, which used the present progressive tense on all verbs, the inflection on the copula encoded the fact that the subject was singular, but did not do the same for the object. This extra information regarding the number of the subject might have been utilized by our participants to restrict the interpretation of the subjects to singular, while the objects remained open to either a singular or plural interpretation. Another possibility is that from a processing perspective, subject position is more prominent than direct object position, so participants were more likely to utilize information regarding the subject than the object in their selection of acceptability ratings. Forthcoming data from the picture-point experiment will help tease apart these potential explanations.

5.2 Definiteness and topicality

We now discuss the second possible explanation of the unpredicted felicity of bare objects in the PAS-I context and describe a second follow-up experiment designed to test this explanation. If our participants *did* interpret bare objects as singular in

12. A reviewer points out that asking participants to interpret sentences with ungrammatical bare singular nouns would be confusing for someone who already knows that English requires articles for singular nouns. Although the results reported in Section 4.1 demonstrate that our participants generally have *not* acquired this knowledge, for any participants who have, the viability of the task will be maintained by asking participants to "Select the picture that *best* matches the sentence, even if the sentence sounds odd, or none of the pictures match it very well".

the PAS-I context, this would suggest, contra the UH, that the patterns we observed in the interpretation of bare nouns are not reflective of (in)definiteness, per se, but rather of topicality.[13] Cross-linguistically, nouns in subject position tend to be more strongly associated with topicality than those in other syntactic positions. Mandarin, in particular, is known to be a topic prominent language where the first constituent, often the subject, functions as a topic (Li & Thompson, 1976).[14] Given that topicality is associated with definiteness, this could explain why bare subjects seemed to be interpreted as definite by our participants, while bare objects apparently admitted both definite and indefinite interpretations.

A follow-up AJT will test this possibility by teasing apart syntactic position from topicality. The design of this experiment will be nearly identical to the experiment reported here, except that we will include bare nouns in two positions that were not included in the first experiment: indirect object position (e.g., 'The boy is giving <u>girl</u> a gift'), and subject position but preceded by an explicit topic (e.g., 'Outside, <u>boy</u> is kicking a ball'). Indirect objects are not canonically associated with topicality. Therefore, if the acceptability of bare nouns in indirect object position tends to pattern with that of bare subjects, this would suggest that the results we have seen so far are driven by the semantic definiteness of bare NPs (as posited by the UH), rather than by discourse topicality. On the other hand, if bare indirect objects tend to pattern with bare direct objects, this observation would suggest that the pattern in the acceptability of bare subjects is driven by the topicality of this syntactic position, rather than by the semantics of NPs more generally.

The second condition of the follow-up AJT will focus on the subject itself. Yang (2001) argues that appearances notwithstanding, Mandarin bare nominals in subject position are not necessarily definite. Following Li and Thompson (1976), she treats Mandarin as a topic prominent language, where the first constituent functions as a topic. She relates the perceived definiteness of subjects to topicality and provides the following minimal pair as evidence:[15]

(15) a. *gou zai-jiao*
dog be-barking
'*A dog/*Dogs/The dog/The dogs are barking'
[$_{\text{TOPIC}}$ dog$_i$ [$_{\text{TP}}$ t$_i$ [$_{\text{VP}}$ is barking]]]

13. Two classic references on the issue of topics are Li & Thompson (1976) and Reinhart (1981).

14. Yang (2001) has argued that Mandarin subject bare NPs can be interpreted as indefinite when they are preceded by another expression; see the final paragraph in this subsection.

15. This is in keeping with the parallel drawn between English bare plurals and Mandarin bare nominals which are unmarked for number (see footnotes 4 and 6).

b. *waimian/yuanchou gou zai-jiao* (Yang, 2001, p. 32)
outside/far away dog be-barking
'Dogs/A dog/The dog/The dogs are barking outside'
[$_{\text{TOPIC}}$ outside [$_{\text{TP}}$ dog [$_{\text{VP}}$ is barking]]]

As (15) indicates, the indefinite reading of subject bare NPs peeks out if another expression functions as a topic. Given this fact, it would also be revealing to see if the bias towards definiteness for subjects shifts in the L2 grammar for English sentences with explicit topics in first position.

6. Conclusion

We examined the interpretation of ungrammatical bare NPs in L2 English by intermediate learners from an L1 Mandarin background. Our objective was to test the extension of the unambiguity hypothesis (UH), which proposes that bare NPs in article-less languages are semantically definite, to L2 acquisition, under the assumption that L1 patterns would guide this process. To our knowledge, the experiment reported here constitutes the first study of the *comprehension,* rather than *production,* of ungrammatical bare NPs in L2 English by L1 speakers of an article-less language. Moreover, although there is evidence for asymmetries in the distribution of L2 English bare NPs (see Section 2.2), this study is the first to attempt to explain these asymmetries in terms of the semantics of NP (in)definiteness in the L1.

We contrasted the UH to the AH, which posits ambiguity in the (in)definiteness of bare nouns. Under the AH we would expect a random distribution of bare noun acceptability, irrespective of discourse context or syntactic position. The results of our study were clearly not consistent with the AH because the acceptability of ungrammatical English bare NPs was not equivalent across discourse contexts and syntactic positions. Rather, ungrammatical bare NPs in subject position were more acceptable when they occurred in contexts that required definites, following the pattern found in article-less languages, while bare objects received equivalent acceptability ratings across discourse contexts.

Although our results clearly demonstrate the presence of unambiguity in bare NP interpretation, they do not constitute unequivocal support for the UH as formulated above. Specifically, the prediction from the UH that bare objects in the PAS-I context would be relatively unacceptable, as would be expected for definite NPs, was not borne out. Although the relatively high acceptability of bare objects in this context could be explained as the result of ambiguity in the interpretation of number (Section 5.1), it is also possible that bare objects failed to demonstrate evidence for a definite interpretation because, unlike bare subjects, they lack a canonical association with topicality (Section 5.2). The follow-up experiments described in those sections will allow us to refine our understanding of the structure and limits of unambiguity in bare NP interpretation.

Importantly, our results demonstrate that factors like discourse context and syntactic position have predictable effects, not only on the use of bare NPs in L2 English, as already seen in the omission data of the production studies cited in Section 2.2, but also on the interpretation, and therefore the comprehension, of L2 bare nouns. Taken together, this set of results points to a distinct systematicity in L2 "errors" in the semantic domain, much like that already uncovered in the syntactic domain.

References

Almor, A. (1999). Noun-phrase anaphora and focus: The information load hypothesis. *Psychological Review*, 106, 748–765. https://doi.org/10.1037/0033-295X.106.4.748

Almor, A. (2005). A computational investigation of reference: Bridging the product and action traditions. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use. Bridging the language-as-product and language-as-action traditions* (pp. 285–301). The MIT Press.

Almor, A., & Nair, V. A. (2007). The form of referential expressions in discourse. *Language and Linguistic Compass*, 1, 84–99. https://doi.org/10.1111/j.1749-818X.2007.00009.x

Brisson, C. (1997). On definite plural NP's and the meaning of *all*. In A. Lawson (Ed.), *Proceedings of 7th Semantics and Linguistic Theory Conference (SALT VII)* (pp. 55–72). Cornell University. https://doi.org/10.3765/salt.v7i0.2791

Brisson, C. (2003). Plurals, *all*, and the nonuniformity of collective predication. *Linguistics and Philosophy*, 26, 129–184. https://doi.org/10.1023/A:1022771705575

Brown, G. (1995). *Speakers, hearers and communication: Explorations in discourse analysis*. Cambridge University Press. https://doi.org/10.1017/CBO9780511620942

Carlson, G. N. (1977). A unified analysis of the English bare plural. *Linguistics and Philosophy*, 1, 413–457. https://doi.org/10.1007/BF00353456

Chierchia, G. (1998). Reference to kinds across languages. *Natural Language Semantics*, 6, 339–405. https://doi.org/10.1023/A:1008324218506

Christensen, R. H. B. (2019). Ordinal – Regression models for ordinal data. *R Package*. Retrieved on 19 October 2019 from https://CRAN.R-project.org/package=ordinal

Cowart, W. (1997). *Experimental syntax*. Sage.

Dayal, V. (1999). Bare NPs, reference to kinds and incorporation. In T. Matthews & D. Strolovitch (Eds.), *Proceedings of Semantics and Linguistics Theory* 9, 35–51. CLC Publications.

Dayal, V. (2004). Number marking and (in)definiteness in kind terms. *Linguistics and Philosophy*, 27, 393–450. https://doi.org/10.1023/B:LING.0000024420.80324.67

Dayal, V. (2011). Hindi pseudo-incorporation. *Natural Language and Linguistic Theory*, 29, 123–167. https://doi.org/10.1007/s11049-011-9118-4

Dayal, V. (2017). Determining (in)definiteness in the absence of articles. In V. Hohaus & W. Rothe (Eds.), *Proceedings of TripleA*, 3 (pp. 85–99). Tübingen University Press.

Dayal, V. (2018). (In)definiteness without articles: Diagnosis, analysis, implications. In G. Sharma & R. Bhatt (Eds.), *Trends in Hindi linguistics* (pp. 1–26). Mouton de Gruyter. https://doi.org/10.1515/9783110610796-001

Dayal, V., & Jiang, L. J. (2021). The puzzle of anaphoric bare nouns in Mandarin: A counterpoint to index! *Linguistic Inquiry*, 1–42. https://doi.org/10.1162/ling_a_00433

Enç, M. (1991). The semantics of specificity. *Linguistic Inquiry*, 22, 1–25.

Goad, H., & White, L. (2006). Prosodic transfer: L1 effects on the production of L2 determiners. In D. Bamman, T. Magnitskaia, & C. Zaller (Eds.), *Proceedings of the 30th annual Boston University Conference on Language Development* (pp. 213–224). Cascadilla Press.

Goad, H., & White, L. (2009). Prosodic transfer and the representation of determiners in Turkish-English interlanguage. In N. Snape, Y.-k. I. Leung, & M. Sharwood-Smith (Eds.), *Representational deficits in SLA: Studies in honor of Roger Hawkins* (pp. 1–26). John Benjamins. https://doi.org/10.1075/lald.47.04goa

Heim, I. (1982). The semantics of definite and indefinite noun phrases (Unpublished doctoral dissertation). University of Massachusetts.

Heim, I. (2011). Definiteness and indefiniteness. In C. Maienborn, K. von Heusinger, & P. Portner (Eds.), *Semantics: A handbook of natural language meaning*. De Gruyter.

Huang, H. (2018). Revisiting pseudo-incorporation: Post-verbal non-referential bare NPs in Mandarin. In W. G. Bennett, L. Hracs, & D. R. Storoshenko (Eds.), *Proceedings of the 35th West Coast Conference on Formal Linguistics* (pp. 203–213). Cascadilla Proceedings Project.

Huebner, T. (1983). *A longitudinal analysis of the acquisition of English*. Karoma.

Huebner, T. (1985). System and variability in interlanguage syntax. *Language Learning*, 35, 141–163. https://doi.org/10.1111/j.1467-1770.1985.tb01022.x

Ionin, T., & Montrul, S. (2010). The role of L1 transfer in the interpretation of articles with definite plurals in L2 English. *Language Learning*, 60, 877–925. https://doi.org/10.1111/j.1467-9922.2010.00577.x

Ionin, T., Ko, H., & Wexler, K. (2004). Article semantics in L2 acquisition: The role of specificity. *Language Acquisition*, 12, 3–69. https://doi.org/10.1207/s15327817la1201_2

Jenks, P. (2018). Articulated definiteness without articles. *Linguistic Inquiry*, 49, 501–536. https://doi.org/10.1162/ling_a_00280

Jiang, L. J. (2012). Nominal arguments and language variation (Unpublished doctoral dissertation). Harvard University.

Jiang, L. J. (2018). Definiteness in Nuosu Yi and the theory of argument formation. *Linguistics and Philosophy*, 41, 1–39. https://doi.org/10.1007/s10988-017-9219-6

Kamp, H. (1981). A theory of truth and semantic representation. In J. A. G. Groenendijk, T. M. V. Janssen, & M. B. K. Stokhof (Eds.), *Formal methods in the study of language* (pp. 277–322). Mathematisch Centrum Amsterdam.

Lasersohn, P. N. (1999). Pragmatic halos. *Language*, 75, 522–551. https://doi.org/10.2307/417059

Law, J., & Syrett, K. (2017). Experimental evidence for the discourse potential of bare nouns in Mandarin. In A. Lamont & K. Tetzloff (Eds.), *Proceedings of the 47thannual North East Linguistics Society (NELS)* (Vol. 2, pp. 231–240). GLSA.

Lenth, R. (2020). Emmeans: Estimated marginal means, aka least-squares means. *R Package (1.4.4)*. Retrieved on 24 February 2022 from https://CRAN.R-project.org/package=emmeans

Li, C., & Thompson, S. (1976). Subject and topic: A new typology of language. In C. Li (Ed.), *Subject and topic* (pp. 459–489). Academic Press.

Liu, M., & Keusch, F. (2017). Effects of scale direction on response style of ordinal rating scales. *Journal of Official Statistics*, 33(1), 137–154. https://doi.org/10.1515/jos-2017-0008

Löbner, S. (1985). Definites. *Journal of Semantics*, 4, 279–326. https://doi.org/10.1093/jos/4.4.279

Master, P. (1987). A cross-linguistic interlanguage analysis of the acquisition of the English article system (Unpublished doctoral dissertation). University of California, Los Angeles.

Parrish, B. (1987). A new look at methodologies in the study of article acquisition for learners of ESL. *Language Learning*, 37, 361–383. https://doi.org/10.1111/j.1467-1770.1987.tb00576.x
R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved on 24 February 2022 from https://www.R-project.org/
Reinhart, T. (1981). *Pragmatics and linguistics: An analysis of sentence topics*. Indiana University.
Robertson, D. (2000). Variability in the use of the English article system by Chinese learners of English. *Second Language Research*, 16, 135–172. https://doi.org/10.1191/026765800672262975
Sanchez, L., & Giménez, M. J. (1998). The L2 acquisition of definite determiners: From null to overt. In A. Greenhill, M. Hughes, H. Littlefield, & H. Walsh (Eds.), *Proceedings of the 22nd annual Boston University Conference on Language Development* (pp. 640–650). Cascadilla Press.
Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime (2.0)*. [Computer software and manual]. Psychology Software Tools.
Schönenberger, M. (2014). Article use in L2 English by L1 Russian and L1 German speakers. *Zeitschrift für Sprachwissenschaft*, 33, 77–105. https://doi.org/10.1515/zfs-2014-0003
Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research* 12, 40–72. https://doi.org/10.1177/026765839601200103
Schwarz, F. (2013). Maximality and definite plurals – Experimental evidence. In E. Chemla, V. Homer, & G. Winterstein (Eds.), *Proceedings of Sinn und Bedeutung* 17 (pp. 509–526). Ecole Normale Supérieure.
Singmann, H., & Kellen, D. (2018). An introduction to mixed models for experimental psychology. In D. H. Spieler & E. Schumacher (Eds.), *New methods in cognitive psychology*. Psychology Press.
Thomas, M. (1989). The acquisition of English articles by first- and second-language learners of English. *Applied Psycholinguistics*, 10, 335–355. https://doi.org/10.1017/S0142716400008663
Trenkic, D. (2007). Variability in L2 article production: Beyond the representational deficit vs. processing constraints debate. *Second Language Research*, 23, 289–327. https://doi.org/10.1177/0267658307077643
Trenkic, D. (2009). Accounting for patterns of article omissions and substitutions in second language production. In R. Hawkins & M. P. G. Mayo (Eds.), *Second language acquisition of articles: Empirical findings and theoretical implications* (pp. 115–143). John Benjamins. https://doi.org/10.1075/lald.49.09tre
Walter, D., & Hentschel, J. (2013). *CaMLA English Placement Test (EPT) forms D-F: Development report*. CaMLA.
White, L. (2003). Fossilization in steady state L2 grammars: Persistent problems with inflectional morphology. *Bilingualism: Language and Cognition*, 6, 129–141. https://doi.org/10.1017/S1366728903001081
White, L. (2008). Different? Yes. Fundamentally? No. Definiteness in the L2 English of Mandarin speakers. In R. Slabakova, J. Rothman, P. Kempchinsky, & E. Gavruseva (Eds.), *Proceedings of the 9th Generative Approaches to Second Language Acquisition Conference (GASLA 2007)*, 251–261. Cascadilla Proceedings Project.
Yang, R. (2001). Common nouns, classifiers, and quantification in Chinese (Unpublished doctoral dissertation). Rutgers University.
Zdorenko, T., & Paradis, J. (2011). Articles in L2 English: When L1 and L2 acquisition meet at the interface. *First Language*, 32, 38–62. https://doi.org/10.1177/0142723710396797

L1-Mandarin L2-English learners' acquisition of English double-quantifier scope

Mien-Jen Wu and Tania Ionin
National Chung Cheng University / University of Illinois at Urbana-Champaign

English double-quantifier configurations such as *A dog scared every man* are ambiguous between a surface-scope reading (there exists one specific dog which scared every man) and an inverse-scope reading (each man was scared by a possibly different dog), while the Mandarin equivalent only has the surface-scope reading. Therefore, if L1-transfer is at work, L1-Mandarin L2-English learners would not initially allow inverse-scope readings in English, but may acquire them through exposure to relevant input. We tested learners in the U.S. and native English speakers on their acceptance of surface-scope and inverse-scope readings and found that learners disallowed inverse-scope readings of English double-quantifier sentences. This suggests that positive evidence alone is not sufficient for the L2-acquisition of inverse scope.

Keywords: second language acquisition, quantifier scope, positive evidence, English, Mandarin

1. Introduction

When learning a second language (L2), L2 learners may encounter linguistic properties in the L2 that are represented differently from those in their first language (L1), or even are completely absent in their L1. These situations can potentially cause learning difficulty for learners as they cannot depend on their L1 knowledge in those cases. With properties that are represented differently in the L1 and the L2, positive and/or negative evidence may be needed for learners to adjust their Interlanguage Grammar. When the L2 has structures or form – meaning mappings that do not exist in the L1, positive evidence informing learners that these properties exist in the L2 is crucial for acquisition.

https://doi.org/10.1075/lald.67.04wu

The learning scenario in this study involves a form – meaning mapping that exists in the learners' L2 (English) but is absent in the learners' L1 (Mandarin): inverse-scope readings of double-quantifier sentences. In English, such configurations, which contain quantifier phrases (QPs) in both subject and object positions, are ambiguous between two readings. We focus here on sentences with an indefinite quantifier in subject position, and a universal quantifier in object position, as in (1) and (2).

(1) A dog scared every man.

(2) One dog scared every man.

When the two quantifiers are interpreted in their surface order, the indefinite takes scope over the universal, resulting in the surface-scope reading paraphrased in (3a). On the inverse-scope reading paraphrased in (3b), the universal quantifier scopes over the indefinite. In a scenario where each man was scared by a different dog, the surface-scope reading in (3a) is false, but the inverse-scope reading in (3b) is true.

(3) Two Readings of (1) & (2):
  a. Surface-scope (a/one>every): There exists one specific dog that scared every man.
  b. Inverse-scope (every>a/one): For every man, there exists one (possibly different) dog which scared him.

English native speakers can access both scopal readings of (1) and (2) (though the surface-scope reading is preferred – more on this below). However, Mandarin is a scope-rigid language (Aoun & Li, 1989, 1994; Huang, 1982, 1983), which allows only the surface-scope reading (3a) for the Mandarin equivalent of (1)–(2). Therefore, L1-Mandarin L2-English learners have to add the inverse-scope reading of sentences such as (1)–(2) to their Interlanguage Grammar. Since English scope is not taught in L2 English classrooms, learners can only acquire inverse scope from naturalistic input. However, when Chu, Gabriele, and Minai (2014) tested L1-Mandarin L2-English learners in Taiwan, they found that even advanced learners had failed to acquire English inverse scope with English double-quantifier configurations. This suggests that the English input L1-Mandarin L2-English learners in Taiwan had received was not sufficient for them to learn that English double-quantifier configurations possess two readings. It is possible that learners may successfully acquire inverse scope if they receive more positive evidence.

To examine this hypothesis, this study tests L1-Mandarin L2-English learners immersed in an English environment (studying in the U.S.) on English double-quantifier configurations to see whether they successfully acquire English inverse scope. Learners in the U.S. are assumed to be exposed to more English input than those in an EFL (English as a Foreign Language) setting and should thus have a higher chance of successfully acquiring English inverse scope. The results should inform us as to whether a supposedly higher amount of input makes a difference in this population's acquisition of inverse scope.

2. Scope in English, Mandarin and L2-acquisition

2.1 English and Mandarin scope

English double-quantifier configurations such as (1) and (2) are known to be ambiguous between the two readings in (3a) and (3b). On standard accounts of quantifier scope (Heim & Kratzer, 1998; May, 1977, 1985), the surface-scope reading in (3a) is derived when the subject (e.g., *a dog* in (1)) goes through VP-internal subject movement and adjoins to IP, while the object undergoes short type-driven Quantifier Raising (QR) to a VP-adjoined position at LF. This is schematized in (4) (we are abstracting away from whether there is a vP projection above the VP, and only focus on the positions of the two quantifiers relative to one another). Thus, at LF, the subject c-commands and scopes over the object.

(4) $[_{IP}\ \underline{\text{a dog}}_i \ldots \underline{\text{every man}}_j\ [_{VP}\ t_i \text{ scared } t_j]]$

The inverse-scope reading in (3b) can be generated through further covert QP movement at LF. On one derivation (May, 1977), the object undergoes covert QR to a higher position than that of the subject at LF; in this case, *every man* scopes over *a dog* at LF, as schematized in (5a). On an alternative derivation (Fox, 1998), the subject QP lowers from its surface subject position back to its VP-internal subject position at LF and thus scopes below *every man*, which has undergone short movement to a VP-adjoined position, as in (5b). Thus, the two options for deriving inverse-scope readings of English double-quantifier sentences are covert quantifier raising (QR) and covert quantifier lowering (or reconstruction). For our purposes, it does not matter which of these options is adopted.

(5) a. $[_{IP}\ \underline{\text{every man}}_j\ [_{IP} \text{ a dog}_i \ldots \underline{\text{every man}}_j\ [_{VP}\ t_i \text{ scared } t_j]]]$

b. $[_{IP}\ \underline{\quad\quad} \ldots \text{every man}_j\ [_{VP}\ \underline{\text{a dog}} \text{ scared } t_j]]]$

In a context where one specific dog scared every man, (1) and (2) are true on both surface-scope and inverse-scope readings (since in this configuration, surface scope entails inverse scope: if one specific dog scared all the men, it follows that each man was scared by a dog). On the other hand, in a context where each man was scared by a different dog, (1) and (2) are false on the surface-scope reading but true on the inverse-scope reading; this is the crucial context for teasing apart the two readings.

Unlike English, Mandarin is a scope-rigid language, so the Mandarin translation of (2), in (6), only has the surface-scope reading (3a).[1]

(6) Yǒu yīzhī gǒu xiàle měigèrén
there exists one-classifier dog scare-ASP every man
'One dog scared every man.'

To explain the scope rigidity of Mandarin, S.-F. Huang (1981) proposed that quantifiers within a Mandarin sentence can only be interpreted linearly based on their surface word order. In response to S.-F. Huang's proposal, C.-T. J. Huang (1982, 1983) argued that Mandarin quantifiers are interpreted based on the c-commanding order at LF rather than on their linear order at the surface level: if QP1 c-commands QP2 at the LF, QP1 scopes over QP2. Aoun and Li (1989, 1994) later adopted C.-T. J. Huang's proposal and proposed *The Scope Principle* (see (7), Aoun & Li, 1989, p. 141; Aoun & Li, 1994, p. 11).

(7) *The Scope Principle*
A quantifier A has scope over a quantifier B in case A c-commands a member of the chain containing B.

Thus, the account of scope rigidity in Mandarin is framed in rather different terms than current accounts of inverse scope in English; we are not aware of any theoretical literature that explains why Mandarin does not allow covert quantifier movement as a way of deriving inverse scope. Ultimately, what matters for the purposes of our study is that the Mandarin sentence in (6) only has the surface-scope reading, regardless of what is causing the scope rigidity.

2.2 Experimental studies on English and Mandarin scope

2.2.1 *Experiments on English scope*

Experimental studies testing adult English native speakers (Anderson, 2004; Chu et al., 2014; Kurtzman & MacDonald, 1993; Scontras, Tsai, Mai, & Polinsky, 2014; Scontras, Polinsky, Tsai, & Mai, 2017) have found that though both the surface-scope and the inverse-scope readings are available in double-quantifier configurations, inverse-scope readings are dis-preferred. One explanation (see, e.g., Anderson, 2004; Tunstall, 1998) is that the covert movement required for deriving inverse-scope readings induces extra processing costs.

1. Mandarin does not have (in)definite articles, so (1) does not have a counterpart in Mandarin.

Online psycholinguistic studies have found that inverse-scope readings induce longer judging and/or reading times relative to surface-scope readings, in addition to being judged as less acceptable. Kurtzman and MacDonald (1993) tested adult English native speakers using a speeded continuation judgment task, in which participants read a double-quantifier sentence followed by a continuation sentence describing a context which supported either the surface-scope reading or the inverse-scope reading (e.g., *A kid climbed every tree.–>The kid was full of energy./ The kids were full of energy.*). Double-quantifier configurations with both existential QP subjects (*A kid climbed every tree)* and universal QP subjects (*Every kid climbed a tree*) were included. However, in the latter case, the inverse-scope reading entails the surface-scope reading (if there is one specific tree that was climbed by every kid, it also means that for every kid, there is one tree that he/she climbed), and, therefore, this sentence type cannot provide evidence for availability of inverse scope. With configurations such as *A kid climbed every tree*, participants accepted inverse-scope continuations (*The kids were full of energy*) to a lesser extent than surface-scope continuations (*The kid was full of energy*), which suggests that under time pressure, the inverse-scope readings were harder to access. Kurtzman and MacDonald proposed the Single Reference Principle to explain why the inverse-scope readings were dis-preferred. As readers encounter an indefinite QP at the beginning of the sentence, they assume that there exists only one agent. Holding this assumption, from the processing point of view, makes it harder for readers to access the inverse-scope readings, which in this case involve multiple agents.

Anderson (2004) administered a series of self-paced reading tasks to adult English native speakers and found that they read English double-quantifier sentences more slowly within contexts that disambiguated in favor of the inverse-scope reading compared to those that disambiguated in favor of the surface-scope reading; this suggests that inverse-scope readings are still hard to process even within supporting contexts. Anderson suggested that inverse-scope structures are syntactically more complex than surface-scope ones (cf. Fox, 1998; May, 1977) and proposed the principle of *Processing Scope Economy*, which states that speakers prefer (syntactically) simpler configurations because those configurations are easier to process.

Offline studies (Chu et al., 2014; Scontras et al., 2014, 2017) have shown that inverse-scope readings are dis-preferred even when participants are not time-pressured. Chu et al. (2014) tested adult English native speakers (as controls for their L1-Mandarin L2-English learners) on English active and passive double-quantifier configurations such as *Someone dropped every plate* and *Every plate was dropped by someone* in a Context/Picture-Matching Task. Chu et al. found that active sentences such as *Someone dropped every plate* were judged as significantly less acceptable in contexts that are only compatible with inverse-scope readings compared to those that are compatible with surface-scope readings. Passive configurations such as *Every plate was dropped by someone* were judged equally acceptable in both contexts. However, in these configurations, inverse-scope readings entail surface-scope ones (if there is one specific person who dropped every plate, it follows that for every plate, it was dropped by a person), so these results do not indicate availability of inverse scope.

Scontras et al. (2017) tested adult English native speakers (as controls for their English-dominant Mandarin heritage speakers) on English double-quantifier sentences such as *A/one shark attacked every pirate* and *Every shark attacked a/one pirate* (in which, again, the inverse-scope readings entail the surface-scope ones) in a Picture-based Acceptability Judgment Task.[2] With *A/one shark attacked every pirate*, they found significantly lower ratings with pictures supporting only the inverse-scope readings compared to those supporting the surface-scope readings. In addition, inverse-scope readings were found to be less acceptable when the indefinite subject QP contained *one* than when it contained *a*. Scontras et al. referred to Kurtzman and MacDonald (1993)'s Single Reference Principle in their explanation, as follows. When the participants encounter *a* or *one* at the beginning of a double-quantifier configuration, they might assume that there is just one agent in the context. However, *one* is more salient and noticeable than *a*, and thus *one* may induce a stronger preference than *a* for a single agent (and hence for surface scope, in this configuration), even in an offline task. The finding of a greater surface-scope preference with *one* relative to *a* in subject position on the part of native English speakers was also replicated in Ionin and Luchkina (2019) (native English speakers were tested in this study for comparison with English-dominant learners of Russian).

2.2.2 *Experiments on Mandarin scope*

Previous studies (Chu et al., 2014; Scontras et al., 2014, 2017) have also provided experimental evidence that inverse-scope readings are not allowed with Mandarin double-quantifier configurations. Chu et al. found that Mandarin double-quantifier configurations, such as *Mǒugèrén shuāipòle měigè pánzǐ* ('Someone dropped every plate'), were only accepted in contexts supporting the surface-scope reading. Similarly, Scontras et al. (2014, 2017) found that Mandarin double-quantifier configurations such as *Yǒuyītiáo shāyú gōngjīle měiyīgè hǎidào* ('A/One shark attacked every pirate') only had surface-scope readings. Therefore, these results clearly

2. Scontras et al. (2014) used a Truth-Value Judgment Task with pictures to test the same scope configurations; the results of the two studies were convergent, so we discuss only the 2017 results here.

indicate that, consistent with the theoretical literature, Mandarin is scope-rigid in double-quantifier configurations.

Since these studies tested both native English speakers and native Mandarin speakers using the same methodology, their results provide conclusive evidence that the two languages behave differently: in English, the inverse-scope readings were dis-preferred relative to the surface-scope ones, but were still allowed, whereas in Mandarin, the inverse-scope readings were completely disallowed. For example, on a 7-point Acceptability Judgment Task in Scontras et al. (2017), native English speakers gave a mean rating of 4.46 to sentences with *a* in the inverse-scope scenario, and a mean rating of 2.11 to sentences with *one*; native Mandarin speakers tested on Mandarin gave a mean rating of 1.56 (essentially at-floor) to sentences in the inverse-scope scenario. While inverse scope was rated fairly low for English sentence with *one*, these ratings were still significantly higher than the corresponding ratings in Mandarin.

Taken together, these findings indicate that surface scope of double-quantifier sentences is a preference in English, but a requirement in Mandarin.

2.2 Previous L2 and heritage language studies on scope

The only study on quantifier scope that we have found with L1-Mandarin L2-English learners is Chu et al. (2014). They found that intermediate and advanced L1-Mandarin L2-English learners in Taiwan experienced difficulty in acquiring inverse scope of English double-quantifier configurations like *Someone dropped every plate*. The learners accepted these configurations in contexts that supported the surface-scope readings but rejected them in contexts that supported only the inverse-scope readings; the learners behaved the same when tested in both English and Mandarin. Chu et al. suggested that learners in an EFL setting may not receive sufficient positive evidence for the existence of inverse scope in English. In addition, Chu et al. proposed that the learners may simply dis-prefer inverse-scope readings due to *Processing Scope Economy*. A problem for this explanation, however, is the fact that English native speakers' ratings of the inverse-scope readings were still higher than those of the learners'. If the results of both groups are exclusively due to processing preferences, we would not expect such a difference; the fact that one exists suggests that learners are affected by their L1, since Mandarin does not allow inverse scope. Testing another group of L2-English learners with an L1 that allows inverse scope would help establish whether L1 transfer is indeed at work here, or whether a dis-preference for inverse scope is a feature of L2-acquisition independently of the L1. Still, whatever the reason, Chu et al.'s (2014) results indicate that advanced L1-Mandarin L2-English learners in an EFL context are not target-like on English inverse scope.

Having English and Mandarin native speakers' performance as the baseline (see previous sections), Scontras et al. (2017) tested English-dominant Mandarin heritage speakers on both English and Mandarin double-quantifier sentences such as *A/One shark attacked every pirate*. They found that the heritage speakers rated the inverse-scope readings in English significantly lower than English native speakers (2.55 on a 7-point scale, compared to 4.46 for native speakers, for sentences with *a*). At the same time, the heritage speakers were more accepting than Mandarin native speakers of inverse scope in Mandarin. However, Scontras et al. did not attribute this to transfer from either English or Mandarin. Since the heritage speakers differed from English native speakers and largely rejected inverse-scope readings, it was unlikely that they had fully incorporated inverse scope into their English grammar, and hence unlikely that they were subject to transfer of inverse scope from English to Mandarin. As for the possibility of transferring from Mandarin, since Mandarin was the weaker language of the heritage speakers, it was more likely that the heritage speakers were simply applying the more easily accessible scope configuration (i.e., the surface-scope configuration, which happens to also be the only one available in Mandarin). Also, given that heritage speakers usually lack confidence in judgments in their weaker language, their slightly higher ratings of inverse scope in Mandarin (relatively to native Mandarin speakers) could be due to yes-bias.

We now turn to findings on the L2 acquisition of inverse scope in other languages. Marsden (2009) conducted a cross-sectional study with L1-Korean and L1-English L2-Japanese learners on Japanese double-quantifier sentences; the results provide evidence for L1 transfer in the L2 acquisition of scope. As Korean and Japanese behave similarly in terms of double-quantifier scope, positive transfer from Korean to Japanese was expected. In both languages, canonical SOV double-quantifier sentences only have surface-scope readings, while scrambled OSV double-quantifier sentences are ambiguous between surface-scope and inverse-scope readings. In contrast, there should be more difficulty for L1-English L2-Japanese learners because scope is represented differently between the two languages. In Marsden's study, only some advanced L1-English L2-Japanese learners were target-like, while L1-Korean L2-Japanese learners mastered Japanese double-quantifier scope even at low L2 proficiency; therefore, Marsden attributed L1-English L2-Japanese learners' learning difficulty (specifically, their over-acceptance of inverse scope) to L1 transfer.

In summary, previous studies have shown that advanced L1-Mandarin L2-English learners (in an EFL setting) have difficulty acquiring English inverse scope, and that L2 acquisition of scope is subject to L1 transfer. While there has been no study directly comparing L1-Mandarin L2-English learners to a different L1-group, it is reasonable to suppose that L1-transfer is operative with L1-Mandarin L2-English learners, given Marsden's finding of L1-transfer with scope in other populations. Alternatively, it could be that L1-Mandarin L2-English learners have difficulty with inverse scope in English because surface scope is easier to process, as Scontras et al. (2017) suggest for English-dominant Mandarin heritage speakers. Either way, L1-Mandarin L2-English learners start out assuming that English does not allow inverse scope. In order to learn that English does allow inverse scope, these learners need to attend to positive evidence in the input. We turn to this next.

3. Scope in L2-English: The role of positive evidence

3.1 Positive evidence and L2-acquisition

As discussed above, it is possible that the L1-Mandarin L2-English learners in Chu et al. (2014) were affected by negative transfer from their L1. Under the Full Transfer/Full Access model (Schwartz & Sprouse, 1996), learners can potentially recover from L1-transfer as long as the particular L1/L2 combination does not result in a learnability problem. In this particular case, there is no learnability problem, since the learners' task is to add a new reading to their grammar (rather than to unlearn an existing reading), and such learning should in principle be able to proceed on the basis of positive evidence in the input. Specifically, learners need to attend to double-quantifier sentences that are used in contexts which support the inverse-scope but not the surface-scope readings; this is the evidence indicating that English double-quantifier sentences have inverse-scope readings.

However, there are certain limitations to positive evidence in L2 acquisition. For instance, if the relevant examples are not frequent or robustly available in the input, learners may not notice the target property and thus fail to acquire it. For instance, Montrul (2001) showed that L1-Spanish L2-English learners failed to acquire the transitive structure of some English manner-of-motion verbs (e.g., *The captain marched the soldiers to the tent*) due to it being less typical than the intransitive one (e.g., *The soldiers marched*).

In addition, White (1991) pointed out that if the L2 property and the L1 property form a superset-subset relationship, learners may be misled to believe that the two languages behave the same, and more so when the overlapping structures/interpretations happen to be the more frequent ones in the L2. For instance, Inagaki (2002) found that L1-Japanese L2-English learners failed to acquire the directional reading of English prepositions. Some English prepositions (e.g., *under*) are ambiguous between a directional reading (e.g., *John swam under the bridge,* meaning that John swam from under one side of the bridge to the other side) and a locational reading (e.g., *John swam under the bridge,* meaning the location where John swam was under the bridge) after a manner-of-motion verb. The corresponding prepositions are unambiguously locational in Japanese, so L1-Japanese L2-English learners have to learn the additional directional readings of the English prepositions. Inagaki (2002) found that intermediate learners failed to learn the directional readings and attributed this result to insufficient positive evidence as well as negative L1 transfer. Because the target property forms a superset-subset relationship between the two languages (the English case is a superset of the Japanese one), and the more frequent readings (the locational ones) happen to be the ones also shared with the L1, learners may assume that English behaves just like Japanese, with only the locational readings.

L1-Mandarin L2-English learners learning English inverse scope should experience a similar difficulty. Though both the surface-scope and the inverse-scope readings are available in English, the surface-scope ones are preferred and happen to be the only readings available in Mandarin. Given that inverse-scope readings are dis-preferred for English native speakers, the relevant positive evidence may be infrequent (at least compared to that of surface-scope readings), so learners may think that English only has the surface-scope readings, like Mandarin.

Nevertheless, by testing advanced learners in the U.S., we can at least examine whether exposing learners to naturalistic input from English native speakers can lead to acquisition of English inverse scope. The results can then be compared to learners who are in an EFL setting, as in Chu et al. (2014).

3.2 Research question

Chu et al. (2014) found that even advanced L1-Mandarin L2-English learners in Taiwan failed to successfully acquire English inverse scope. Our study examines whether advanced learners immersed in English in the U.S., and hence possibly receiving more positive evidence than learners in an EFL setting, are able to acquire English inverse scope. This study thus has the following research question:

RQ: Can intermediate-to-advanced L1-Mandarin L2-English learners in the U.S. successfully acquire inverse scope in English double-quantifier configurations?

4. Method

4.1 Task materials

We conducted a Context-based Acceptability Judgment Task (similar to the one in Chu et al. 2014), in which participants rated each sentence on a scale from 1 (totally unacceptable) to 4 (totally acceptable) against the context of a short story accompanied by a picture. Two types of contexts were used (see Figure 1): the "Single Agent"

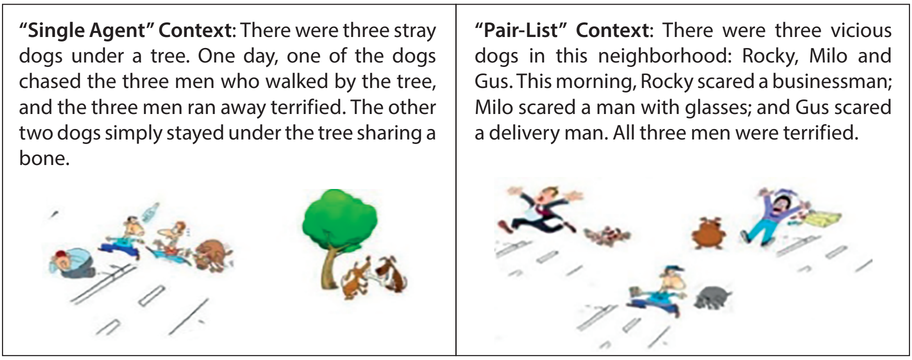


Figure 1. Sample contexts

context, in which only one agent does the action; and the "Pair-List" context, in which multiple agents do the action. The pictures accompanying the stories were in part based on the pictures used in Ionin and Luchkina (2019)'s work on scope in Russian.

There were two types of target sentences: *a-every*, as in (1), repeated in (8); and *one-every*, as in (2), repeated in (9). Each context was paired with both sentence types, so that both (8) and (9) were presented with each context in Figure 1.

(8) A dog scared every man.

(9) One dog scared every man.

The surface-scope reading of (8) and (9), on which there exists one specific dog that scared every man, makes the sentences true only in the "Single Agent" context. The inverse-scope reading of (8) and (9), on which a (potentially different) dog scared each man, makes the sentences true in both the "Single Agent" context and the "Pair-List" context. Since the "Pair-List" context is only true under the inverse-scope reading, this context type is the key for determining whether participants can access the inverse-scope readings.

The two sentence types (*a-every* and *one-every*) were distributed across two lists using a Latin-square design; within each list, each sentence occurred with both contexts, six tokens per context/sentence mapping. Each list thus contained 24 target items (four conditions of six tokens each, where a condition corresponds to one of the two context types paired with one of the two target sentence types), as well as 72 fillers. The fillers tested participants on unrelated phenomena, such as scalar implicature with numeral NPs (e.g., The sentence *Two triangles were colored* with a context indicating that three triangles were colored.).

It was predicted that English native speakers would accept *a-every* and *one-every* sentences in both the "Single Agent" and the "Pair-List" contexts, though ratings should be higher in the "Single Agent" contexts than the "Pair-List" contexts, given the known preference for surface scope in English. In addition, given the prior findings of Scontras et al. (2014, 2017) as well as Ionin and Luchkina (2019), the surface-scope preference was expected to be stronger for *one* than for *a*. Therefore, the acceptance rate in the "Pair-List" context for *one-every* sentences was expected to be lower than that for *a-every* sentences (see Table 1).

Table 1. Sample English target sentences, with predictions for English native speakers' performance

Sample target items	"Single agent" context	"Pair-list" context
a-every: A dog scared every man.	√	?
one-every: One dog scared every man.	√	??

Note. Acceptable:√
Acceptable but dis-preferred: ?
Acceptable but strongly dis-preferred ??

In addition to being tested on English, L1-Mandarin L2-English learners were also tested on Mandarin with the corresponding Mandarin sentences and contexts translated from the English Task (see Table 2). In the Mandarin Task, because Mandarin does not have articles, *a* was translated as *yǒu+classifier* (*yǒu* is an existential marker in Mandarin), and thus the type *a-every* in English is referred to as the type *classifier-every* in Mandarin. *One* was translated as *yǒu+yī+classifier* (*yī* means *one* in Mandarin), and thus the type *one-every* in English is referred to as the type *one+classifier-every* in Mandarin. Mandarin *classifier-every* and *one+classifier-every* sentences are unambiguous and only have the surface-scope readings; therefore, they were predicted to be unacceptable in the "Pair-List" contexts (see Table 2).

Table 2. Sample Mandarin target sentences, with predictions for Mandarin native speakers' performance

Sample target items	"Single agent" context	"Pair-list" context
classifier-every: *Yǒu zhī gǒu xiàle měigèrén.* "A dog scared every man."	√	×
one+classifier-every: *Yǒu yīzhīgǒu xiàle měigèrén.* "One dog scared every man."	√	×

Note. Acceptable: √
Unacceptable: ×

The L1-Mandarin L2-English learners' judgments in Mandarin were used as the baseline for their performance in English, allowing us to compare learners' acceptance of inverse-scope readings in the two languages.

4.2 Procedure

The English native participants in this study completed the English version of the Context-based Acceptability Judgment Task and a language background questionnaire. The L1-Mandarin L2-English learners took part in two sessions. In the first session, the learners completed the English version of the Judgment Task, a language background questionnaire, and an English proficiency test; in the second session, they completed the Mandarin version of the Judgment Task. To minimize any priming effects from English to Mandarin, we separated the two testing sessions by two weeks. All test materials were implemented via the online survey tool Alchemer (formerly SurveyGizmo).

The English proficiency test that L1-Mandarin L2-English learners completed was a multiple-choice English cloze test. The cloze test contained 40 multiple-choice items, and for each item, the participants had to choose the best answer out of three. The same test was used in Ionin and Montrul (2010), who found that it was reliable (*Cronbach* $\alpha = 0.817$). Since Ionin and Montrul found that native English speakers performed at ceiling on this test, we did not administer this test to our English native speaker participants.

4.3 Participants

Data from 26 English native speakers and 24 L1-Mandarin L2-English learners were collected. All participants were recruited in a university in the U.S. However, participants were excluded from data analysis if they failed to accept (that is, to rate as 3 or 4 on a 4-point scale) at least 10 out of 12 unambiguously true filler sentences with numeral NPs; a sample filler is given in (10). L1-Mandarin L2-English learners were excluded from the analysis of both tasks if they failed to reach the criterion on either the English Task or the Mandarin Task.

(10) Joanna was making sandwiches for her kids. She buttered three sandwiches and put them in the lunchboxes.
Sentence: *Three sandwiches were buttered.*

Based on this criterion, the participants included in the data analysis were 19 English native speakers (Age Range: 18–23; 1 male and 18 females) and 18 L1-Mandarin L2-English learners (Age Range: 18–24; 5 males and 13 females). The L1-Mandarin L2-English learners included in the analysis had been living in the U.S. between two months and six years at the time of the study; 10 learners had been in the U.S. for less than six months, three had been there for one year, two for two years, one for three years, one for five years, and one for six years. One came to the U.S. at age 16, while the others came after age 18. Regarding their English proficiency, learners had an average score of 34.17 out of 40 (Range: 31–38; SD: 2.36) on the English proficiency test and were thus considered to be intermediate-to-advanced English learners.

5. Results

5.1 Descriptive results

Figures 2 through 4 show the descriptive results of the English native speakers on the English Judgment Task (Figure 2), the L1-Mandarin L2-English learners on the English Judgment Task (Figure 3), and the L1-Mandarin L2-English learners on the Mandarin Judgment Task (Figure 4). As shown in Figure 2, English native speakers allowed both *a-every* and *one-every* sentences in both the "Single Agent" and "Pair-List" contexts. As predicted, the average ratings in the "Pair-List" contexts were lower than those of the "Single Agent" contexts, while the average ratings of *one-every* sentences in the "Pair-List" contexts were lower than those of *a-every* sentences. As shown in Figure 3, L1-Mandarin L2-English learners allowed both *a-every* and *one-every* sentences in the "Single Agent" contexts. However, the average ratings of both *a-every* and *one-every* sentences in the "Pair-List" contexts were much lower than those of English native speakers, while at the same time the average ratings of *one-every* sentences in the "Pair-List" contexts were also lower than those of *a-every* sentences in the "Pair-List" contexts. Finally, as shown in Figure 4, L1-Mandarin L2-English learners allowed both *classifier-every* and *one+classifier-every* sentences in the "Single Agent" contexts. The ratings of *classifier-every* and *one+classifier-every* sentences were almost at-floor in the "Pair-List" contexts.

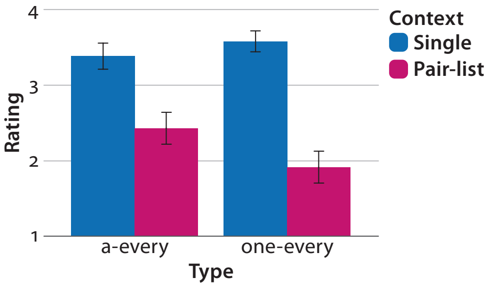


Figure 2. Average ratings in the English Task (English native speakers)[3]

3. The error bar represents standard error, which indicates how concentrated the data points are. The shorter the error bar the more concentrated the data points are, in which case the average rating better reflects participants' judgment/performance.

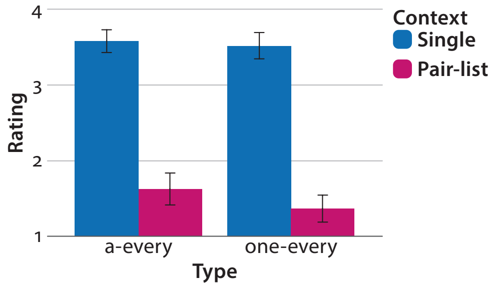


Figure 3. Average ratings in the English Task (L1-Mandarin L2-English learners)

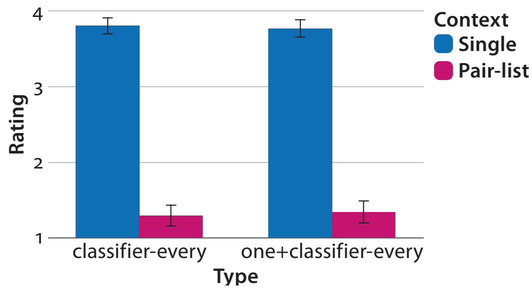


Figure 4. Average ratings in the Mandarin Task (L1-Mandarin L2-English learners)

5.2 Statistical analysis

The data from the target items in the Context-based Acceptability Judgment Task were analyzed using the Cumulative Link Mixed Model in R (Christensen, 2018), which analyzes data similarly to Linear Mixed-Effects Models. The only difference is that the dependent variable is ordinal, such as ratings on a scale. Three models were run. Model 1 analyzed the data of both native speakers and learners on the English Judgment Task; Model 2 analyzed the data of learners on both the English and the Mandarin Judgment Tasks; and Model 3 analyzed the data of learners on the English Judgment Task, this time with proficiency scores as a factor. The fixed effects in Model 1 were type (*a-every* and *one-every*; 2 levels), context ("Single Agent" and "Pair-List"; 2 levels) and group (English native speakers and L1-Mandarin L2-English learners; 2 levels). The fixed effects in Model 2 were type (*a/classifier-every* and *one/one+classifier-every*; 2 levels), context ("Single Agent" and "Pair-List"; 2 levels) and language (English and Mandarin; 2 levels). The fixed effects in Model 3 were type (*a-every* and *one-every*; 2 levels), context ("Single Agent" and "Pair-List"; 2 levels) and proficiency (numerical, range 31–38). In each model, the random effects were participants ($N = 37$ in Model 1; $N = 18$ in Models 2 and 3) and items ($N = 24$).

In the Model 1 output (see Table 3), there was a significant effect of context, but no significant effect of type or group. Since there was a significant interaction of context and group, pairwise comparisons were run on Model 1 output using the 'emmeans' package in R (Lenth, 2018).

Table 3. Model 1 output results

	Estimate	Std. error	Z value	P value
type	0.0840	0.3575	0.235	0.814
context	−4.1424	0.3718	−11.142	<.0001 *
group	−0.5670	0.4329	−1.310	0.190
type*context	−0.8470	0.5195	−1.631	0.103
type*group	0.4323	0.4383	0.986	0.324
context*group	2.1681	0.4236	5.119	<.0001 *
type*context*group	−0.4204	0.6150	−0.684	0.494

Note. * indicates $p < 0.5$.

The pairwise comparison results, in Table 4, show that both English native speakers and L1-Mandarin L2-English learners rated English *a-every* and *one-every* sentences significantly lower in the "Pair-List" contexts than in the "Single Agent" contexts. However, learners' ratings of those sentences in the "Pair-List" contexts were significantly lower than those of English native speakers, with no corresponding difference in the "Single Agent" context. This indicates that learners did not accept English inverse-scope readings as much as English native speakers did. Interestingly, there were no statistical differences in the ratings of sentences with *a* vs. with *one* in the "Pair-List" context, even though descriptively, sentences with *a* were rated higher in this context than those with *one*.

Table 4. Model 1 pairwise comparison results

Pair	Estimate	P value
a, single, native > *a*, pair-list, native	1.9743	<.0001 *
one, single, native > *one*, pair-list, native	3.2417	<.0001 *
a, single, learner > *a*, pair-list, learner	4.1424	<.0001 *
one, single, learner > *one*, pair-list, learner	4.9894	<.0001 *
a, single, native = *a*, single, learner	−0.5670	0.8955
one, single, native = *one*, single, learner	−0.1346	1.0000
a, pair-list, native > *a*, pair-list, learner	1.6011	<.05 *
one, pair-list, native > *one*, pair-list, learner	1.6130	<.05 *
a, single, native = *one*, single, native	−0.5164	0.7665
a, pair-list, native = *one*, pair-list, native	0.7510	0.1911
a, single, learner = *one*, single, learner	−0.0840	1.0000
a, pair-list, learner = *one*, pair-list, learner	0.7629	0.4851

Note. * indicates $p < .05$.

In Model 2 output (see Table 5), there were significant effects of context and language, but no significant effect of type. Since there was a significant interaction of context and language, pairwise comparisons were run on Model 2 output using 'emmeans'.

Table 5. Model 2 output results

	Estimate	Std. error	Z value	P value
type	0.0616	0.3604	0.171	0.864
context	−4.1036	0.3745	−10.959	<.0001 *
language	0.8234	0.3927	2.097	<.05 *
type*context	−0.9984	0.5232	−1.908	0.056
type*language	−0.1008	0.5990	−0.168	0.866
context*language	−1.9471	0.5476	−3.556	<.001 *
type*context*language	1.1695	0.8361	1.399	0.162

Note. * indicates $p < .05$.

As shown in Table 6, L1-Mandarin L2-English learners rated both types of sentences significantly lower in the "Pair-List" contexts than in the "Single Agent" contexts, in both the English and the Mandarin tasks. The significant interaction between context and language in Model 2 stemmed from a greater difference in the ratings of sentences between the "Single Agent" and "Pair-List" contexts in the Mandarin Task, relative to the English Task (see Figures 3 vs. 4). At the same time, as shown in Table 6, the pairwise comparisons yielded no significant differences in the ratings of any condition in the two languages. Thus, the learners gave the same patterns of ratings in their two languages, consistent with L1-transfer.

Table 6. Model 2 pairwise comparison results

Pair	Estimate	P-value
a, single, English > *a*, pair-list, English	4.1036	<.0001 *
one, single, English > *one*, pair-list, English	5.1020	<.0001 *
cl, single, Mandarin > *cl*, pair-list, Mandarin	6.0508	<.0001 *
one+cl, single, Mandarin > *one+cl*, pair-list, Mandarin	5.8797	<.0001 *
a, single, English = *cl*, single, Mandarin	−0.8234	0.417
one, single, English = *one+cl*, single, Mandarin	−0.7226	0.604
a, pair-list, English = *cl*, pair-list, Mandarin	1.1237	0.072
one, pair-list, English = *one+cl*, pair-list, Mandarin	0.0551	1.0000
a, single, English = *one*, single, English	−0.0616	1.0000
a, pair-list, English = *one*, pair-list, English	0.9367	0.228
cl, single, Mandarin = *one+cl*, single, Mandarin	0.0392	1.0000
cl, pair-list, Mandarin = *one+cl*, pair-list, Mandarin	−0.1319	1.0000

Note. * indicates $p < .05$; "cl" stands for "classifier".

In Model 3 output (see Table 7), there was a significant effect of context, but no significant effect of type or proficiency. There was a significant interaction of context and proficiency. To explore the interaction, we plotted the relationship between context and proficiency (see Figure 5).

Table 7. Model 3 output results

	Estimate	Std. error	Z value	P value
type	−2.9132	4.8985	−0.595	0.552
context	−17.5601	4.8333	−3.633	<.001*
proficiency	−0.0435	0.1274	−0.342	0.733
type*context	−3.7495	7.4868	−0.501	0.617
type*proficiency	0.0858	0.1435	0.598	0.550
context*proficiency	0.4092	0.1395	2.933	<.01*
type*context*proficiency	0.0751	0.2152	0.349	0.727

Note. * indicates $p < .05$.

Figure 5 shows that learners with proficiency scores of 37 and 38 (5 out of 18 learners) had higher average ratings than learners with lower proficiency scores, which suggests that higher proficiency correlates with higher acceptance of inverse scope. However, as a group, learners still failed to perform in a native-like manner on inverse scope (see Model 1 pairwise comparison results in Table 4).

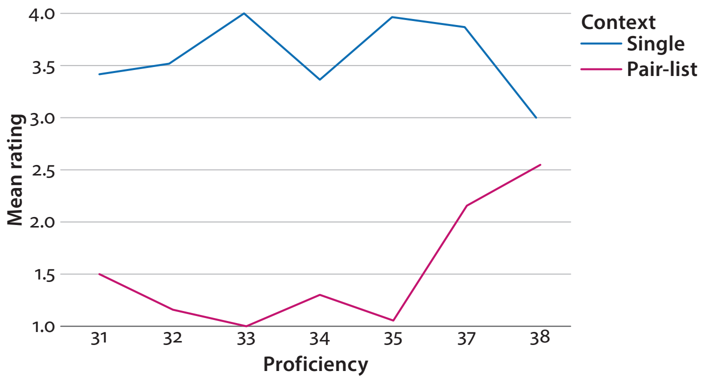


Figure 5. Interaction plot (context*proficiency)

6. Discussion

First, English native speakers performed largely as expected in this study. Both the surface-scope and the inverse-scope readings were available with English double-quantifier configurations, but the inverse-scope readings were dis-preferred compared to the surface-scope ones. This is fully consistent with prior literature. The one place where our findings diverged from prior literature was in the lack of a significant difference in the ratings of *one-every* vs. *a-every* sentences in the context that matched inverse scope (the "Pair-List" context). Prior literature (Ionin & Luchkina, 2019; Scontras et al., 2014, 2017) found inverse scope to be significantly more acceptable for sentences with *a* than with *one* in the subject position. It is not clear why we did not replicate this finding, though we note that descriptively, sentences with *a* were rated higher than sentences with *one* in the "Pair-List" context.

Turning to the L1-Mandarin L2-English learners' performance in the English Judgment Task, we see that the ratings in the "Pair-List" contexts were significantly lower than in the "Single Agent" contexts, and that, crucially, the learners gave significantly lower ratings than English native speakers only in the "Pair-List" context and not in the "Single Agent" context. This indicates that the learners overall were not accessing the inverse-scope reading in English to the same extent as the native English speakers, despite having high English proficiency and residing in the U.S.

Comparing the learners' performance on the English and Mandarin Judgment Tasks, we see largely the same patterns of performance. Numerically, the ratings of *a-every* sentences in the "Pair-List" contexts were higher in English than the corresponding ratings of *classifier-every* sentences in Mandarin. However, this difference failed to reach significance in the pairwise comparison results. This suggests that the learners largely treated English as Mandarin, disallowing inverse scope. It is reasonable to suppose that the L1-Mandarin L2-English learners' performance was due to negative L1-transfer from Mandarin. However, we cannot completely exclude the possibility that their rejection of inverse scope was instead due to the Scope Processing Economy principle (as explained in Section 2.3); however, as discussed earlier, this principle is supposed to apply to native speakers as well, so it is not clear how the L2/NS difference can be explained. Nevertheless, in order to definitively test the L1-transfer explanation, we would need to compare L1-Mandarin L2-English learners' performance to that of another group of L2-English learners whose L1 is not scope-rigid. If L1-transfer is at work, then the learner group whose L1 is not scope-rigid should accept English inverse scope more than L1-Mandarin L2-English learners do.

The findings in this study imply that the naturalistic input available to learners immersed in an English environment is still not sufficient for L1-Mandarin L2-English learners to realize that English double-quantifier configurations are ambiguous between a surface-scope and an inverse-scope reading. However, since around half of the learners in this study had only resided in the U.S. for less than six months, our learners might not have received much more input than Chu et al. (2014)'s learners did. Also, as Chu et al. suggested, one possible reason that inverse scope is so difficult to acquire could be because even English native speakers dis-prefer the inverse-scope readings, which means that there is probably insufficient input regarding the relevant readings available to learners. The dis-preference for inverse scope on the part of adult native English speakers means that children acquiring English as their L1 also receive less input with inverse-scope readings compared to surface-scope ones; however, children still manage to acquire inverse scope (even if it is dis-preferred). Unlike L1-acquiring children, L1-Mandarin L2-English learners face the additional obstacle of negative L1 transfer, resulting from the subset-superset relationship between the two languages, which is possibly why even advanced learners fail to acquire inverse scope. Therefore, future research should test learners with at least several years of U.S. residence, to see if they would be more successful on inverse scope. If even highly advanced learners with many years of exposure to naturalistic English input are unable to acquire inverse scope, this would suggest that adult L2-learners aren't able to learn less frequent structures from the input the way that young L1-learners do.

Convergent evidence regarding difficulty with inverse scope comes from Wu and Ionin (2019), a study that tested advanced L1-Mandarin L2-English learners in the U.S. on English quantifier-negation configurations (e.g., *Every sheep didn't jump over the fence*). In this study, learners also consistently rejected inverse-scope readings, while native English speakers accepted these readings quite readily. This suggests that advanced L1-Mandarin L2-English learners experience difficulty in acquiring inverse scope in English across two different configurations. Therefore, this learner population may need additional input or instruction in order to incorporate inverse-scope readings in their Interlanguage Grammar.

The question now becomes what kind of input or evidence is needed for learners to acquire the inverse-scope readings. If our learners failed because they did not notice the inverse-scope readings due to low amount of positive evidence, providing learners with artificially increased input (in the form of Input Flooding, cf. Hernández, 2011; Trahey, 1996) could potentially lead to successful acquisition. Of course, in this case, how much positive evidence learners require is uncertain, and the amount may vary by individual. If Input Flooding fails to work, it is possible that inverse scope would only be acquired with explicit instruction, when learners are explicitly taught about the availability of inverse scope. These possibilities were pursued in an intervention study (Wu, 2021) with advanced L1-Mandarin L2-English learners.

7. Conclusion

This study found that intermediate-to-advanced L1-Mandarin L2-English learners immersed in an English environment failed to perform in a target-like manner with English double-quantifier configurations. The findings were largely the same as in Chu et al. (2014), even though our learners were in the U.S., while Chu et al.'s were in Taiwan. These results suggested that naturalistic input in an English environment is not sufficient for learners to incorporate inverse-scope readings into their Interlanguage grammar.

References

Anderson, C. (2004). The structure and real-time comprehension of quantifier scope ambiguity (Unpublished doctoral dissertation). Northwestern University.

Aoun, J., & Li, Y.-H. (1989). Scope and constituency. *Linguistic Inquiry*, 20, 141–172.

Aoun, J., & Li, Y.-H. (1994). *Syntax of scope*. The MIT Press.

Christensen, R. H. B. (2018). *Ordinal – Regression models for ordinal data*. R package version 2018. 8–25. Retrieved on 24 Febrary 2022 from http://www.cran.r-project.org/package=ordinal/

Chu, C.-Y., Gabriele, A., & Minai, U. (2014). Acquisition of quantifier scope interpretation by Chinese-speaking learners of English. In C.-Y. Chu, C. E. Coughlin, B. Lopez Prego, U. Minai, & A. Tremblay (Eds.), *Selected Proceedings of the 5th Conference on Generative Approaches to Language Acquisition North America (GALANA 2012)* (157–168). Cascadilla Proceedings Project.

Fox, D. J. (1998). Economy and semantic interpretation – A study of scope and variable binding (Unpublished doctoral dissertation). MIT.

Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar*. Blackwell.

Hernández, T. A. (2011). Re-examining the role of explicit instruction and input flood on the acquisition of Spanish discourse markers. *Language Teaching Research*, 15(2), 159–182. https://doi.org/10.1177/1362168810388694

Huang, C.-T. (1982). Logical relations in Chinese and the theory of grammar. (Unpublished doctoral dissertation). MIT.

Huang, C.-T. (1983). On the representation of scope in Chinese. *Journal of Chinese Linguistics*, 11(1), 36–91.

Huang, S.-F. (1981). On the scope phenomena of Chinese quantifiers. *Journal of Chinese Linguistics*, 9, 226–243.

Inagaki, S. (2002). Japanese learners' acquisition of English manner-of-motion verbs with location/directional PPs. *Second Language Research*, 18(1), 3–27. https://doi.org/10.1191/0267658302sr196oa

Ionin, T., & Luchkina, T. (2019). Scope, syntax, and prosody in Russian as a second or heritage language. In M. Cabrera & J. Camacho (Eds.), *Exploring interfaces* (pp. 141–170). Cambridge University Press. https://doi.org/10.1017/9781108674195.008

Ionin, T., & Montrul, S. (2010). The role of L1 transfer in the interpretation of articles with definite plurals in L2 English. *Language Learning*, 60(4), 877–925. https://doi.org/10.1111/j.1467-9922.2010.00577.x

Kurtzman, H. S., & MacDonald, M. C. (1993). Resolution of quantifier scope ambiguities. *Cognition*, 48, 243–279. https://doi.org/10.1016/0010-0277(93)90042-T

Lenth, R. (2018). Package "emmeans". R package version 1.3.3. Retrieved on 24 February 2022 from https://cran.r-project.org/web/packages/emmeans/emmeans.pdf

Marsden, H. (2009). Distributive quantifier scope in English-Japanese and Korean-Japanese interlanguage. *Language Acquisition*, 16(3), 135–177. https://doi.org/10.1080/10489220902967135

May, R. (1977). The grammar of quantification grammar (Unpublished doctoral dissertation). MIT.

May, R. (1985). *Logical Form: Its structure and derivation*. The MIT Press.

Montrul, S. (2001). Agentive verbs of manner of motion in Spanish and English as second languages. *Studies in Second Language Acquisition*, 23, 171–206. https://doi.org/10.1017/S0272263101002030

Schwartz, B., & Sprouse, R. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research*, 12(1), 40–72. https://doi.org/10.1177/026765839601200103

Scontras, G., Polinsky M., Tsai, C.-Y., & Mai, K. (2017). Cross-linguistic scope ambiguity: When two systems meet. *Glossa: A Journal of General Linguistics*, 2(1), 1–28. https://doi.org/10.5334/gjgl.198

Scontras, G., Tsai, C.-Y., Mai, K., & Polinsky, M. (2014). Chinese scope: An experimental investigation. In U. Etxeberria, A. Fălăuş, A. Irurtzun, & B. Leferman (Eds.), *Proceedings of Sinn und Bedeutung* 18 (pp. 396–414). Retrieved on 24 February 2022 from https://semanticsarchive.net/sub2013/ProceedingsSuB18.pdf

Trahey, M. (1996). Positive evidence in second language acquisition: Some long-term effects. *Second Language Research*, 12(2), 111–139. https://doi.org/10.1177/026765839601200201

Tunstall, S. L. (1998). The interpretation of quantifiers: Semantics & processing (Unpublished doctoral dissertation). MIT.

White, L. (1991). Argument structure in second language acquisition. *French Language Studies*, 1, 189–207. https://doi.org/10.1017/S0959269500000983

Wu, M.-J., & Ionin, T. (2019). L1-Mandarin L2-English speakers' acquisition of English quantifier-negation scope. In M. Brown & B. Dailey B. (Eds.), *Proceedings of the 43rd annual Boston University Conference on Language Development* (pp. 716–729). Cascadilla Press.

Wu, M.-J. (2021). Acquisition of Inverse Scope by L1-Mandarin L2-English Learners (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.

L2 acquisition at interfaces

Feature dependency and the poverty of the stimulus in the acquisition of L2 German plural allomorphy

John Archibald
University of Victoria

In this study, I investigated the acquisition of L2 German plural allomorphy via a written production task of classroom learners in North America. Trommer (2015) has argued that the mutual exclusivity of marking plural by either an [n] suffix or an umlauted (i.e. [CORONAL]) stem vowel derives from a universal property of phonological representations. The crossing of association lines, which would result from doubly linking the [CORONAL] feature (which marks the plural), is banned. The data show that the learners made many errors, but that the number of forms which suggested a violation of this universal property were statistically insignificant. I argue that these data are consistent with models showing that interlanguage phonologies are governed by universal phonological principles.

Keywords: L2 phonology, allomorphy, universal grammar, feature geometry, interfaces

1. Introduction

In this chapter, I present the results of the first stage of a project which investigates the acquisition of German plural affixes in second language learners. While there have been studies which have looked at the child acquisition of plurality (Zaretsky et al., 2013), there have been fewer on SLA (Parodi et al., 2004).

The literature on German phonology (e.g., Wiese, 1996) extensively documents the various patterns, and notes the challenges of determining an adequate explanation for the observed alternations. Wiese (1996) noted that the German plural forms are prosodically homogenous in that they all consist of a (σ_{full} . σ_{weak}) pattern, but they are segmentally heterogeneous as illustrated in Table 1.

https://doi.org/10.1075/lald.67.05arc

Table 1. German plural variants

Singular form	Plural form	Affixation
Pelz (fur)	Pelz[ə]	Affixation of -ə
Kind (child)	Kind[ɐ]	Affixation of -ɐ
Held (hero)	Helden	Affixation of -n
Stecken (stick)	Stecken	ø-Affixation

In addition to affixation (e.g., -ə, -ɐ, -n), the plural may also be marked by umlauting (Wunderlich, 1999), as shown in (1).

(1)	a.	Thron	→ Thron[ə] (throne)	
	b.	Sohn	→ Söhn[ə] (son)	'ö' = [ø]
	c.	Mund	→ Münd[ɐ] (mouth)	'ü' = [y]
	d.	Bund	→ Bünd[ə] (federation)	
	e.	Land	→ Länd[ɐ] (land)	'ä' – [ɛ]

In 1(b) and 1(d), we see that the umlauted form can co-occur with the schwa suffix. In 1(c) and 1(e), we note that the umlauted form can co-occur with the dark-schwa suffix [-ɐ]. We can also note a learnability challenge in this tiny data set by comparing the segmentally-similar forms in (a) and (b) and noting that one of them umlauts while the other does not. Compare this with (c) and (d), where the forms are again segmentally-similar but both umlaut while one takes the schwa and the other takes the dark schwa. For the phonologist, it is a challenge to state the generalization (though the dark schwa is often signaled by an orthographic '-r'). For the L2 learner, it is a challenge to learn how the plurals are formed.

In addition to this productive variation, there is an interesting constraint on plural formation, which is that plural *-n* cannot co-occur with umlauting (i.e., vowel fronting). So, the plural form for *godparent* is generated as shown in (2).

(2) Pat[ə] → Pate-n/*Päte-n

Note that the sequence of non-umlauted vowel ([a]) followed by the plural suffix *-n* is well-formed, while the umlauted vowel ([ɛ], orthographically 'ä') followed by plural *-n* is ill-formed. The plural *-n* can also co-occur with a front, rounded vowel in the base form, as shown in (3a). What is banned is the process of umlauting of a vowel as part of plural formation. By contrast, non-plural or lexical *-n* can co-occur with an umlauted vowel, as shown in (3b).

(3)	a.	Börse → Börsen	*stock markets*
	b.	Laden → Läden	*stores*

From an acquisition perspective, these patterns are interesting because they look as if they might involve the acquisition of *multiple exponence*. As defined by Harris (2017, p. 9), "Multiple exponence is the occurrence of multiple realizations of a single morphosemantic feature, bundle of features, or derivational category within a word." This can be shown in the third row of Table 2.

Table 2. Multiple exponence in German plural

Singular	Plural	
Arm	Arme	'arm'
Vater	Väter	'father'
Hals	Hälse	'neck'

The plural form of *neck* is different from the singular in that it has both an affix *and* the umlauted vowel, an example of multiple exponence. The focus of this chapter, though, is on the acquisition implications of the interaction between the *-n* suffix and the umlauting process. We see from the form *Hälse* in Table 2 that multiple exponence occurs in German. However, Harris (2017, p. 9) points out that "an alternation introduced by a phonological rule is not considered an exponent, and hence the alternation cannot involve this as one of the two morphemes in a relation of multiple exponence." Thus, phonologically-conditioned morphological phenomena are *not* instances of multiple exponence. As acquisitionists, the question that interests us is *what are the learners acquiring when they are acquiring the knowledge that umlaut and plural* -n *cannot co-occur?* Ultimately, I argue that what is being acquired is phonologically-conditioned allomorphy, *not* multiple exponence. To prepare to answer this question, we need to probe the Phonology/Morphology interface.

2. Interfaces in SLA

While there has been considerable interest in grammatical interfaces in SLA (Franceschina, 2001; Montrul, 2011; Sorace, 2012; White, 2011), the phonology/morphology interface has been relatively understudied. One major exception of course is the work addressing the Prosodic Transfer Hypothesis (Goad & White, 2009, 2019) focusing on why L2 morphemes are omitted by L2 learners. Archibald (2016) explored a Distributed Morphology (see also Stefanich et al., 2019) analysis of morpheme competition, and of root competition between bilingual lexical items adopting a model of the phonology/morphology interface espoused by Embick (2010).

3. Multiple exponence versus allomorphy

However, this study is guided by the work of Trommer (2015, 2018). Trommer looks at the phonology/morphology interface within the model of Coloured Containment Theory (see Zimmermann, 2017). The details are not particularly relevant for this empirical study, so I will be parsimonious in my description. Distinct morphemes have different morphological *colours*. Colours function as diacritic features where all phonological objects affiliated with a given morpheme wear its colour (i.e., have a feature indicating the originating morpheme). Morphological structure is minimally reflected (by colour only) in phonological representations. That is to say, phonology can tell only whether two phonological objects are part of the same morpheme or not; phonological constraints cannot target specific morphemes (e.g., 3sg). Morphological colour is the only morphological information visible to phonological constraints. The theoretical machinery is designed to handle the spell out of a grammatical feature. Such a relationship of exponence is shown in (4).

(4) $[\mathrm{F}] \Leftrightarrow \phi$

The schema in (4) is read as: the feature [F] has exponent phi; e.g., if the grammatical feature in English is the present participle, then realize this feature as [ɪŋ]. We also need to be able to handle cases of contextual allomorphy (Bonet & Harbour, 2012) as shown in (5).

(5) $$[\mathrm{F}] \begin{cases} \phi_1 \ \mathrm{Context}_1 \\ \phi_2 \ \mathrm{Context}_2 \end{cases}$$

In such instances, the spell out of a single grammatical feature can vary predictably. This is not multiple exponence. According to Trommer (2015, 2018), there is only a single plural affix. The relationship is shown in (6).

(6) $$\begin{array}{ll} [\mathrm{+pl}] \ \mathrm{affix} \Leftrightarrow & \bullet \\ & | \\ & [\mathrm{COR}] \end{array}$$

The plural is spelled out as an underspecified coronal node. Feminine plural is spelled as shown in (7).

(7) [+pl +fem] ⇔ [NASAL].

This then leads directly to the question of why an umlauted vowel and the plural *-n* suffix cannot co-occur. The answer has to do with basic properties of phonological architecture, such as feature dependency and association lines. In what follows, I build the argument that interlanguage phonological grammars respect phonological universals, and that second language learners do not hypothesize impossible grammars. In order to do that, let us explore the nature of the representation of phonological features.

3.1 Feature geometry

Feature geometric representations (Avery & Rice, 1989; Rice, 1994, 1996) encode various structural options sanctioned by natural languages. For example, dependency relations are encoded in (8).

(8)

/t/
Root
|
Place

/p/
Root
|
Place
|
Peripheral

/k/
Root
|
Place
|
Peripheral
|
Dorsal

/n/
Root
SV Place

/l/
Root
SV Place
|
Oral

Consider the representation of /k/. The presence of the [DORSAL] feature implies the presence of [PERIPHERAL]; a representation with [DORSAL] but without [PERIPHERAL] would be ill-formed. This is because certain features are dependent on other higher features. Consider an analogy with syntactic structure. If there is an X^0 then there must be an XP above it; the head is a dependent of the phrase. Similarly, a feature like [strident] is only found under the [CORONAL] node; there are no non-coronal stridents. The presence of the [DORSAL] feature also implies that it can spread to another segment. This is unlike the /t/ which has no dependent of the [PLACE] feature to spread in an assimilation process.[1] See Özçelik & Sprouse (2016) for a thorough discussion of these issues in SLA.

1. The relationship between specification (or redundancy) and phonological activity (in a process like spreading) is a complex one. Itô, Mester, and Padgett (1995) situate these issues within an Optimality Theory context and question the types of formalisms adopted here. However, work such as Dresher (2009), Hall (2017), and Mackenzie (2011) demonstrate how the Successive Division Algorithm (SDA) can solve some of these problems. Mackenzie also demonstrates how the contrastive hierarchies generated by the SDA can be recast within OT.

Of particular relevance to this study is the question of how universal phonological properties influence an interlanguage grammar. Some have argued that IL grammars are 'impaired' (Hawkins & Chan, 1997), 'fundamentally different' (Bley-Vroman, 1990, 2009), or 'shallow' (Clahsen & Felser, 2017), whereas others (Schwartz & Sprouse, 1996) have argued that IL grammars are constrained by Universal Grammar. The L2-plural data set contributes to this discussion as well (see also Archibald & Croteau, 2021). Crucial is the behavior of the [CORONAL] node in marking the German plural.

Remember that Trommer (2015) noted the two spell outs of features given in (9).

(9) [+pl] ↔ [COR] in German.
[+pl +fem] ↔ [NAS]

In terms of feature geometric representations, this would lead to the representations shown in (10).

(10)

Coronal consonant	**Coronal vowel**
CPlace	CPLace
|	|
[COR]	VPlace
|	|
[n]	[COR]
	|
	[i]

Crucially, the [CORONAL] feature can either be expressed as a vowel *or* as a consonant, but not both. The question for second language learners (or perhaps more accurately, for their grammars) is *will [CORONAL] be linked to more than one segment*? If it is, then this would violate a universal phonological principle related to the association lines that connect features to segments. The learnability question is how L2 learners would come to know this, given that we can't hear association lines in the acoustic input. I address these broad theoretical questions by analyzing the data from classroom learners of German, arguing that they do *not* allow multiply-linked [CORONAL] in their grammars.

4. SLA and allomorphy

In this study, I gathered data from 154 learners of L2 German in an introductory university class in North America. The data were collected as part of the regular instruction of the university class. No other information on the subjects was collected. The tests were taken in class and marked by the regular instructional team. I analyzed the sections of the anonymous test papers which asked students about the plural. My main research question was: Will they produce umlaut]$_{\text{Root}}$ + -*n* forms? If so, they might be violating certain phonological universals which will be discussed later in Section 5.3.2.

4.1 Methodology

The data were gathered via two fill-in-the-blanks tests (approximately one month apart) which included sections on German plurals. The tests were administered in class and later marked by the professors and teaching assistants. At time 1, the test question provided the article + noun and asked the student to write in the plural of the noun, as shown in (11).

(11) Der Apfel

At time 2, the test provided an image of the target noun and asked the student to write the plural form of the depicted noun (including the article), as shown in Figure 1.

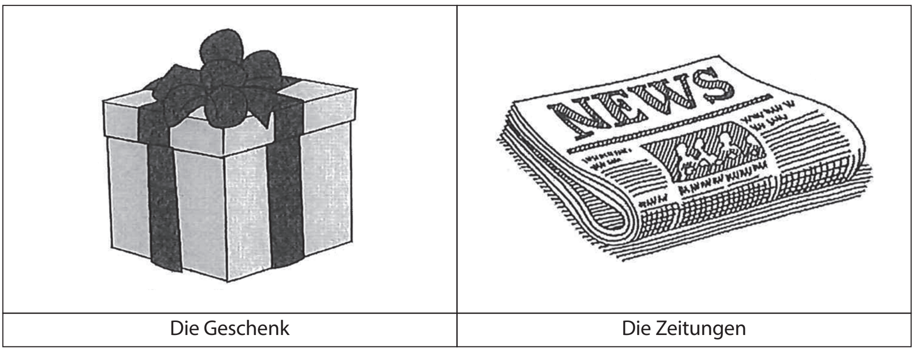


Figure 1. Picture task for plurals

The tests were designed by the German instructional team and, clearly, have different methodologies. While not a strictly controlled experiment, I nonetheless think the data from the two tests reveal the same patterns.

4.1.1 *Test items T1*

The test items at time 1 are given in Table 3. Interestingly (and most likely coincidentally), all but one of the forms is potentially umlautable (i.e., there are back vowels which can be fronted), and it seems to be the one that is not (Ei, 'egg') which had the highest error rate.

Table 3. Test items at time 1

Der Apfel (apple)	Die Äpfel
Die Wurst (sausage)	Die Würste
Die Tomate (tomato)	Die Tomaten
Die Suppe (soup)	Die Suppen
Der Salat (salad)	Die Salate
Das Ei (egg)	Die Eier

4.1.2 *Test Items T2: The test items at time 2 are given in Table 4*

Table 4. Test items at time 2

Die Haltestelle ((bus) stop)	Die Haltestellen
Der Zug (train)	Die Züge
Das Geschenk ((birthday) present)	Die Geschenke
Das Taxi (taxi)	Die Taxis
Der Koffer (suitcase)	Die Koffer
Die Zeitung (newspaper)	Die Zeitungen
Das Flugzeug (airplane)	Die Flugzeuge

Of the twelve lexical items tested, there were 5 feminine, 4 masculine and 3 neuter nouns. One reviewer has raised the point that having more non-feminine nouns than feminine nouns might directly lead to fewer umlaut+-*n* errors if the subjects knew that -*n* was the feminine, plural marker and would, therefore, not add it to non-feminine nouns. This is an interesting point that I will take into account in future data collection; however, in looking at the error patterns for these forms, combining both T1 and T2 data, there were 27% errors made on feminine nouns and 57% errors made on non-feminine (i.e., masculine or neuter) nouns. Thus, it seems that subjects were making errors on non-feminine items.

4.1.3 *Data analysis*

Students' answers were categorized by type of error. I classified the errors in one of two ways. A Type A error consisted of choosing the wrong (but a *possible*) allomorph as shown in (12).

(12) Wursten; Wurste ⇽ Würste

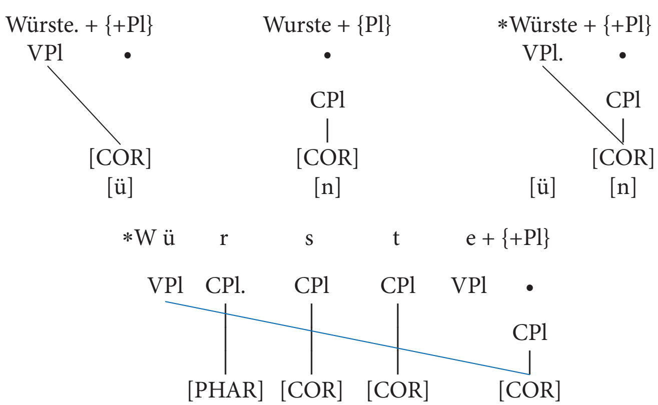
In this case, for the singular *Wurste*, the student would not choose the actual plural *Würste* but would write either the *-n* suffix without umlauting (*Wursten*), or no umlauting and no *-n* suffix (*Wurste*); both forms are phonologically well-formed (i.e. possible German words). A Type B error on the other hand, consisted of choosing the wrong (and an *impossible*) allomorph as shown in (13).

(13) *Würsten ⇽ Würste

In this case, the student produces a form which has *both* the umlauted vowel *and* the *-n* suffix; this is not phonologically well-formed and is an impossible German plural. Interestingly, this is not to say that the string *Würsten* might not occur in the input to the students, as this is the pattern for the dative case for a significant set of nouns[2] (*der Wurst ~ den Würsten*). If learners were merely tracking input frequency, then the existence of these dative plurals might mislead the learners to hypothesize the well-formedness of such forms. But, as we shall see, subjects do *not* produce such forms, demonstrating that the dative plural data does not mislead them as universal phonological properties block the generation of such forms.

Let us explore, then, why some answers are possible (Type A) but other answers are impossible. (Type B). The forms in (12) are well-formed as they follow the conventions of feature dependency, and none of the association lines cross (Goldsmith, 1976; Hammond, 1988). However, the form in (13) would be ill-formed because the doubly-linked [CORONAL] feature would entail the crossing of association lines as shown in (14).

(14)



2. Thanks to a caring reviewer for pointing this out to me.

Let me say more about the motivation of the ill-formedness of the crossing of association lines by discussing an example from tone languages. In tone languages, tones are associated with segments (usually vowels). To take a hypothetical example, a language might allow vowels to have a high tone, a low tone, or a neutral tone. This can be represented as in (15) where H stands for a high tone, L for a low tone, and a vowel with neither H nor L is implemented as a neutral tone.

```
(15)  k a t o b e
        |     |
        H     L
```

In many languages, a neutral tone can be filled in by the spreading of another tone. If the High tone spreads to the neutral tone, then the resulting structure would be as shown in (16).

```
(16)  k a t o b e
        |/    |
        H     L
```

However, if there is an intervening low tone between the high and the neutral tones, as shown in (17) then the high tone cannot spread as it would result in crossed the association lines. Of course, these graphic representations are meant to capture differences in well-formed and ill-formed mental representations.

```
(17)  *k a m a t o b e
         |   |     |
         H   L     L
```

With this grammatical architecture in place, we can articulate the first pass at a hypothesis: *if* the L2 learners' interlanguage phonological grammars respect the principles of coronal licensing and association lines *then* forms with doubly associated coronals should not occur systematically.

4.1.4 *The task*

However, in order for that hypothesis to be coherent, we want to be confident that the task taps into phonological knowledge. My short answer is *yes, it does*.

4.1.5 *Phonology, silent reading, and lexical activation*

There is much literature which demonstrates that silent reading activates phonology. Corcoran (1966, 1967) showed that when asked to cross out the letter *e* in sentences, there were more errors when the *e* was not pronounced. If the task was only visual, this effect would not arise. McCutchen and Perfetti (1982) compared the silent reading of tongue twisters to control sentences and found that tongue twisters take longer to read than matched regular sentences. If sounds were not activated, then this effect would not arise. Fodor (2002) showed that phonology can influence relative-clause (RC) attachment preferences. Short RCs (as in (a)) tend to be interpreted as having low attachment while long RCs (as in (b)) have high attachment as shown in (18 a and b) where the referent of the relative pronoun is bolded:

(18) a. Someone shot the servant of the **actress** [**who** was on the balcony]
b. Someone shot the **servant** of the actress [**who** was on the balcony with her husband]

Bader (1998) showed that it is more costly to the parser to reanalyze longer (i.e., more phonological content) embedded clauses (as in 19(a)) as shown in the following sentences where the initial parse is bracketed:

(19) a. [In order to help the little boy] put down the package he was carrying.
b. [Peter knew the answer] would be false.

Finally, McGuian and Dollins (1989) showed by means of electromyography that muscular activity is triggered during silent reading. This can only be explained if the phonological entries are being activated in silent reading tasks.

4.1.6 *Bilingual lexicon and non-selective access*

There is also a large body of psycholinguistic literature on the architecture of the bilingual lexicon which shows the role of phonology in silent reading. I will briefly summarize some of the work which demonstrates the behaviour of what are called interlingual homographs and interlingual homophones in lexical processing. Dijkstra et al. (1999) used lexical decision tasks to reveal the complexity of what had been known simply as 'cognates' before. This research showed that if we were interested in a question like "are words in both languages activated in a monolingual environment," then we needed to ask a subtler question than that. Not all 'words' were created the same. Table 5 sets up different categories of cross-linguistic relation

Table 5. Interlingual (IL) homographs and homophones

SOP cognates	**SO cognates**	**SP cognates**
hotel	fruit [frøyt]	news/nieuws
film	chaos [xaɔs]	boat/boot
lip	jury [ʒyri]	wheel/wiel
OP false friends	**IL homographs (O)**	**IL homophones (P)**
step (scooter)	glad	[lif]
arts (doctor)	[xlɑt] (slippery)	'leaf' 'lief'
kin (chin)		(dear)

between English and Dutch where the acronyms reveal what is shared across lexical items (*S* = Semantics; *O* = Orthography; *P* = Phonology). Of most interest at this point is the difference between the items which share only Orthography (interlingual *homographs*) and the items which share only Phonology (interlingual *homophones*).

In lexical decision tasks, Dijkstra et al. found that interlingual homographs had faster decision times than control items while interlingual homophones had slower decision times than control items.

Nakayama and Archibald (2005) replicated this finding via a task with more ecological validity than a lexical decision task: eye tracking. In an eye tracking task, participants read sentences on a computer screen. During this task, where and how long the eyes fixate on a sentence can be measured. The central question is: do interlingual homographs and interlingual homophones behave differently in a monolingual, silent reading task? We had *14* Dutch/English bilinguals in Calgary read sentences which had either IL homographs or homophones embedded in English sentences. The fixation times on each were compared to control words, which were then matched by word length, frequency, and predictability. Examples of the two experimental categories (before the solidus with the control item after) are shown in (20).

(20) IL homographs: An **angel/elbow** can be damaged easily.
IL homophones: I had never seen a single **oar/oat** before.

The results of fixation time for the interlingual *homophones* are given in (21).

(21) Mean Control fixation: 239 ms
Mean Experimental fixation: 280 ms

This difference of +41 ms *inhibition* on the experimental items was significant ($p < .05$).

The results of fixation time for the interlingual *homographs* are given in (22).

(22) Mean Control fixation: 284 ms
Mean Experimental fixation: 255 ms

This difference of −29 ms *facilitation* on the experimental items was significant ($p < .05$).

Thus, the interlingual homographs *facilitate* lexical access while the interlingual homophones *inhibit* lexical access. In a monolingual English context, bilinguals cannot suppress access to the other language. Properties of these *Dutch* words explain the *English* reading behaviour. So, lexical activation (including by silent reading) taps into phonology. I thus assume that the classroom tasks reported here activate phonological representations.

5. Results

Let us now look at the responses given by our classroom learners of German, shown in Table 6.

Table 6. Accuracy rates and error patterns

	Time 1	Time 2		Totals	
Subjects	87	67		154	
Test Items	522	469		991	
Correct	292	239		593	
Type A Error	225 (43%)	71 (15%)	159 (33%)	296 (30%)	159 (16%)
Type B Error	5 (.9%)	2 (.4%)		7 (.7%)	

Note. The second cell for Type A errors at T2 and Total indicate that the test item was left blank.

As would be expected from introductory-level students, they made plenty of errors. Remember that the T2 test had only a picture as the stimulus (not the singular word). This different task accounts for the high number of blank responses at T2. I have scored them as errors but have separated them out in the accounting. Note that the Type B (i.e., impossible) errors are very infrequent at seven out of 991 items (.7%). The actual Type B errors are given in (23), where the doubly-assigned [coronal] is in bold.

(23) Die Z**ü**ge**n** (*trains*)
Die Zeit**ü**nge**n** (*newspapers*)
Die T**ö**mate**n** (*tomatoes*)
Die W**ü**rste**n** (*sausages*)
Die S**ü**ppe**n** (*soups*)

The question we must ask is: are these *errors* or *mistakes*? I am drawing here on the traditional applied linguistics terminology (Corder, 1967) in which systematic deviations are referred to as errors (and are assumed to reflect competence) while non-systematic deviations are referred to as mistakes (and are assumed to be performance phenomena). Before looking at the statistical analysis, it is worth noting that none of these forms involved umlauting a front vowel; they seem to know what umlauting is/does.

In order to determine if there was a significant difference in the frequency of Type A versus Type B errors, I ran a Chi-squared test, and the results are given in Table 7.

Table 7. Chi-squared results of Type A versus Type B errors

	Observed N	Expected N	Residual
Type A	296	151.5	144.5
Type B	7	151.5	−144.5
Total	303		
Chi-square	275.64		
Df	1		
Asymp. Sig.	.0001		

The chi-squared results clearly show that the Type B errors occurred significantly less often than the Type A errors. For this reason, it would be more accurate to refer to these deviations from the native target as *mistakes*.

5.1 Errors X word type: T1

In order to explore the data more thoroughly, let us look at the patterns for different allomorphs, as shown in Tables 8 and 9.

Table 8. Accuracy in different plural allomorphs (n = 87)

Umlaut only		Umlaut + e		−n		−e		−er	
Apfel (apple)		Wurst (sausage)		Tomate (tomato) Suppe (soup)		Salat (salad)		Ei (egg)	
Correct	Incorrect	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect	Correct	Incorrrect
48 (55%)	39 (44%)	38 (43%)	49 (56%)	123 (70%)	51 (29%)	60 (68%)	27 (31%)	28 (32%)	59 (67%)

Table 9. Accuracy in different plural allomorphs (n = 67)

Null		Umlaut + e		−n		−e		−en		−s	
Koffer		Zug		Haltstelle		Flugzeug Geschenk		Zeitung		Taxi	
Correct	Incorrect	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect
25	17	31	12	35	2	52	20	30	13	59	5
	(25%)		(18%)		(3%)		(15%)		(19%)		(8%)
	(25)	(14)	(30)		(62)	(34)	(3)				

Note. The scores in parentheses indicate the number of answers left blank.

As my concern is solely between Type A and Type B differences, I do not explore the variation with Type A allomorphic patterns in depth. It is interesting to note, however, that the form which caused the most difficulty involved neither umlauting nor *-n* suffixation.

5.2 Errors X word type: T2

The pattern that we see here, I would argue, is *not* 'they don't produce things they don't hear, and they never hear umlaut+n.

Rather, it would be more accurate to note that 'they don't produce *illegal* structures that they don't hear but they produce *legal* structures that they don't hear.' The legal structures are structures licensed by universal properties of phonological representations. In Table 10, the target form is in the right column while the erroneous forms produced by the subjects (which are never heard in the input) are in the left column. These erroneous forms are, however, well-formed possible plurals.

Table 10. Well-formed (type A) errors which are not in the input

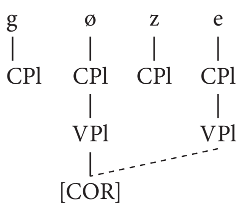
Wursten; Wurste	⇽	Würste
Apfels; Apfelen	⇽	Äpfel
Süppe	⇽	Suppen
Tomate; Tomates	⇽	Tomaten

But what of the ill-formed items (which occur .7% of the time)? In order to delve into why those are not considered in the interlanguage grammars of our subjects, let us turn to the question of what makes an impossible grammar impossible. Ultimately, this is what will allow me to argue that the .7% of the forms are, in essence, noise in the data, i.e., performance errors which do not reflect grammatical competence.

5.3 No impossible Turkish grammars

Özçelik and Sprouse (2016) investigated the acquisition of vowel harmony in L2 Turkish learners. Vowel harmony is an example of what looks like a non-local spreading operation but when viewed within a particular phonological model; the harmony process is governed by tier-based locality. The example in (24) shows how the [CORONAL] feature under the VPlace node spreads to the right so that the two vowels in the word agree in place (i.e., are both front ([CORONAL]) vowels).

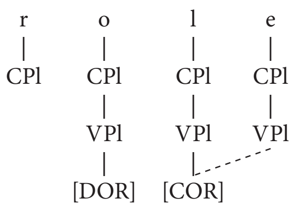
(24)



5.3.1 *Secondary feature spreading*

The example in (25) shows that the [CORONAL] feature can also be found on a CPlace node to indicate secondary articulation. A secondary [CORONAL] feature represents palatalized consonants.

(25)

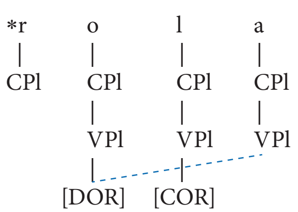


In this form, the word contains both a back vowel and a front vowel. The place feature of the front vowel [e] is the result of the spreading of the [CORONAL] feature from the adjacent palatalized [l].

5.3.2 *No crossing constraint*

There is a general phonological principle that blocks the crossing of association lines (Goldsmith, 1976). Thus, the form given in (26) with two back ([DORSAL]) vowels is ill-formed as it would require the crossing of association lines.

(26)



5.4 No impossible German grammars

I argue that the classroom learners in this study are not hypothesizing 'impossible' German grammars which contain doubly-associated [CORONAL] features with crossing association lines as we saw in (14). Furthermore, I argue that the blocking of the multiple association of [CORONAL] (which would result in the co-occurrence of an umlauted vowel with a plural *-n* suffix) is an instance of the poverty of the stimulus. Let us consider other possible explanations. To rule out the most obvious possibility, the ungrammaticality of umlaut+plural[n] is not taught in class. For the most part, students are told that they need to memorize the plural form of a given singular noun. But the students make mistakes, and the mistakes generate forms that they have not heard from their teacher or read in the textbook. So, what licenses a certain class of mistake but bans another? This is potentially a problem of indirect negative evidence (see also Schwartz & Goad, 2016). If it is not a delimiting of the hypothesis space which blocks the Type B errors, then we would have to ascribe a reasoning process to the learner something along the following lines.

– Sometimes I hear umlaut in plural words (e.g., *Würste*). And the plural word can include an umlaut and a suffix.
  Sohn → S**ö**hn**e** {umlaut + [ə]}
  Mund → M**ü**nd**er** {umlaut + [ɐ]}
– Sometimes I hear *-n* in plural words (e.g. *Suppen*). But wait; I never hear umlaut AND plural *-n* together in the same word; but, wait again, I do hear umlaut + [n] in monomorphemic roots (e.g. *Läden*). So, if I never hear umlaut and plural *-n* together in the same word, it must be because that is ungrammatical.

5.5 Poverty of the stimulus

The argument for the poverty of the stimulus succeeds when we note that the environmental evidence for the learner to hypothesize a ban on the [CORONAL] node to the CPlace suffix and the Vplace root vowel is without a doubt impoverished. And yet, this is the ban that seems to describe their grammar.

Trommer's (2018) version of the No Crossing constraint adopted by Özçelik and Sprouse (2016) includes the two constraints given in (27).

(27) – No Crossing Lines (aka NoSkipCPlace)
  – A Place node dominated by a CPlace node may not associate across another CPlace node
  – No multiple linking
  – [COR] links to EITHER CPlace OR VPlace

The behaviour of the subjects in this study is consistent with this principle. In acquiring this instance of phonologically-conditioned allomorphy (not multiple exponence), their IL grammars respect the rules of the morphology/phonology interface.

6. Conclusion

The data collected from the 154 introductory German students confirms that their interlanguage grammars do not allow the crossing of phonological association lines, thus suggesting that their grammars are constrained by universal phonological principles. It has often been suggested that linguistics is a bit of an outlier in the social sciences or humanities in that data on what is *not* allowed, or *not* observed is paramount, but that is the critical pattern in the data here. As would be expected, students in an introductory foreign language class made errors, 239 of them. With all their errors, though, it is the correct forms (593) and Type A errors (232) that tell the story. The 7/991 (.7%) Type B errors (or more accurately 'mistakes') are more like marginalia – notable in their absence. I have argued that this absence is the result of universal constraints on phonological representations (found also in interlanguage grammars). The learners are not hypothesizing impossible grammars with respect to the behaviour of [CORONAL]. Of course, a corpus is limited in that it does not tap into the well-formedness judgements of the subjects. I hope to be able to administer a judgement task to learners in the future during which they could be given Type B forms like *Würsten* (in both written and spoken forms) to see whether they would reject them as possible plurals. However, even in the absence of such data, this pilot study suggests this to be a fruitful avenue to pursue. In the meantime, the production of these L2 German allomorphs by classroom learners strongly suggests that their interlanguage phonological grammars respect universal representational principles.

Acknowledgments

I would like to thank the GASLA audience in Reno for useful feedback. Particularly, Holger Hopp, and Rex Sprouse. Thanks to Hossein Nassaji for a discussion of data analysis. Finally, I would like to thank the reviewers of this chapter. They say that we should review in the way we would like to be reviewed, and I have to say that the reviews I received of this manuscript were careful, detailed, and supportive – models of the reviewer's craft.

References

Archibald, J. (2016). Phonology at the interface: Late insertion and spell out in L2 morphophonology. Paper presented at News Sounds 2016, Aarhus, Denmark.

Archibald, J., & Croteau, N. (2021). Acquisition of L2 Japanese WH questions: Evidence of phonological contiguity and non-shallow structures. *Second Language Research*, 37(4), 649–679. https://doi.org/10.1177/0267658319897786

Avery, P., & Rice, K. (1989). Segment structure and coronal underspecification. *Phonology*, 6, 179–200. https://doi.org/10.1017/S0952675700001007

Bader, M. (1998). Prosodic influences on reading syntactically ambiguous sentences. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 1–46). Kluwer. https://doi.org/10.1007/978-94-015-9070-9_1

Bley-Vroman, R. (1990). The logical problem of foreign language learning. *Linguistic Analysis*, 20, 3–49.

Bley-Vroman, R. (2009). The evolving context of the Fundamental Difference Hypothesis. *Studies in Second Language Acquisition*, 321, 175–198. https://doi.org/10.1017/S0272263109090275

Bonet, E., & Harbour, D. (2012). Contextual allomorphy. In J. Trommer (Ed.), *Morphology and phonology of exponence* (pp. 195–235). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199573721.003.0007

Clahsen, H., & Felser, C. (2017). Some notes on the shallow structure hypothesis. *Studies in Second Language Acquisition*, 40(3), 693–706. https://doi.org/10.1017/S0272263117000250

Corcoran, D. W. J. (1966). An acoustic factor in letter cancellation. *Nature*, 210, 658. https://doi.org/10.1038/210658a0

Corcoran, D. W. J. (1967). Acoustic factor in proof reading. *Nature*, 214, 851–852. https://doi.org/10.1038/214851a0

Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics*, 5, 160–170. https://doi.org/10.1515/iral.1967.5.1-4.161

Dijkstra, T., Grainger, J., & Van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language*, 41, 496–518. https://doi.org/10.1006/jmla.1999.2654

Dresher, B. E. (2009). *The contrastive hierarchy in phonology*. Cambridge University Press. https://doi.org/10.1017/CBO9780511642005

Embick, D. (2010). *Localism versus Globalism in Morphology and Phonology*. MIT Press.

Fodor, J. D. (2002). Prosodic disambiguation in silent reading. In M. Hirotani (Ed.), *Proceedings of the North East Linguistic Society*, 32 (pp. 112–132). GSLA.

Franceschina, F. (2001). Morphological or syntactic deficits in near – native speakers? An assessment of some current proposals. *Second Language Research*, 17(3), 213–247. https://doi.org/10.1177/026765830101700301

Goad, H., & White, L. (2019). Prosodic effects on L2 grammars. *Linguistic Approaches to Bilingualism*, 9, 769–808. https://doi.org/10.1075/lab.19043.goa

Goad, H., & White, L. (2009). Ultimate attainment in interlanguage grammars: A prosodic approach. *Second Language Research*, 22, 243–268. https://doi.org/10.1191/0267658306sr2680a

Goldsmith, J. (1976). Autosegmental phonology (Unpublished doctoral dissertation). MIT.

Hall, D. C. (2017). Contrastive specification in phonology. In M. Aronoff (Ed.) *Oxford research encyclopedia in linguistics*. Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.26

Hammond, M. (1988). On deriving the well – formedness condition. *Linguistic Inquiry*, 19(2), 319–325.

Harris, A. (2017). *Multiple exponence*. Oxford University Press.

Hawkins, R., & Chan, C. Y-H. (1997) The partial availability of Universal Grammar in second language acquisition: The 'Failed Functional Features Hypothesis.' *Second Language Research*, 13, 187–226. https://doi.org/10.1191/026765897671476153

Itô, J., Mester, A., & Padgett, J. (1995). Licensing and underspecification in Optimality Theory. *Linguistic Inquiry*, 26(4), 571–613.

Mackenzie, S. (2011) Contrast and the evaluation of similarity: Evidence from consonant harmony. *Lingua*, 121, 1401–1423. https://doi.org/10.1016/j.lingua.2011.03.004

McCutchen, D., & Perfetti, C. (1982). The visual tongue – twister effect: Phonological activation in silent reading. *Journal of Verbal Learning and Verbal Behavior*, 21, 672–687. https://doi.org/10.1016/S0022-5371(82)90870-2

McGuian, F.J. and Dollins, A.B. 1989. Patterns of covert speech behavior and phonetic coding. *Pavlovian Journal of Biological Science*, 24, 19-26.

Montrul, S. (2011). Multiple interfaces and incomplete acquisition. *Lingua*, 121, 591–604. https://doi.org/10.1016/j.lingua.2010.05.006

Nakayama, M. & J. Archibald (2005). Eyetracking and interlingual homographs. *Proceedings of the Canadian Linguistic Association*.

Özçelik, Ö., & Sprouse, R. (2016). Emergent knowledge of a universal phonological principle in the L2 acquisition of vowel harmony in Turkish: A 'four' – fold poverty of the stimulus in L2 acquisition. *Second Language Research*, 33(2), 179–206. https://doi.org/10.1177/0267658316679226

Parodi, T., Schwartz, B., & Clahsen, H. (2004). On the L2 acquisition of the morphosyntax of German nominals. *Linguistics*, 42(3), 669–705. https://doi.org/10.1515/ling.2004.022

Rice, K. (1994). Peripheral in Consonants. *Canadian Journal of Linguistics/Revue Canadienne De Linguistique*, 39(3), 191-216. https://doi.org/10.1017/S000841310001522X

Rice K. (1996). *Default variability: The coronal-velar relationship Natural Language and Linguistic Theory*. 14: 493-543. https://doi.org/10.1007/Bf00133597

Schwartz, B. D., & Sprouse, R. (1996). L2 cognitive states and the full transfer/full access model. *Second Language Research*, 12, 40–72. https://doi.org/10.1177/026765839601200103

Schwartz, M., & Goad, H. (2016). Indirect positive evidence in the acquisition of a subset grammar. *Language Acquisition*, 24(3), 234–264. https://doi.org/10.1080/10489223.2016.1187616

Sorace, A. (2012). Pinning down the concept of interface in bilingual development. *Linguistic Approaches to Bilingualism*, 2(2), 209–216. https://doi.org/10.1075/lab.2.2.04sor

Stefanich, S., Cabrelli, J., Hilderman, D., & Archibald, J. (2019). The morphophonology of intraword codeswitching: Representation and processing. *Frontiers in Communication: Language Sciences*, 4, 54. https://doi.org/10.3389/fcomm.2019.00054

Trommer, J. (2015). Moraic affixes and morphological colors in Dinka. *Linguistic Inquiry*, 46(1), 77–112. https://doi.org/10.1162/LING_a_00176

Trommer, J. (2018). The subsegmental structure of German plural allomorphy. Talk given at the University of Victoria.

White, L. (2011). Second language acquisition at the interfaces. *Lingua*, 121, 577–590. https://doi.org/10.1016/j.lingua.2010.05.005

Wiese, R. (1996). *The phonology of German*. Oxford University Press.

Wunderlich, D. (1999). German noun plural reconsidered. *Behavioral and Brain Sciences* 22(6), 1044–1045. https://doi.org/10.1017/S0140525X99542222

Zaretsky, E., Lange, B., Euler, H., & Neumann, K. (2013). Acquisition of German pluralization rules in monolingual and multilingual children. *Studies in Second Language Learning and Teaching*, 3(4), 551–580. https://doi.org/10.14746/ssllt.2013.3.4.6

Zimmermann, E. (2017). *Morphological length and prosodically defective morphemes*. Oxford University Press. https://doi.org/10.1093/oso/9780198747321.001.0001

Development of L2 prosody

The case of information focus

Atsushi Fujimori[1], Noriko Yamane[2], Noriko Yoshimura[1], Mineharu Nakayama[3], Brian Teaman[4] and Kiyoko Yoneyama[5]
[1]University of Shizuoka / [2]Hiroshima University / [3]The Ohio State University / [4]Osaka Jogakuin University / [5]Daito Bunka University

This study examines the development of L2 English prosody as associated with information focus by Japanese EFL (JEFL) learners. Comprehension, perception, and production tasks were conducted with 54 participants in three sub-groups: low proficiency level JEFLs, high proficiency level JEFLs, and native English speakers. Results show that the low-level JEFLs could identify which parts of sentences require focus but they could not successfully perceive nor produce three different prosodic patterns associated with information focus. In contrast, high proficiency JEFLs demonstrated native-like performance, with a slight difficulty in producing post-focal compression. Our findings have implications for the syntax-discourse-prosody interface: (i) the interface knowledge develops through multiple stages, and (ii) the discourse-prosody interface is challenging for L2 learners.

Keywords: prosody, information focus, comprehension, perception, production, Japanese EFL learners

1. Introduction

Many early second language (L2) acquisition studies within the Principles and Parameters approach dealt with accessibility to Universal Grammar. They examined cases where L2 differed from first language (L1) by examining whether L1 transfer occurred at the initial stage of L2 (Schwartz & Sprouse, 1994) and whether parameter-resetting took place in the course of L2 acquisition (White, 1989, 2003). With recent developments in the Minimalist theory of grammar (Chomsky, 1993, 1995), research concerns have shifted to how individual linguistic modules interact with each other in L2 acquisition. These studies have attributed L2 difficulties to the

https://doi.org/10.1075/lald.67.06fuj

interface between modules such as syntax/semantics, syntax/discourse, and syntax/morphophonology (Belletti, Bennati, & Sorace, 2007; Nakayama & Yoshimura, 2015; Slabakova, 2009; Sorace, 2004, 2011; Tsimpli & Sorace, 2006; White, 2011). For instance, it has been proposed that knowledge of referential pronouns at the interface between syntax and discourse is more vulnerable than that knowledge of constructions involving syntactic derivations alone (Sorace & Filiaci, 2006).

However, how syntax-discourse phenomena are associated with phonology has not been researched thoroughly in L2 acquisition. Phonological constituents match up with their corresponding syntactic constituents: prosodic phrases to syntactic phrases and intonational phrases to syntactic clauses (Selkirk, 2009). Intonational phrases in particular are affected by discourse-related focal pitch rises by dephrasing the following pitch range compressions (Venditti, Maekawa, & Beckman, 2008). Therefore, it is expected that their prosodic properties are associated with discourse and pragmatic properties. This study explores the interface between prosody, discourse, and syntax in L2 acquisition. To investigate how these modules interact with each other, we examine English prosodic marking of information focus by Japanese speakers learning English as a foreign language (EFL).

This paper is organized as follows: Section 2 outlines the linguistic background of information focus marking in English and Japanese. Section 3 explains the experimental procedure and presents the results of the experiments. Our discussion is provided in Section 4 and concluding remarks are found in Section 5.

2. Realizing information focus in English and Japanese

2.1 Information focus marking in English

Focus is a linguistic phenomenon associated with Information Structure. There are, minimally, two types of foci: information focus and contrastive focus (Jackendoff, 1972; Rochemont, 1986; Zimmermann & Onea, 2011; Zubizarreta, 1998; among many others).[1] Of particular interest here is information focus. Information focus identifies the information that is most salient and important in a sentence, i.e., the non-presupposed part of the sentence in discourse, typically observed in responses to *wh*-questions. Foci can vary in domain and place. Taking an English transitive structure, for instance, the domain of focus in (1) is narrowed down to a part of the sentence, namely, the subject determiner phrase (DP) appearing at the beginning

1. A contrastively focused expression facilitates the update of the common ground, under the assumption that the focus content is unexpected for the hearer from the speaker's viewpoint. (Zimmermann, 2008; Zimmermann & Onea, 2011).

of the sentence (early narrow focus, ENF). In (2), the domain of focus is the object DP at the end of the sentence (late narrow focus, LNF). In (3), the domain of focus is the entire clause (broad focus, BRF). We follow the accepted practice of marking focused constituents with $[\ ]_F$ (Jackendoff, 1972; Selkirk, 1984).

(1) Q: Who met Mary?
A: $[$ Bob $]_F$ met Mary. (early narrow focus, ENF)

(2) Q: Whom did Bob meet?
A: Bob met $[$ Mary $]_F$. (late narrow focus, LNF)

(3) Q: What happened?
A: $[$ Bob met Mary $]_F$. (broad focus, BRF)

While focused elements must be maximally prominent (Truckenbrodt, 1995), how foci are realized varies from language to language; they can be realized by pitch accent, prosodic phrasing, constituent ordering, or a mix of all of these (Büring, 2012). English and Japanese information foci are phonetically realized by prosodic prominence including pitch and its shape, which can be derived from language-specific parameters such as prosodic headedness built upon prosodic phrasing. Match Theory (Selkirk, 2009) requires syntactic phrases to correspond to phonological counterparts; syntactic phrases to phonological phrases (pP) and syntactic clauses to intonational phrases (IP). In English, pPs are right-headed, conforming to a rule traditionally known as the Nuclear Stress Rule (Chomsky & Halle, 1968; Liberman & Prince, 1977). In the transitive mono-clause in (4), the subject DP and VP constitute a different pP, which contains its head on the right. The pPs join together to constitute an IP where the rightmost prosodic word becomes its head. This is by default observed in broad focus (BRF), as an answer to an out-of-the blue question (e.g., "what happened").

(4) $(\quad\quad\quad\quad *)_{IP}$ (BRF)
$(*) (\quad\quad\quad *)_{pP}$
$[$ Bob met Mary $]_F$

In the early narrow focus (ENF) in (5), the subject DP, but not the other elements, is focused and becomes the head of IP.

(5) $(* \quad\quad\quad\quad)_{IP}$ (ENF)
$(*) (\quad\quad\quad *)_{pP}$
$[$ Bob $]_F$ met Mary.

In the late narrow focus (LNF) in (6), the object DP, but not the other elements, is focused. Hence, the rightmost focused element becomes the head of the IP.

(6) (*)$_{IP}$ (LNF)
(*) (*)$_{pP}$
Bob met [Mary]$_F$

It turns out that BRF and LNF are identical in the headedness of IP, and one might wonder whether the two types of foci are phonetically the same or how they are differentiated. According to Jannedy (2002), they differ in phonetic realization. In BRF, the sentence yields two high pitch targets, namely, the subject and object nouns, with a slight interpolation between them to create a so-called "hat pattern." In LNF, in contrast, the stressed syllable of the object noun receives a nuclear prominence (i.e., being perceived as the most prominent) with an immediate steep pitch fall to the immediately following syllable, i.e., post-focal compression (PFC). In ENF, the focused subject noun is assigned the prosodic prominence with a PFC, and the remainder stays low (see also Beckman (1986) for the English phonetic trait of pitch declination).

2.2 Information focus marking in Japanese

Information Focus marking in Japanese and English are similar in that the focused word is marked by pitch raising. However, there are some phonetic and phonological differences. pPs, a lower constituent on the prosodic hierarchy, are left-headed in Japanese (Büring, 2009; Selkirk & Tateishi, 1991; Truckenbrodt, 1995), which is a mirror image of the English and reflects the universal tendency of syntax-phonology interaction that the complement attracts prosodic prominence (Nespor, Shukla, van de Vijver, Avesani, Schraudolf, & Donati, 2008). In contrast, Japanese IPs, which are hierarchically above pPs, are right-headed, like in English (Truckenbrodt, 1995). As shown in BRF (7), the pitch accent * on the object 'Mary' in the rightmost pP is projected as the head of the IP by default. In the case of ENF (8), the focused subject noun 'Bob' becomes the head of the IP.[2] In LNF (9), the focused object noun becomes the head of the IP. Thus, BRF and LNF share the same structure in the prosodic hierarchy (i.e., at the phonological level rather than phonetic level). It turns out that Japanese, like English, has a two-way distinction at this level, regarding the three types of information focus.

2. Another way to state it would be that the focused subject noun overrides the right-headed IP assuming that recursive IPs correspond to CPs and above, along the lines of Ito and Mester (2012) (see also Sugahara, 2003, for this overriding rule). The same applies to English ENF cases.

(7) (*)$_{IP}$ (BRF)
(*)(*)$_{pP}$
[Bob-ga Mary-o mita]$_F$
Bob-NOM Mary-ACC saw
'Bob saw Mary.'

(8) (*)$_{IP}$ (ENF)
(*)(*)$_{pP}$
[Bob-ga]$_F$ Mary-o mita

(9) (*)$_{IP}$ (LNF)
(*)(*)$_{pP}$
Bob-ga [Mary-o]$_F$ mita

While Japanese is similar to English in the headedness of IPs across all types of BRF, ENF, and LNF, the phonetic realization of Japanese information focus is different from its English counterpart (see Table 1). The global phonetic pitch shape of BRF (7) of Japanese is not a "hat pattern", but a "zigzag pattern," with multiple high targets due to fixed lexical pitch accents and the subsequent downstep (Kubozono, 1993). The focused word is also marked with pre-focal rising (PFR), triggering pitch reset appearing as F0 rise at the phrase-initial position (Ishihara, 2011; Pierrehumbert & Beckman, 1988). Thus, LNF differs from BRF in that there is a larger pitch increase on the object.

Table 1. Information focus marking in English and Japanese

	English	**Japanese**
Phonetic realization of nuclear prominence	Post-focal compression (PFC) F	Pre-focal rising (PFR) F
Intonational Phrase (IP)	Right-headed	Right-headed
Phonological Phrase (pP)	Right-headed	Left-headed

2.3 Research questions

Many studies on the L2 acquisition of English prosody by Japanese EFL learners are concerned with prosodic properties per se, paying attention to those properties found in a single sentence, but not in discourse (Hosaka, 1998; Mori, Hori, & Erickson, 2014; among others). A few studies have investigated the acquisition of L2 English focus marking by Japanese EFL learners, however. In Fujimori, Yamane, and

Yoshimura (2017), novice Japanese learners of English participated in a read-aloud task. In producing question-answer pairs, the learners typically misplaced the highest pitch on the sentence-initial word (like a subject noun), even though the information focus was a verb or object. While their production scores were significantly lower than native speakers, their comprehension and perception scores were comparable to native speakers of English.[3] Their study suggests that novice learners already had at least some knowledge of the syntax-discourse interface. However, the question remains as to whether the knowledge of discourse-prosody is acquirable in the course of L2 development.

Takeda (2018) tested advanced and novice Japanese EFL learners with regard to English focus marking. In a perceptual rating task, the advanced and novice learners could correctly identify contrastive focus, but not information focus, on a prenominal adjective. In a semi-spontaneous production task, both groups were asked to read aloud a target sentence with the appropriate prosody for a particular context. Results showed learners could not associate the appropriate prosody with information focus in test tokens similar to those used in the perceptual rating task. Takeda concluded that the Japanese EFL learners had difficulty perceiving and producing appropriate prosodic patterns of information focus in L2 English.[4] While she paid attention to the particular focus placement, namely, prenominal adjectives, to compare the two types of foci, focus placement varies from context to context in natural conversation. Thus, it is important to examine how Japanese EFL learners perform in information focus marking on a variety of focus locations.

To explore how L2 knowledge of information focus develops at the interfaces between syntax, discourse, and prosody, we designed our study to answer the following research questions: (i) how do Japanese EFL learners of low- and high-proficiency levels perform in comprehension, perception and production tasks involving information focus marking?, and (ii) how do Japanese EFL learners mark three types of information focus in English (namely, BRF, ENF and LNF)? Given that focus placement is flexible and context-dependent in both English and Japanese, we expect that positive L1 transfer will occur. That is, Japanese EFL learners will not only comprehend the semantically and pragmatically salient parts

3. The test tokens in Fujimori et al.'s (2017) perception task were uncontrolled in focus placement.

4. Takeda's (2018) results of the low proficiency level learners might also be attributed to task difficulties. Rating tasks are challenging particularly for L2 novice learners (Kunevich & Alexopoulou, 2014). In her production task, more than a word had to be emphasized for information focus whereas a single word for contrastive focus; Information focus placed a greater cognitive load on production than contrastive focus.

of speech, but will also produce and perceive the nuclear prominence flexibly on the appropriate place, regardless of focus placement, in L2 English. Unsuccessful prominence production and perception would indicate that this discourse-prosody interface phenomenon is challenging for L2 learners.

3. Experiments

In this experimental study, 36 adult Japanese learners of English participated in three tasks involving the comprehension, perception, and production of information focus marking. To determine the existence of developmental stages, learners were divided into two experimental groups based on their scores in Test of English for International Communication (TOEIC): 18 of high proficiency level (average TOEIC 787.5, CEFR B2 level) and 18 of low proficiency level (average TOEIC 552.8, CEFR A2). Eighteen native speakers (NSs) of English also participated in the study as the control group. The same groups of participants completed the three tasks and a background questionnaire during their one-time visit to a lab in the order shown in (10), so that each participant's production was not affected by the model provided in the perception task. After each task was completed, documents were collected so they could not refer to their own previous answers in a subsequent task.

(10) Order in which tasks were administered
1. Comprehension task
2. Background questionnaire
3. Production task
4. Irrelevant production task
5. Perception task

3.1 Comprehension task

The Comprehension task examined whether participants could understand which part(s) of the sentence constituted information focus in a question-and-answer dialog preceded by a context. In the task, participants were given nine question-answer pairs (3 focus types x 3 tokens) in written form. They were then asked to mark semantically salient words (shown as capitalized in (11) but in normal mixed case on their mark sheet) in a written test sentence to elicit an appropriate response to the written question. Each test sentence contains a transitive construction with an adjunct prepositional phrase (PP). Note that (11) contains three distinct questions for an expository purpose although each participant encountered one of the three questions in the actual test token (see Appendix for test tokens). In the task, necessary and sufficient markings of semantically salient words were counted as correct (for example in 11a we expected "Ben" to be underlined or circled for ENF). Any other markings than those predicted were counted as incorrect answers in the analysis.

(11) Ben is one of Taro's friends. He is an interesting guy. He wore boots to Taro's wedding.
a. Q: Who wore boots to the wedding?
A: BEN wore boots to the wedding. (ENF)
b. Q: What did Ben wear to the wedding?
A: Ben wore BOOTS to the wedding. (LNF)
c. Q: What happened?
A: BEN wore BOOTS to the WEDding. (BRF)

We turn to the results of the Comprehension task for each focus type. As Table 2 shows, not only the NS group but also the high- and low-proficiency groups performed well and correctly marked focused words in a response to the written question; 77.8% of the NS group, 80% of the high-proficiency group and 77.8% of the low-proficiency group correctly chose the subject (S) as the only semantically salient element in ENF type sentences.[5] The other answers such as the subject and object (SO), the subject, verb and object (SVO), and the subject, object and prepositional phrase (SOPP) were considered incorrect.

Table 2. Answer types and frequencies for ENF phrases (%)

Answers	NS	HIGH	LOW
S	77.8	80.0	77.8
SO	13.0	9.3	1.7
SVO	7.4	1.9	1.9
SOPP	1.9	1.9	0
Others	0	6.9	18.6

All three groups also correctly chose the object (O) alone as new information in LNF type sentences at the rate of 72.2%, as in Table 3.

5. Of the remaining responses for ENF in the NS group, many chose not only the subject but also other content words such as the object (SO 13.0%) and the verb (SVO 7.4%). The participants may have confused such content words with semantically salient words of a sentence in response to the *wh*-question.

Table 3. Answer types and frequencies for LNF phrases (%)

Answers	NS	HIGH	LOW
O	72.2	72.2	72.2
SO	7.4	11.1	7.4
VO	7.4	3.7	7.4
OPP	1.9	7.4	5.6
Others	11.1	7.3	7.4

Table 4 shows the results of BRF type sentences; the NS group's answers showed a variety of choices while there were fewer cases where only the subject (S) was chosen (0%), only the verb (V) (9.3%) or only the object (O) (13.0%). Thus, all answer choices which included multiple elements were considered to be accurate. This constituted 77.7% of answers for the NS group. The answer choices by the learner groups were also widely distributed with those including multiple elements were 81.4% for the high-proficiency group and 66.7% for the low-proficiency group.

Table 4. Answer types and frequencies for BRF phrases (%)

Answers	NS	HIGH	LOW
VO	25.9	11.1	18.5
SO	20.4	18.5	20.4
O	13.0	9.3	18.5
SVO	9.3	7.4	18.5
V	9.3	7.4	11.1
VOPP	7.4	0	3.7
SOPP	5.6	18.5	3.7
OPP	5.6	3.7	5.5
SVOPP	3.7	18.5	0
SV	0	1.9	1.9
VPP	0	1.9	7.4
PP	0	1.9	3.7
SPP	0	0	3.7

Table 5 summarizes the accuracy response rates of the three tokens for the three focus types. An Analysis of Variance (ANOVA) was conducted with correct (1)/incorrect (0) responses as the dependent variable, and groups and focus types as the fixed factors. There were no main effects for group (F (2, 477) = 1.011, p = .365) and focus type (F (2, 477) = 0.844, p = .430), and there was no interaction (F (4, 477) = 0.444, p = .776). That is, there were no significant differences among the groups and the focus types.

Table 5. Correct response rates for comprehension

Group	Focus type		
	ENF	LNF	BRF
NS	.72	.78	.78
HIGH	.74	.81	.81
LOW	.72	.78	.67

3.2 Perception task

In the Perception task, participants heard a *wh*-question through a headset and selected the best choice from a set of three potential recorded answers that were structurally the same but with different pitch patterns of either ENF, LNF, or BRF. An example answer is found in (12), where the prominent syllable is indicated in upper case. The test tokens that the participants heard were the same as those used in the Comprehension task. Crucial to the task is that the clause contains an adjunct prepositional phrase so that BRF and LNF can clearly be perceived in post-object compression; a PFC is expected immediately after the focused object DP in LNF (12A). An object PFC is not expected in BRF (12B) as pitch remains high until the end of the entire clause, i.e., the focus domain.

(12) Context: Mary and Jennifer went shopping at a department store last weekend. Mary bought boots for her mother's birthday.
Question: Who bought boots for her mother?
A. Mary bought [BOOTS]$_{\text{LNF}}$ [for her mother]$_{\text{adjunct}}$.
B. [MAry]$_{\text{ENF}}$ bought boots [for her mother]$_{\text{adjunct}}$.
C. [MAry bought BOOTS [for her MOther]$_{\text{adjunct}}$]$_{\text{BRF}}$.
D. I don't know.

Although only the context was displayed on a monitor, participants heard the question with the three answer choices. After they heard each set twice, they made their choice by circling either A, B, C, or "I don't know." There was only one "I don't know" response and it was excluded from the analyses. The task was administered via PowerPoint and the set of each test token was automatically played on a laptop after a 10 second interval.

Test tokens were recorded by an adult male native speaker of North American English in a sound-proofed room. He deliberately read aloud the test tokens slightly slower than a normal speech rate in order to generate clear pitch contours. The experimenters confirmed that each token was uttered with the appropriate prosody, particularly in terms of pitch and duration.

In the Perception task, the NS- and high-proficiency groups performed well across the board; their accuracy rate for each focus type was above 80%, as shown in Table 6. The correct response rates of the low-proficiency group were at around 60%, lower than those of the NS- and high-proficiency groups. An ANOVA was conducted with correct (1)/incorrect (0) responses as the dependent variable, and groups and focus types as the fixed factors. There was a main effect of group (F (2, 477) = 15.936, p < .001) but no main effect of focus type (F (2, 477) = 2.238, p = .108) nor interaction (F (4, 477) = 0.872, p = .481). Within each focus type of ENF and LNF, the low-proficiency group's correct response rates were significantly lower than the NS's as shown by a post-hoc Bonferroni (all p<.001).

Table 6. Accuracy rates for choosing the correct audio response

Group	**Focus type**		
	ENF	**LNF**	**BRF**
NS	.91	.87	.80
HIGH	.89	.87	.81
LOW	.72	.56	.65

Individual data of the low-proficiency group showed biases to particular focus types; four participants chose a BRF type answer for LNF in more than two out of three tokens. Two participants chose a BRF type answer for ENF. Three participants chose an LNF type answer for BRF. These data indicated that some participants in the low-proficiency group had difficulty differentiating among the three focus types in a perception task.

3.3 Production task

In the Production task, participants read question-answer sentences aloud in a natural way. They were given several minutes to practice in pairs before recording, to ensure that they understood the conversation, and that they became familiar with the task. The goal was to help participants read the texts fluently without stuttering, unnatural pausing and fillers, which should have positive effects consistently across all of the three focus types of tokens. They were given no instructions regarding pronunciation, rhythm, or intonation during the course of the Production task. Each participant was individually recorded in a sound-proofed room.

Sentences included transitive verbs with sentence-final prepositional phrases, the same as those each participant saw in the Comprehension and the Perception tasks, such that pitch compression was measured for not only the subject but also for the object. All the tasks included nine question-answer pairs, each of which was preceded by a context. Three tokens were provided for each focus type of ENF, LNF and BRF making a total of 9 (3 focus types x 3 tokens) utterance pairs. Tokens were presented in a Latin square such that participants did not see the three conditions of the same item (lexicalization).

Each test token was controlled in the following ways. The subject and object nouns consisted of vowels, sonorants, and voiced obstruents. This is because vowel qualities are affected by onset voiceless obstruents (e.g., fricative /s/), often making it impossible to measure vowel pitches for the acoustic analyses of the test stimuli.

All recorded sentences were digitized at a sampling rate of 16 KHz and normalized in semitones (st) with Praat software in order to assure a gender-neutral analysis. After segmenting the test tokens for nine participants in each group, post-focal compressions, i.e., the f0 (fundamental frequency) height of a focused word minus that of its post-focal word, was measured at the mid-point of each vowel (Cooper, Eady, & Mueller, 1985). As a reviewer noted, other prosodic properties such as duration, intensity, peak alignment, and format transition are also important cues to distinguish different focus types in English (Sityaev & House, 2003; Xu & Xu, 2005; among many others). However, this study paid attention to pitch height for the following reasons. First, Jannedy (2002) points out that pitch height itself distinguishes between the three focus types of ENF, LNF and BRF in English. Second, Japanese behaves in a way similar to English in that pitch accent plays a major role in information focus marking.

In the post-subject compression (from the subject to the verb), the performance of the NSs showed a larger declining pitch movement with ENF than with LNF or BRF, as in Figure 1. A similar tendency was found in the high-proficiency group, as in Figure 2, but not in the low-proficiency group, shown in Figure 3. An ANOVA was conducted with post-subject compressions in semitone as the dependent variable, and groups and focus types as the fixed factors. There was a main effect of focus type (F (2, 207) = 12.055, $p < .001$) and a marginal main effect of group (Group F (2, 207) = 2.337, $p = .099$) but no interaction (F (4, 207) = 1.864, $p = .118$). Within each group of the NS- and high-proficiency groups, there were significant differences between ENF vs LNF and BRF, as revealed by a post-hoc Bonferroni.

As for the post-object compression, there were main effects of focus type (F (2, 192) = 14.812, $p < .001$) and group (F (2, 192) = 4.960, $p = .008$) but no interaction (F (4, 192) = 1.320, $p = .264$). Post-hoc Bonferroni tests indicated that there were significant differences among the three focus types for the NS group. That is, the NS group demonstrated a three-way distinction between LNF, BRF, and ENF, with the largest pitch movement from the object to the preposition in LNF. Meanwhile, there was a two-way distinction between LNF and BRF vs ENF for the high-proficiency group. The low-proficiency group did not differentiate among the three focus types.

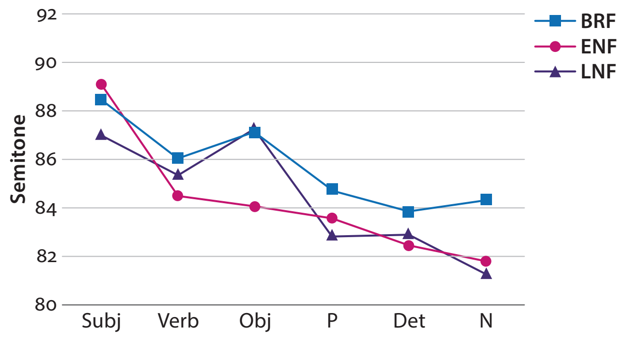


Figure 1. Average pitch patterns for NSs

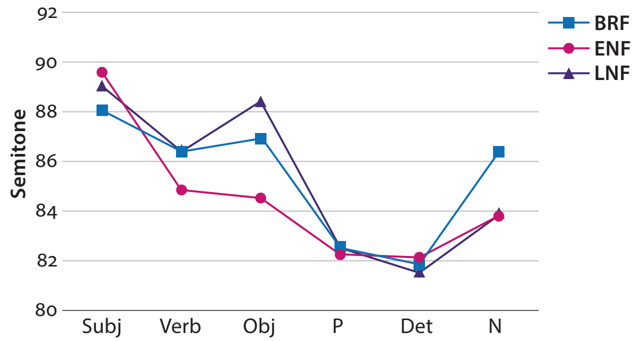


Figure 2. Average pitch patterns for high-proficiency learners

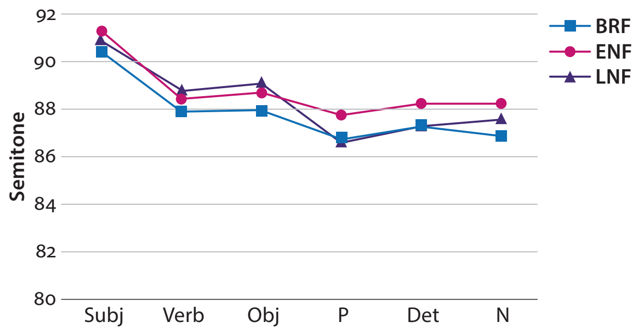


Figure 3. Average pitch patterns for low-proficiency learners

4. Discussion

The experimental results of this study showed that the low-proficiency L2 learners could determine which parts of a sentence were the most salient constituents in discourse, but they could not perceive the different prosodic patterns of three focus types ENF, LNF, and BRF, as summarized in Table 7. They also produced flat pitch patterns with no difference among the focus types.

Table 7. Summary of the results of the three experimental tasks

Group	Tasks		
	Comprehension	Perception	Production
LOW	✓	×	×
HIGH	✓	✓	△
NS	✓	✓	✓

High-proficiency L2 learners performed well not only in comprehension but also in perception. Particularly, they showed to be sensitive to post-focal compression. These facts indicate that each linguistic component of prosody and discourse develops independently in L2. In particular, the knowledge of question-and-answer congruence at the syntax-discourse interface is acquired prior to the discourse-prosody knowledge in the development of L2.

Results also showed that the high-proficiency L2 learners overall demonstrated a native-like performance in production, as summarized in Table 8; the learners' post-subject compressions were quite similar to those of the NS group with steep declination in ENF, but not LNF nor BRF.

Table 8. Post-focal compressions by the NS and the HIGH groups

Post-focal compression	NS	HIGH
Post-subject	ENF > LNF, BRF	ENF > LNF, BRF
Post-object	LNF > BRF > ENF	LNF, BRF > ENF

Unlike the NS group, however, learners could not differentiate between LNF and BRF in post-object compression. This difficulty lies partially at the syntax-discourse-prosody interface. The NS group planned their speech, targeting the right-edge of focus domain in BRF and resulting in a gradual pitch declination or interpolation between the object and the PP-noun, in contrast to steep declination in LNF. High-proficiency learners prioritized the syntax-prosody knowledge, stressing the focused content words and destressing the function words. As a result, they produced a steep declination after the object targeting a preposition in BRF, like in LNF. In other words, even the high proficiency learners at their stage were not as sensitive to the discourse-prosody knowledge as the native speakers of English. Then the question arises as to their ultimate attainment; we need to examine whether near-native learners of English can ultimately demonstrate the appropriate discourse-prosody knowledge in production.

Another possible factor that could account for the difficulty of differentiating LNF from BRF is that Japanese is sensitive to pre-focus pitch rising, as opposed to post-focal compression. Hence, LNF and BRF are differentiated by pitch height of the object while the pitch-target of Japanese post-focal compression lies at around 100Hz regardless of the focus types.[6] Hence, L2 English prosody might be affected by phonetic L1 transfer in post-focal compression. We note that this problem was found only in production, but not in perception. The asymmetry might support the claim that perception precedes production in L2 (Flege, Bohn, & Jang, 1997). The claim should be tested with intermediate level learners.

The finding of the developing L2 English prosody of high-proficiency L2 learners poses a pedagogical question of whether there is any effective way of instructing L2 English prosody. The high-proficiency learners continued encountering some difficulties in producing English prosody while prosody as it is associated with not only syntax but also discourse plays a crucial role in communication. A clue can be found in our data. The pitch range was expanded with pitch lowering on non-focused words as the learners' proficiency improved (see Figures 2 and 3). Also, much attention needs to be paid to the pitch target of post-focal compression and pitch declination in general. It might be effective to instruct L2 learners to concentrate on the falls rather than the rises in producing accurate English prosody. This should be studied further in future research.

5. Conclusion

The present study has shown evidence of Japanese EFL learners' developmental stages of comprehending, perceiving, and producing the L2 English information focus. The finding of developmental stages at the interfaces in the linguistic components has a further implication for L2 acquisition study. Prosody components are acquired late because it is a syntax-external or clause-external component and is placed onto meaningful linguistic units associated with discourse (see also Goss

6. This fact is not clearly stated in Ishihara (2011) although he shows pitch patterns of BRF and LNF in Japanese.

& Nakayama, 2011). We need to investigate further whether near-native learners of English can ultimately demonstrate this discourse-prosody knowledge in production.

Acknowledgements

We thank two anonymous reviewers, the book editors and the series editor, Lydia White, for their insightful comments which helped refine the paper. Finally, we are grateful to the participants who provided the data for this study.

Funding

This study has been partially supported by the Japan Society for the Promotion of Science Grant-in-Aid for Scientific Research (C) 17K02969 to the first author. Their support is gratefully acknowledged.

Appendix. Nine test tokens used in the experiment

1. Ben is one of Taro's friends. He is an interesting guy. He wore boots to Taro's wedding.
   Q: Who wore boots to the wedding? (triggering ENF)
   Q: What did Ben wear to the wedding? (triggering LNF)
   Q: What happened? (triggering BRF)
   A: Ben wore boots to the wedding.
2. Alex and Ken jogged to a park yesterday. Alex bought a bottle of water and drank it on his way home.
   Q: Who drank water on his way home?
   Q: What did Alex drink on his way home?
   Q: What happened?
   A: Alex drank water on his way home.
3. Mary and Jennifer went shopping at a department store last weekend. Mary bought boots for her mother's birthday.
   Q: Who bought boots for her mother?
   Q: What did Mary buy for her mother?
   Q: What happened?
   A: Mary bought boots for her mother.
4. Emma and Naomi liked their classmate Tom very much. One day, Naomi wrote a love letter to him.
   Q: Who wrote a love letter to Tom?
   Q: What did Naomi write to Tom?
   Q: What happened?
   A: Naomi wrote a love letter to him.

5. Mike went to a bar with Bob last Friday. Mike drank wine while Bob drank a beer.
   Q: Who drank wine at the bar?
   Q: What did Mike drink at the bar?
   Q: What happened?
   A: Mike drank wine at the bar.
6. Alice's daughter Nancy got a fever. So Alice brought ice and placed it on her forehead.
   Q: Who brought ice to her daughter?
   Q: What did Alice bring to her daughter?
   Q: What happened?
   A: Alice brought ice to her daughter.
7. Eric invited friends to his house. He made lemonade for them.
   Q: Who made lemonade for his friends?
   Q: What did Eric make for his friends?
   Q: What happened?
   A: Eric made lemonade for his friends.
8. There are two trains from Ted's town to Boston every day. He planned to take the 7:00 a.m. train. But he overslept and missed the train in the morning. So he had no choice but had to take the 1:00 p.m. train.
   Q: Who missed the train in the morning?
   Q: What did Ted miss in the morning?
   Q: What happened?
   A: Ted missed the train in the morning.
9. Mary called Ben about her party. Ben said he was willing to join Mary's party.
   Q: Who invited Ben to her party?
   Q: Who did Mary invite to her party?
   Q: What happened?
   A: Mary invited Ben to her party.

References

Beckman, M. (1986). *Stress and non-stress accent*. Foris. https://doi.org/10.1515/9783110874020

Belletti, A., Bennati, E., & Sorace, A. (2007). Theoretical and developmental issues in the syntax of subjects: Evidence from near-native Italian. *Natural Language and Linguistic Theory*, 25, 657–689. https://doi.org/10.1007/s11049-007-9026-9

Büring, D. (2009). Towards a typology of focus realization. In M. Zimmermann & C. Féry (Eds.), *Information structure: Theoretical, typological and experimental perspectives* (pp. 177–205). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199570959.003.0008

Büring, D. (2012). Focus and intonation. In G. Russell & D. G. Fara (Eds.), *The Routledge companion to the philosophy of language* (pp. 103–115). Routledge.

Chomsky, N. (1993). A minimalist program for linguistic theory. In K. Hale & S. J. Keyser (Eds.), *The view from building 20. Essays in linguistics in honor of Sylvain Bromberger* (pp. 1–52). The MIT Press. https://doi.org/10.7551/mitpress/9780262527347.001.0001

Chomsky, N. (1995). *The minimalist program*. The MIT Press.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Harper & Row. https://hdl.handle.net/2027/heb.08419

Cooper, W., Eady, S., & Mueller, P. (1985). Acoustical aspects of contrastive stress in question-answer contexts. *The Journal of the Acoustical Society of America*, 77, 2142–2156. https://doi.org/10.1121/1.392372

Flege, J. E., Bohn, O. S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25, 437–470. https://doi.org/10.1006/jpho.1997.0052

Fujimori, A., Yamane, N., & Yoshimura, N. (2017). The acquisition of prosodic focus marking in English dialogs and narratives. In M. Hirakawa, J. Matthews, K. Otaki, N. Snape, & M. Umeda (Eds.), *Proceedings of PacSLRF* 2016 (pp. 53–57). Chuo University.

Goss, S., & Nakayama, M. (2011). Prosody and comprehension in oral reading by L2 Japanese learners. *Second Language*, 10, 33–50. https://doi.org/10.11431/secondlanguage.10.0_33

Hosaka, H. (1998). Discourse intonation in English and Japanese EFL speakers: A pilot study. *ARELE*, 9, 107–116. https://doi.org/10.20581/arele.9.0_107

Ishihara, S. (2011). Japanese focus prosody revisited: Freeing focus from prosodic phrasing. *Lingua*, 121, 1870–1889. https://doi.org/10.1016/j.lingua.2011.06.008

Ito, J., & Mester, A. (2012). Recursive prosodic phrasing in Japanese. In T. Borowsky, S. Kawahara, T. Shinya, & M. Sugahara (Eds.), *Prosody matters: Essays in honor of Elisabeth Selkirk* (pp. 280–303). Equinox.

Jackendoff, R. (1972). *Semantic interpretation in generative grammar*. The MIT Press.

Jannedy, S. (2002). Hat patterns and double peaks: The phonetics and psycholinguistics of broad versus late narrow versus double focus intonations (Unpublished doctoral dissertation). The Ohio State University. https://doi.org/10.13140/RG.2.2.19343.18089

Kubozono, H. (1993). *The organization of Japanese prosody*. Kurosio.

Kunevich, M., & Alexopoulou, D. (2014). *Acquisition at interfaces: The case of prosodic focus marking*. Paper presented at European Second Language Association 24, University of York, UK.

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, 8, 249–336.

Mori, Y., Hori, T., & Erickson, D. (2014). Acoustic correlates of English rhythmic patterns for American versus Japanese speakers. *Phonetica*, 71, 83–108. https://doi.org/10.1159/000365594

Nakayama, M., & Yoshimura, N. (2015). The modularity of grammar in L2 acquisition. In M. Nakayama (Ed.), *Handbook of Japanese psycholinguistics* (pp. 235–570). De Gruyter Mouton. https://doi.org/10.1515/9781614511212-015

Nespor, M., Shukla, M., van de Vijver, R., Avesani, C., Schraudolf, H., & Donati, C. (2008). Different phrasal prominence realizations in VO and OV languages. *Lingue e Linguaggio*, 7, 139–168. https://doi.org/10.1418/28093

Pierrehumbert, J., & Beckman, M. (1988). *Japanese tone structure*. The MIT Press.

Rochemont, M., (1986). *Focus in generative grammar*. John Benjamins. https://doi.org/10.1075/sigla.4

Schwartz, B., & Sprouse, R. (1994). Word order and nominative case in non-native language acquisition: A longitudinal study of German interlanguage. In T. Hoekstra, & B. Schwartz (Eds.), *Language acquisition studies in generative grammar* (pp. 317–368). John Benjamins. https://doi.org/10.1075/lald.8.14sch

Selkirk, E. (1984). *Phonology and syntax: The relation between sound and structure*. The MIT Press. https://doi.org/10.11435/gengo.136.0_35

Selkirk, E., & Tateishi, K. (1991). Syntax and downstep in Japanese. In C. Georgopoulos & R. Ishihara (Eds.), *Interdisciplinary approaches to language: Essays in honor of S.-Y. Kuroda* (pp. 519–543). Kluwer. https://doi.org/10.1007/978-94-011-3818-5_28

Selkirk, E. (2009). On clause and intonational phrase in Japanese: The syntactic grounding of prosodic constituent structure. *Gengo Kenkyu*, 136, 35–73.

Sityaev, D., & House, J. (2003). Phonetic and phonological correlated of broad, narrow and contrastive focus in English. In M. J. Sole, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1819–1822). Causal Productions.

Slabakova, R. (2009). What is easy and what is hard to acquire in a second language? In M. Bowles, T. Ionin, S. Montrul, & A. Tremblay (Eds.), *Proceedings of the 10th Generative Approaches to Second Language Acquisition Conference (GASLA 2009)* (pp. 280–294). Cascadilla Proceedings Project.

Sorace, A. (2004). Native language attrition and developmental instability at the syntax-discourse interface: Data, interpretations and methods. *Bilingualism: Language and Cognition*, 7, 143–145. https://doi.org/10.1017/S1366728904001543

Sorace, A. (2011). Pinning down the concept of 'interface' in bilingualism. *Linguistic Approaches to Bilingualism*, 1, 1–33. https://doi.org/10.1075/lab.1.1.01sor

Sorace, A., & Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Language Research*, 22, 339–368. https://doi.org/10.1191/0267658306sr271oa

Sugahara, M. (2003). Downtrends and post-focus intonation in Tokyo Japanese (Unpublished doctoral dissertation). University of Massachusetts, Amherst. https://scholarworks.umass.edu/dissertations/AAI3079597

Takeda, A. (2018). English prosodic marking of information structure by L1-Japanese second language learners (Unpublished doctoral dissertation). University of Hawai'i, Manoa. http://hdl.handle.net/10125/62787

Truckenbrodt, H. (1995). Phonological phrases – Their relation to syntax, focus, and prominence (Unpublished doctoral dissertation). MIT.

Tsimpli, I., & Sorace, A. (2006). Differentiating interfaces: L2 performance in syntax-semantics and syntax-discourse phenomena, In D. Bamman, T. Magnitskaia, & C. Zaller (Eds.), *Proceeding of the 30th annual Boston University Conference on Language Development* (pp. 653–664). Cascadilla Press.

Venditti, J. J., Maekawa, K., & Beckman, M. E. (2008). Prominence marking in the Japanese intonation system. In S. Miyagawa & M. Saito (Eds.), *The Oxford handbook of Japanese linguistics* (pp. 456–512). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195307344.013.0017

White, L. (1989). *Universal grammar and second language acquisition*. John Benjamins. https://doi.org/10.1075/lald.1

White, L. (2003). *Second language acquisition and Universal Grammar*. Cambridge University Press. https://doi.org/10.1017/CBO9780511815065

White, L. (2011). Second language acquisition at the interfaces. *Lingua*, 121, 577–590. https://doi.org/10.1016/j.lingua.2010.05.005

Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, 33, 159–197. https://doi.org/10.1016/j.wocn.2004.11.001

Zimmermann, M. (2008). Contrastive focus and emphasis. *Acta Linguistica Hungarica*, 55, 347–360. https://doi.org/10.1556/ALing.55.2008.3-4.9

Zimmermann, M., & Onea, E. (2011). Focus marking and focus interpretation. *Lingua*, 121, 1651–1670. https://doi.org/10.1016/j.lingua.2011.06.002

Zubizarreta, M. (1998). *Prosody, focus, and word order*. The MIT Press.

Transfer of prosodic representation

L1 Bengali production of L2 English regular simple past tense

Jacqueline Ingham
The University of Sheffield

This study examines the spoken suppliance of inflectional morphology by L1 Bengali speakers of L2 English in the phonological framework of the Prosodic Transfer Hypothesis (Goad, White, & Steele, 2003; Goad & White, 2004; Goad & White, 2006, et seq.). Data from a semi-spontaneous elicitation task suggests that, at lower levels of proficiency, production of inflection is partially conditioned by the stem vowel in terms of vowel length and the voicing status of the stem-final consonant. This finding is proposed to be indicative not only of transfer of L1 prosodic representations, but also transfer of L1 word minimality requirements and moraic structure below the level of the prosodic word. Evidence of such transfer is arguably visible when there is a mismatch between word minimality and the distribution of syllable weight in the L1 and L2.

Keywords: prosodic transfer, Bengali, prosodic representation, inflectional morphology, minimal word constraints

1. Introduction

The persistent omission of inflectional morphology in the spoken production of adult second language (L2) speakers of English is well-documented in second language acquisition studies, especially within those focused on investigating whether adult learners have continued access to the Universal Grammar (UG) feature inventory or not. One strand within this research has explored whether adult learners can acquire L2 *uninterpretable* features that are not present in the first language (L1), and how restricted access to UG in adulthood may account for the omission of functional morphology in spoken language (Hawkins & Hattori, 2006; Tsimpli & Dimitrakopoulou, 2007). Another strand of research assumes that access to

https://doi.org/10.1075/lald.67.07ing

syntactic functional categories and features continues through adulthood, either via L1 transfer or UG, such that failure to supply inflectional morphology is due to other factors. Within the latter strand, some researchers have highlighted difficulties in mapping L2 syntactic categories to L2 surface morphology (e.g., Prévost & White, 2000), while other researchers point to problems in the assembly and reassembly of syntactic features present in the L1 and the L2 (Lardiere, 2008).

Offering a phonological explanation, the Prosodic Transfer Hypothesis (Goad, White, & Steele, 2003; Goad & White, 2004; Goad & White, 2006) proposes that rates of suppliance or omission of inflectional morphology are related to the availability and transfer of L1 prosodic representations into the interlanguage. According to the Prosodic Transfer Hypothesis (PTH), the way in which L1 affixes are either attached or incorporated into the prosodic word can help determine the extent to which inflectional morphology is supplied in L2 oral production. A number of studies have tested the PTH via the production of L2 functional morphology. This research includes studies on the production of simple past tense, present perfect, and third person singular agreement by L1 Mandarin speakers (Goad, White, & Steele, 2003; Goad & White, 2006) and the acquisition of articles and tense by L1 Turkish speakers (Goad & White, 2004). The PTH has also been tested against other accounts that explain the omission of functional morphology. See, for example, Goad and White (2006) on the transfer of L1 syntactic settings as proposed under the Representational Deficit Hypothesis (e.g., Hawkins & Liszka, 2003; Hawkins & Chan, 1997), and Goad and White (2008) on the role of frequency and input in emergentist approaches to L2 learning (e.g., Bybee, 2001; Ellis, 2002, 2003).

In one such study, Cabrelli Amaro, Campos-Dintrans, and Rothman (2018) investigated the suppliance of regular simple past tense by L1 Mandarin, Japanese, and Spanish speakers in relation to the PTH and Representational Deficit accounts, in both written and spoken modes. All three languages are reported to have constraints against consonant clusters on the right-edge; both Japanese and Mandarin have strict restrictions on which singleton consonants may occur in the rhyme, and Spanish is said to undergo second consonant deletion in non-formal speech (Cabrelli Amaro et al., 2018: 511–513). The authors' analysis of L1 syntactic and prosodic representations indicated that Japanese speakers were best positioned to produce L2 English regular simple past inflectional morphology because of two factors: the instantiation of the [*u*past] feature in their L1 and the availability of prosodic word (PWd) adjoined representation, required in the L2. In comparison, Mandarin has no PWd-adjoined prosodic representation or an instantiated [*u*past] feature, while Spanish has a [*u*past] feature but no PWd-adjoined prosodic representation readily available to transfer. Findings indicated that while there was a statistically significant difference between the oral mean suppliance rates of regular simple past inflectional morphology between the Japanese and Mandarin groups, (83% and 71% respectively), there was no statistical significance between the Japanese and Spanish groups (83% and 77% respectively), nor between the Spanish and Mandarin groups. Thus, according to Cabrelli Amaro et al., (2018), having a readily available PWd-adjoined prosodic representation is advantageous for L1 Japanese speakers of L2 English. However, the ability to construct a PWd-adjoined prosodic representation from existing prosodic structures may also account for the lack of statistical significance between the three L1 groups. Cabrelli Amaro and colleagues (2018, p. 523) concluded that whilst their study does not refute representational deficit accounts, their findings show that suppliance of morphology is greatly influenced by a complex of phonological factors. These phonological factors may exert a single or combinatory influence in terms of prosodic and/or phonotactic constraints.

The current study is concerned with phonological factors of the prosodic kind. In particular, based on the assumption that it is possible for learners to construct the required L2 PWd representations, the study is designed to investigate whether successful suppliance of L2 inflectional morphology is also dependent upon L1 transfer of prosodic constituents below the level of the PWd, specifically, transfer of moraic structure and word minimality requirements. Thus, this investigation studies the extent to which transfer of L1 prosodic structure, both at and below the level of the prosodic word, may interact with the availability of the required L2 prosodic representations. The proposal put forward here is that when there is a mismatch between the L1 and L2 word minimality requirements, a more complex situation emerges: Whether L2 inflection can be incorporated or adjoined to the PWd also depends upon the relative well-formedness of the PWd in relation to the L2. Under this analysis, the suppliance of L2 inflectional morphology is influenced, to varying degrees, by (a) the availability of the required prosodic representation and (b) the equivalence of moraic structure and minimality requirements between the L1 and L2.

This study offers data from an oral production task focusing on the suppliance of regular simple past morphology by L1 Bengali speakers of L2 English. Foreshadowing the results, I propose that a mismatch between L1 Bengali and L2 English word minimality requirements can be captured in the L2 production of English verbs, depending on whether the verb has a long (tense) vowel or a diphthong (LV), and thus considered well-formed according to Bengali word minimality requirements, or a short (lax) vowel (SV), and thus considered subminimal according to Bengali word minimality requirements. Broadly, thus, this investigation shows that further examination within a non-statistically significant trend can provide insight into the role of L1 transfer of prosodic representation in relation to the production of L2 inflectional morphology. These conclusions are based on my analysis of the verb stem, which reveals that, at some levels of proficiency, there is a statistically significant difference regarding whether the L2 English SV verb stems are supplied with inflection according to the voicing status of the stem-final consonant. Verbs including both a short vowel and an unvoiced stem-final consonant are supplied with inflection more consistently than those SV verb stems with a voiced stem-final consonant; an asymmetry that can arguably be explained by the notion of transfer of L1 word minimality requirements and a potential sensitivity to L2 vowel length in relation to the coda environment.

The organisation of this paper proceeds as follows. First, I present an outline of the theoretical background to the study, setting out the prosodic representation required for English regular simple past tense in accordance with the PTH (Goad et al., 2003; Goad & White, 2004, 2006), with special reference to the minimality requirements of the English PWd. This section is followed by an analysis of the prosodic representation available to transfer from L1 Bengali, as well as Bengali word minimality requirements. Finally, a brief summary of the predictions and methodology is followed by a presentation of the results and discussion.

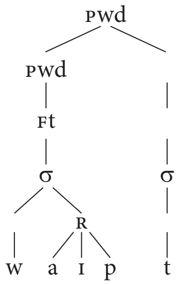
2. Theoretical background to the study

2.1 Prosodic representation of English regular simple past tense

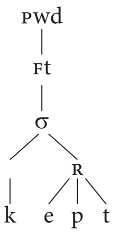
Whether English inflectional morphology is prosodically represented in a PWd-adjoined or PWd-internal structure, the internal structure of the prosodic word adheres to the structural organisation of the prosodic hierarchy (Nespor & Vogel, 2007; Selkirk, 1997), which is constrained by permissible relations between constituents. Goad et al. (2003) propose that target-like production requires a PWd-adjoined prosodic representation, as illustrated in (1), showing the regular past form ‘wiped’. The inflection is adjoined to the stem by a higher PWd, and it is not internal to the PWd. The irregular simple past form ‘kept’ (as shown in (2) for comparison) illustrates inflection incorporated internal to the original PWd.[1]

1. See Goad et al. (2003) and Goad and White (2006) for a detailed analysis of the prosodic representation of English inflectional morphology and the framework of constraints which determine the relationship between constituents.

(1)



(2)



Of relevance to the current study is the number of segments that are permitted in word-final rhymes. Goad et al. (2003, p. 248) illustrate this in relation to monomorphemic words and inflected forms. The authors claim that a maximally ternary rhyme is permissible in word-final position, and this is said to hold across a number of languages (Goad & White, 2019, p. 777). English monomorphemic words conform to this constraint, and allow word-final VCC or VVC. This is illustrated in Example (3) for word-final VCC and Example (4) for word-final VVC sequences. Furthermore, when Class I derivational suffixes are attached, the vowel is shortened in order to conform to the constraint on the number of segments permitted in the rhyme, as shown in Example (5). Examples (3) to (5) are reproduced from Goad and White (2019, p. 777).[2]

(3) a. [hɛlm] 'helm'
b. [hɛmp] 'hemp'
c. *[hɛlmp]

(4) a. [rijm] 'ream'
b. [rijp] 'reap'
c. *[rijmp]

(5) a. [dijp] 'deep'
b. [dɛpθ] 'depth'
c. *[dijpθ]

2. Example (4) shows the phonetic reality of an English long vowel as a vowel plus glide sequence, but it could also be transcribed as a long vowel [riːm].

Despite this patterning, it should be noted that there are exceptions to the maximal three-segment word-final constraint within monomorphemic forms. These can be specifically categorized. Goad and White (2019, p. 777) identify a VVCC-final configuration in cases when the cluster shares coronal feature, such as 'paint' [pejnt], as well as the existence of 'true exceptions', such as 'Manx' [mæŋks].

Returning to the contrast between monomorphemic and inflected forms, inflected regular stems do not conform to the same length constraints that are seen across the non-exceptional monomorphemic forms. This is shown in Example (6) for third person singular agreement, simple past (and past participle) and plural agreement (reproduced from Goad and White, 2019, p. 777).

(6) a. [hɛlmz] 'helms'
b. [rijpt] 'reaped'
c. [dɛpθs] 'depths'

In order to explain this further, it is necessary to turn to irregular verbs. Goad and White (2019) argue that (pseudo-inflected) irregular verbs in the past tense also permit a maximally ternary rhyme, but in this case, unlike with regular verbs, it is achieved by a process of vowel shortening, as shown in (7) with the irregular verb 'keep' (7a) and 'kept' (7b).

(7) a. [ki:p] 'keep'
b. [kɛpt] 'kept'
c. *[ki:pt]

The structure in (7c) is not permissible; the addition of inflection to an already ternary-branching rhyme (7a) violates the maximum number of segments permitted at the right-edge. The legitimate form is the example shown in (7b), the prosodic representation of which is shown in (2). The inflection is internal to the PWd, with a shortened vowel, avoiding violation of a maximal ternary rhyme at the right-edge.

2.2 Interim summary

In accordance with the PTH, the suppliance of simple past inflectional morphology in L2 English requires that the speaker has both a PWd-internal and a PWd-adjoined prosodic representation, required for irregular and regular verb forms respectively. If available in the L1, these prosodic representations may be transferred to the interlanguage. If not, existing L1 representations may be minimally adapted. This may be achieved by (a) licensing L1 structures to a different (L2) morphosyntactic domain, (b) creating the required L2 representation by combining L1 licensing relations, or (c) allowing an L1 structure to be licensed at a different edge in the prosodic domain (Goad & White, 2019, p. 788).

2.3 Prosodic representation of Bengali past tense

2.3.1 *Bengali verbal morphology*

Typologically, the verbal morphology of Bengali may be classified as agglutinating (Kar, 2009), mainly agglutinating (Mazumdar, 1920), or partly agglutinating and fusional (Boyle David, 2015). As shown in (8), there is a strict order of suffixation for Bengali verbal inflection (Ray, Hai, & Ray, 1966).

(8) root – aspect suffix – tense suffix – person suffix

Bengali roots are either monosyllabic and monomoraic, or disyllabic and bimoraic (Klaiman, 1990, p. 501), as illustrated in (9). In (9a), the lax vowel is monomoraic, and the coda consonant does not add weight (Fitzpatrick-Cole, 1996). In (9b), there are two vowels contributing to a bimoraic form.

(9) a. dæk^h- 'see'
b. g^{h}uma- 'sleep'

Bengali monosyllabic verb bases with non-high root vowels, such as *dæk^h-* ('see') and *boj^h-* ('understand') have two alternate forms; a 'low' form and a 'high' form, as shown in (10) and (11). The low root vowel alternates are predictable (i.e., /o/ raises to /u/, /ɔ/ to /o/, /e/ to /i/, /æ/ to /e/ and /a/ to /e/). It should be noted that in the example given in (11) 'oj' is not a diphthong and 'j' is not a glide, but [j^h] is an aspirated palatal.[3]

(10) a. dæk^h-
b. dek^h-

(11) a. boj^h-
b. buj^h-

The inflection patterns in Bengali verbal morphology demonstrate the phonological process of vowel raising, which is relevant in determining whether inflection is prosodified internally, or adjoined to the PWd (see also Ingham (2019) for further discussion on Bengali prosodic representation). For example, verbs are marked for person in the final position (8). The person suffix morpheme is dependent upon the form of the tense or aspect marker it directly follows. When the root verb *boj^h-* 'to

3. Bengali words, for example, *boj^h-* ('understand'), can be transcribed according to (a) the International Alphabet of Sanskrit Transliteration (IAST) *bojh-*, (b) the International Phonetic Alphabet (IPA) *bodʒ^h-,* or (c) a combination of both IAST and IPA *boj^h-*. The system adopted in the current study is that in (c). Thus, aspirated consonants are denoted with a superscript (e.g., *k^h* and *g^h* rather than *kh* and *gh*) and voiced aspirated palatal as *j^h* rather than *dʒ^h*. See Ingham (2019) for further discussion on this and vowel height assimilation.

understand' is inflected in the simple past (-*l*-), and in the first person (-*am*), the root vowel is raised to the high form *buj*^{h}-, as shown in (12).

(12) buj^{h} -l -am
understand -PST 1PERS
'I understood.'

The inflection pattern for the simple present tense, which does not have an overt marker, is illustrated with the verb *lek*^{h}- 'to write' (high form *lik*^{h}-), marked in the third person in (13a), and the first person in (13b).

(13) a. lek^{h} -ø -e
write -PRS -3PERS
'he/she writes.'
b. lik^{h} -ø -i
write -PRS -1PERS
'I write.'

Whilst the examples in (13) clearly show that the first vowel is raised only in the presence of a high vowel in the person morpheme, the source of the raising of the vowel in (12) is less transparent. The vowel in the inflectional morpheme -*am* is not high, but the root vowel has raised from *boj*^{h}*- to buj*^{h}- (Ingham, 2019, p. 59). The diachronic explanation put forward by Lahiri (2000) can account for the presence of the high form in non-present tenses, regardless of the following vowel in the person affix. This explanation demonstrates that in the written form of 'literary Bengali' (Shadhu bhasha), there is an overt inflectional morpheme. This underlying morpheme is a [+high] marker /-*i*-/, which triggers vowel raising in non-present simple tenses. Whilst it is retained and visible in the written form, this morpheme is no longer realized in spoken language (Lahiri, 2000, p. 75). Returning to the example in (12), vowel raising is triggered by the high vowel /-*(i)l*-/, which is not pronounced in the spoken form, as shown in (14).

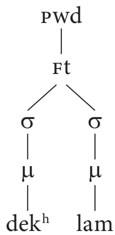
(14) buj^{h} -*i* -l -am (buj^{h}-l-am)
understand (-*i*) -PST -1PERS
'I understood.'

Although the vowel raising in forms such as *buj*^{h}*lam* 'I understood' can be viewed as a result of a diachronic process, the vowel raising in *lik*^{h}*i* 'I write' is evidence of synchronic vowel raising, because whether the root vowel raises or not is dependent upon the height of the vowel in the person marker. Vowel raising is triggered in the simple present tense, but only if the vowel in the person marker has a [+high] feature, providing evidence that first-person marker /-*i*-/ triggers vowel raising within the prosodic word, when there is no intervening prosodic boundary. I will return to this issue in Sections 2.3.2 and 2.3.3. In the following section, it will be seen how the placement of emphatic clitics can indicate a prosodic boundary.

2.3.2 *Bengali prosodic representation: Simple past*

Bengali simple past tense is arguably represented in a PWd-internal structure, indicating a difference in prosodic representation between Bengali simple past and English regular simple past tense. This is illustrated in (15) with the verb *dækʰ-* 'to see' (high alternate dekʰ-) in the first-person past form *dekʰlam* ('I saw'), and with respect to the placement of emphatic clitics (Bayer & Lahiri, 1990; Fitzpatrick-Cole, 1990, 1996; Lahiri, 2000).

(15)



A clitic such as =*o* ('also') cannot stand as a prosodic word; it must be attached to a well-formed prosodic word. Clitics cannot instigate phonological processes, such as vowel height assimilation, and can only be attached once all other phonological processes, and derivational and inflectional affixation are complete (Bayer & Lahiri, 1990; Fitzpatrick-Cole, 1990, 1996; Lahiri, 2000). This is illustrated in (16), showing a finite verb with the clitic attached to the PWd.[4] However, in (17), the clitic is inserted between the root and inflectional morphology and in (18), the clitic is positioned between the tense and person markers. The configurations in (17) and (18) are not grammatical.

(16) ((dekʰ -l -am)$_{PWd}$ =o) $_{PWd}$
see -PST -1PERS =also
'I also saw.'

(17) *(dekʰ =o -l -am) $_{PWd}$
see =also -PST -1PERS
'I also saw.'

(18) *(dekʰ -l =o -am) $_{PWd}$
see - PST =also -1PERS
'I also saw.'

---

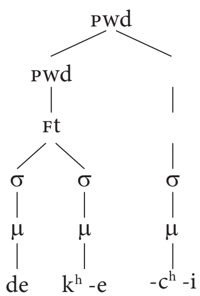
4. Clitics can attach to both finite and non-finite verb forms, as well as most other word classes. Examples (15), (16) and (17) are reproduced and adapted from Ingham (2019, p. 66).

This indicates that clitics cannot attach to a monomoraic root. A monomoraic root is not bimoraic until either (vocalic) inflectional material is added, introducing an additional mora, or the root vowel is lengthened to avoid a subminimal surface form (Fitzpatrick-Cole, 1990). This can be seen in the simple present tense for second person ordinary, where the person suffix /-*o*/ can attach to the root, as in *dækʰo* 'you see' and *bojʰo* 'you understand'.[5] I will return to the fact that clitics attach at the PWd rather than the phonological phrase (PhP) in Section 2.3.3.

2.3.3 *Bengali prosodic representation: Present perfect*

Again, vowel raising and clitic placement can also help determine where prosodic boundaries lie in Bengali present perfect and past perfect tenses (Bayer & Lahiri, 1990; Fitzpatrick-Cole, 1990, 1996; Lahiri, 2000). This is illustrated in (19) with the first person present perfect form *dekʰecʰi* ('I have seen'), constructed from the root verb *dækʰ-* 'to see' (high form *dekʰ-*).

(19)[6]



The Bengali PWd adjoined representation in (19) can be broken down into two steps. First, the perfect participle is formed by addition of the perfect affix /-*e*/, which attaches to the root in a PWd-internal representation. It is the unpronounced /-*i*-/ vowel that arguably triggers the high alternate *dekʰ-*, a process that is internal to the prosodic word. As we will see in Section 2.4.1, this also adds a mora (*de- kʰ-e*), and thus the perfect participle is a well-formed prosodic word (20).

(20) (dekʰ -*i* -e)$_{\text{PWd}}$ (dekʰ-e)
see (-*i*) -PRFP
'seen'

5. Note that verbs in the present simple for second person ordinary do not undergo vowel raising, unlike those in the present simple for second person familiar (e.g., *dækʰo* 'see-2PERS.ORD'; *dekʰis* 'see-2PERS.FAM'; *bojʰo* 'understand-2PERS.ORD'; *bujʰis* 'understand-2PERS.FAM').

6. Example (19) reproduced from Ingham (2019, p. 75).

This forms the base for the present and past perfect forms. Following the suffixation order set out in (8), the perfect marker /-*c*^{h}-/, unpronounced present marker /-ø-/ and first-person marker /-*i*/ form a stem, which is then attached to the perfect participle in a PWd-adjoined representation, as illustrated in (21).

(21) (((dek^h -e)$_{\text{PWd}}$ -c^h -ø -i)$_{\text{STEM})\text{PWd}}$
see.PRF -PRF -PRS -1PERS
'I have seen.'

This is in contrast to the PWd-internal representation illustrated earlier in examples such as *lik^{h}i* 'I write', where the first-person marker /-*i*/ is argued to be internal to the PWd. Evidence that the perfect participle *dek^h-e* is a prosodic word to which the perfect and person markers are adjoined as a stem is from (a) vowel raising (the triggering of the high form within a prosodic domain), as seen in (22), and (b) the placement of clitics. As discussed in Section 2.3.2, clitics can attach to a prosodic word, but not to a root (Bayer & Lahiri, 1990; Fitzpatrick-Cole, 1990, 1996; Lahiri, 2000). This is illustrated in Examples (22) through (25), modified from Fitzpatrick-Cole (1996). The clitic attaches to prosodic words in Examples (22), (24), and (25), but not in (23) because the verb stem does not constitute a prosodic word until the perfect marker /-*e*/ is attached. Once the high alternate form is triggered within the prosodic boundary, as seen in (22), the clitic can then adjoin to the PWd. Unable to operate as either a free clitic or independent PWd, the clitic must therefore attach to the PWd and not the PhP.

(22) ((dek^h -e)$_{\text{PWd}}$ =o) $_{\text{PWd}}$
see -PRFP =also
'seen also'

(23) *((((dek^h =o -e)$_{\text{PWd}}$) -c^h -ø -i)$_{\text{STEM}}$)$_{\text{PWd}}$
see =also -PRF -PRF -PRS -1PERS
'I have also seen.'

(24) (((((dek^h -e) $_{\text{PWd}}$) -c^h -ø -i)$_{\text{STEM}}$)$_{\text{PWd}}$=o) $_{\text{PWd}}$
see .PRF -PRF -PRS -1PERS =also
'I have also seen.'

(25) (((((dek^h -e)$_{\text{PWd}}$ =o)$_{\text{PWd}}$) -c^h -ø -i)$_{\text{STEM}}$)$_{\text{PWd}}$
see .PRF =also -PRF -PRS -1PERS
'I have also seen.'

Furthermore, Example (26) shows that the clitic cannot be placed within the aspect-tense-person stem, indicating that the affixes for aspect, tense and person together constitute a unit, and cannot, therefore, be interrupted by a prosodic boundary. Example (26) illustrates this in the present perfect, modified and adapted from Bayer and Lahiri, 1990, p. 5.[7]

(26) *((((dekʰ -e)$_{\text{PWd}}$) -cʰ =o -ø -i)$_{\text{STEM}}$)$_{\text{PWd}}$
see .PRF -PRF =also -PRS -1PERS
'I have also seen.'

Finally, the same patterns can be seen with the high vowel clitic =*i*, which adds an inclusive emphasis of 'just that and no other' to an utterance (Boyle David, 2015, p. 171). Neither =*i* (=EMPH) nor =*o* (=also) trigger vowel raising, again indicating that the clitic is adjoined to the PWd. This is illustrated in (27), modified from (13), with the verb *lekʰ-* 'to write' (high form *likʰ-*) in the present tense in both the third and first-person.

(27) a. lekʰ -ø -e =i
write -PRS -3PERS =EMPH
'He/she writes.'
b. likʰ -ø -i =i
write -PRS -1PERS =EMPH
'I write.'

In (27a), vowel raising has not been triggered with the addition of the third person morpheme /-e/, but a prosodic word has been formed with the introduction of the vocalic suffix. Recall that in Example (20) the perfect participle *dekʰe* 'seen' is formed with the unpronounced high vowel and the perfect affix /-*(i)e*-/, which triggers vowels raising. In (27), the addition of the high vowel clitic /=i/ does not trigger vowel raising. This is because the clitic is attached outside the PWd, adjoined and not internal to the PWd, and it cannot, therefore, trigger alternation of the root vowel. In contrast, in (27b), the first-person morpheme is a high vowel /-*i*/ which triggers vowel raising and satisfies the word minimality requirements, producing a well-formed PWd.

2.3.4 *Interim summary*

The phonological process of vowel raising and the placement of clitics can help illustrate how inflectional morphology is prosodified in Bengali. English simple past tense is prosodified in both a PWd-internal (irregular verbs) and PWd-adjoined prosodic representation (regular verbs), whereas Bengali past tense morphology is prosodified within the PWd. Assuming transfer from the L1 to the interlanguage,

7. Bayer and Lahiri (1990) refer to the stem and inflection, and do not propose that the stem constitutes a PWd. Following Fitzpatrick-Cole (1996), it is assumed here that the root verb is a PWd once it fulfils the minimal bimoraic requirement.

L1 Bengali speakers of L2 English need to reorganise the L1 past tense prosodic structure to accommodate that required for L2 English regular simple past inflection. As the PWd-adjoined representation is available in the production of Bengali past perfect and present perfect forms, the transferred PWd-adjoined representation is viable for minimal adaptation, to be licensed to a new morphosyntactic domain (Goad & White, 2006; Goad & White, 2019). In Section 2.4, we consider whether a mismatch in minimality requirements between an L1 and L2 may also influence whether L2 inflectional morphology is supplied.

2.4 Minimality requirements in Bengali and English

Whilst it seems that there is opportunity for transfer and minimal adaptation of L1 Bengali prosodic representation of inflection to accommodate the requirements of L2 English simple past tense, there are, however, differences between Bengali and English minimality requirements (see also Ingham (2019) for further discussion). This is in relation to coda consonants, and whether they are moraic or not. According to the moraic theory of syllable weight (Hayes, 1989; Hyman, 1985), Bengali and English share a bimoraic minimal word requirement (Broselow & Park, 1995; Fitzpatrick-Cole, 1990, 1996; Gordon, 2013). A light syllable constitutes a single mora (monomoraic), whilst a heavy, bimoraic syllable is made up of two. A bimoraic minimal word, therefore, requires either two light syllables or one heavy syllable ([μμ]PWd). Moras contribute to the weight of a syllable, but not all segments contribute to weight: some segments may be non-moraic or they may be moraic but extrametrical and therefore non-contributory to weight calculations. For example, onset consonants are non-contributory to syllable weight (Ewen & van der Hulst, 2001), and not all coda consonants are moraic.

2.4.1 *Minimality requirements in Bengali*

Following the McCarthy and Prince 1986 account (1996), Fitzpatrick-Cole (1990, p. 157) proposes that in Bengali the Minimal Word Constraint regulates the phonological processes of glide formation and vowel deletion, invoking a process of vowel lengthening in order to avoid subminimal monomoraic forms and ensure a well-formed output (i.e., a disyllabic form with two light syllables or a monosyllabic form with one heavy syllable). According to Fitzpatrick-Cole (1990), this is accounted for in Prosodic Lexical Phonology (Inkelas, 1991), where bare verbs are bound roots, and nouns are free stems. Glide formation and vowel deletion (targeting certain VV patterns) are not applied across the board; glide formation is blocked in the environment of suffixation or cliticisation with subminimal stems, as shown in Table 1, examples (d) and (e), but not in the suffixation of subminimal roots, examples (a) and (b). Note that 'y' and 'Y' are glides, corresponding to the high and mid vowels 'i' and 'e' respectively. All examples adapted from Fitzpatrick-Cole (1990, pp. 157–161).[8]

Table 1. Examples of glide formation (GF), vowel deletion (VD) and lengthening (VL)

	Monomoraic root+affix	**Monomoraic word**	**Monomoraic stem+affix**	**Monomoraic stem+clitic**
GF	(a)/j^a -i/ → jay go −1PERS	(b)/boi/ → boy book	(c) /ca -e/ → cae tea -OBL *caY	(d)/ca =i/ → ca:i tea -EMPH *ca:y
VD	(f)/k^{h}a -is/ → k^{h}as eat −2PERS.FAM		(e)/pa -er/ → paer leg -GEN	
VL	(g)/k^{h}a -is/ → k^{h}a:s eat −2PERS.FAM	(i)ca/ → ca: tea (j)/rag/ → ra:g anger	(k)ca -e/ → cae (l)/rag -i/ → ragi angry *ca:e/*ra:gi	(h)ca =i/ → ca:i tea -EMPH

Vowel deletion is blocked when VV occurs with an affix following a monomoraic stem, as in example (e). However, the second V is always deleted in the environment of a vowel-final monomoraic root, such as *k^{h}a* 'to eat' (according Fitzpatrick-Cole (1990, p. 164), this is to avoid branching of the weak mora), resulting in a monomoraic form, as shown in example (f). This subminimal form is then repaired with the addition of a mora, which is filled via vowel lengthening and realised as [*k^{h}aːs*] 'eat-2PERS.FAM', as shown in example (g). When clitics are attached to monomoraic stems, as in example (h), vowel lengthening also applies. Vowel lengthening can also operate in monosyllabic forms that are both open (CV) or closed (CVC), as illustrated in examples (i) and (j), but not in the environment of a monomoraic stem and affix, as in examples (k) and (l). The final consonant does not help to satisfy word minimality, see also Example (28), which suggests that consonants are either non-moraic (or extrametrical) in word-final position (Fitzpatrick-Cole, 1990, p. 158), and that in subminimal forms, phonological processes ensure that the bimoraic minimal word requirement ([μμ]PWd) is fulfilled, with vowel lengthening as the final solution. In sum, this suggests a minimal CVV word template.[9]

(28) a. /ca/ [caː] 'tea'
b. /am/[aːm] 'mango'

8. The Bengali honorific system has three levels, -2PERS.FAM indicates the inflectional marker for second-person familiar (e.g., Klaiman, 1990).

9. Whilst Fitzpatrick-Cole (1990) posits that vowels and glides are moraic, but final consonants are not, it is also suggested that consonants in final-position could be moraic but extrametrical. Of importance for the current study is that weight is provided by vowels and glides.

2.4.2 *Interim summary*

Following Fitzpatrick-Cole (1990), the minimal word in Bengali is taken to be a bimoraic prosodic word ([μμ]ᴘᴡd). This is illustrated via phonological processes of glide formation, vowel deletion and vowel lengthening, dependent upon whether the monomoraic form is a stem or a root (and in some conditions whether the root is vowel or consonant final), and whether it is in the environment of an affix or a clitic. Vowel lengthening is seen to be the final course of action to prevent a subminimal form. As this can operate both within open and closed monosyllabic forms, it is assumed that word-final coda consonants are either non-moraic or extrametrical.

2.4.3 *Minimality requirements in English*

English vowels are subject to changes in both quantity and quality in the long-short or lax-tense opposition; English vowel length is associated with tenseness. Tense vowels, such as the /iː/ in 'flea', are closer, higher, and longer than the more open, lower, and shorter lax counterpart /ɪ/ in 'pip'. Tense vowels and diphthongs (29b) are bimoraic, and lax vowels are monomoraic, as illustrated in (29), where <u>m</u> represents mora.

(29) a. fly [fla<u>m</u>ɪ<u>m</u>]
b. flea [fli<u>mm</u>]
c. pip [pɪ<u>m</u>p<u>m</u>]

In English, syllable weight is distributed over the vowel and final consonant, so that final consonants are weight-bearing and the smallest lexical word is CVX (Gordon, 2013, p. 212), where X is either a consonant or vowel. Although there are monosyllabic English CVV words with a diphthong or tense vowel, as in (29a) and (29b), or monosyllabic CVC words with a lax vowel, as in (29c), there are no CV words such as *fi.

2.4.4 *Pre-lenis lengthening in English but not Bengali*

In addition to having a lax-tense distinction, English vowels are also sensitive to the surrounding consonantal environment (the vowel in 'dip' is shorter than that in 'dig') dependent upon whether the word-final consonant is voiced (e.g., /b, d, g/) or unvoiced (e.g., /p, t, k/). Whether vowel length mutation in the environment of a following voiced or unvoiced consonant is a language universal is open to discussion (see, amongst others, Maddieson, 1997). Whilst it is possible that both of these rules may operate within a language, Roberts, Kotzor, Wetterlin, and Lahiri (2014) find no evidence to support a 'pre-fortis lengthening' rule in the perception of English monosyllabic forms; a short vowel is not accepted before a voiced [g], a long or short vowel is accepted before an unvoiced [k]. Bengali does not appear to be influenced by such perceptual environmental demands, where potential monomoraic forms are averted by the process of (phonetically) lengthening the vowel, irrespective of the voicing (or even the presence) of the following consonant (Roberts et al., 2014). Whether L1 Bengali speakers of L2 English are sensitive to such variation in English vowel length will be addressed in Section 5.

2.5 Interim summary and discussion

Although both Bengali and English have a PWd-adjoined prosodic representation, in Bengali it is licensed to the present and past perfect tenses rather than the simple past tense, as it is in English. There are two expected possibilities for the production of L2 English regular simple past tense by L1 Bengali speakers. For advanced levels of proficiency: (a) the PWd-adjoined representation is relicensed and the utterance is produced as target-like, and (b) at lower levels the simple past regular tense is supplied in a non-target-like PWd-internal representation. Unlike English, Bengali has no phonemic distinction in vowel length contrast (Ferguson & Chowdhury, 1960); Bengali vowel lengthening is one way to ensure a minimally bimoraic output. Bengali and English share a bimoraic minimal word requirement but differ in how the mora contributes to syllable weight (i.e., whether final consonants are moraic). Bengali has a CVV minimal word compared to English CVX. A fundamental difference between Bengali and English minimal word requirements is with respect to the distribution of syllable weight. In Bengali, syllable weight is distributed in the nucleus and not across the final consonant. The addition of a mora and subsequent vowel lengthening is applied when the output would otherwise result in a subminimal form. In English, weight is distributed across the vowel and final consonant. Bengali vowel length is not phonetically differentiated according to the voicing status of the following consonant, but in relation to fulfilling the bimoraic minimal word requirement, whereas in English there is both a perceptual and production difference in vowel length dependent upon the voicing quality of the following consonant.

A further difference is in relation to word-final consonant clusters, which are permitted in English but which are not permitted in Bengali.[10] Although a prosodic boundary exists between the stem and the inflectional morpheme in a PWd-adjoined representation, if L1-Bengali speaking learners attempt to produce the target L2 English simple past form in a PWd-internal representation, then there is a possibility that the inflectional morpheme may be omitted due to a strategy of

10. One of the strategies used to incorporate loanwords with final clusters into Bengali is the inclusion of an epenthetic vowel (e.g., *pʰilim* 'film').

cluster reduction. With these similarities and differences in mind, the question is raised as to whether minimal word requirements influence the suppliance of L2 inflectional morphology. This is considered in the context of L1 Bengali speakers of L2 English, particularly in the production of inflectional morphology which requires a PWd-adjoined representation.

3. The study: Suppliance of regular simple past inflection by L1 Bengali speakers of L2 English

Assuming that the transfer of prosodic representation transpires as delineated by the PTH and that the Bengali PWd-adjoined representation can be relicensed to a new position, the current study tests whether the suppliance of inflectional morphology on regular simple past verbs is, at least at some levels of proficiency, subject to differences in word minimality requirements when there is a mismatch between the L1 and L2.

3.1 Predictions

Allowing that L1 Bengali prosodic representation of simple past tense (PWd-internal) and the present/past perfect tense (PWd-adjoined) are both transferred, then the PWd-adjoined representation, which is required for target-like production of L2 English regular simple past tense, will be available in the interlanguage. However, if the L1 PWd-adjoined representation is not licensed to a new position, then it will not be available for the production of L2 English simple past tense. In this case, it is predicted that omission of inflection (and the instantiation of complex coda cluster reduction) will be evident across all verb types, regardless of vowel length or stem-final consonant. Subsequently, L2 English regular simple past forms will be produced via the L1 PWd-internal representation, and omission of inflection will result from phonotactic constraints. If, however, learners have access to both PWd-internal and adjoined prosodic representations in the domain of simple past inflection, it is predicted that L1 Bengali speakers of L2 English will, depending on the level of proficiency, produce inflection on regular simple past tense verbs more frequently under some conditions than others due to the moraic structure of the verb stem with respect to L1 word minimality requirements. It is also predicted that English verb stems that fulfil the L1 word minimality requirements (i.e., those with a long vowel or diphthong) will be produced without inflection at lower levels of proficiency. The reason for this is that because the stem would satisfy L1 minimality requirements, it would be produced in its bare, uninflected form. For example, the simple past form of a LV verb, such as 'clean', may surface as [kliːn], satisfying the L1 minimal word constraint, but with the omission of the required L2 inflectional morphology. On the other hand, it is predicted that English verb stems which do not meet the L1 minimality requirements (i.e., those with a short vowel) will require repair to avoid a subminimal form. It is proposed that this repair can take one of two forms: (a) the introduction of an epenthetic vowel (as a manifestation of the L1 strategy to break up consonant clusters on loanwords), providing a vowel to which a consonantal affix could then be attached, or (b) the addition of a mora and subsequent vowel lengthening, operating as in the L1 as a last course of action. For example, LV verb 'pick' may surface as [pɪkət] in the simple past, or, after vowel lengthening, as [piːkt], arguably drawing the learner's attention to tense, encouraging suffixation. The predictions are summarised as follows:

i. Inflection will be omitted across the board, regardless of verb type (SV or LV) or stem-final consonant, if learners do not have access to a relicensed PWd-adjoined prosodic representation.
ii. If learners have access to a relicensed PWd-adjoined prosodic representation, regular English verbs with a SV verb stem will (at some levels of proficiency) be produced with inflection more frequently than verbs with a LV verb stem.
iii. SV verb stems will be identified as subminimal, with potential for suffixation, and repair either by:
    a. The introduction of an epenthetic vowel.
    b. The addition of a mora and subsequent vowel lengthening.

3.2 Participants

A total of 28 L1 Bengali speakers of L2 English participated in this study, recruited in the U.K. and Bangladesh. The participants in the U.K. were recruited in the Greater Manchester area in northwest England. Proficiency levels were established using the Oxford Quick Placement Test (OQPT), and the participants were banded into four groups as follows: Beginner (n = 7), Elementary (n = 8), Intermediate (n = 8), and Advanced (n = 5). A number (but not all) of the participants were also speakers of the Sylheti dialect.[11] Whether a subject was a Sylheti speaker or not was included in the statistical analysis as a predictor to ensure that being a speaker of both Sylheti and Bengali would not influence suppliance of L2 inflectional morphology.

11. It is proposed that simple past tense is also represented in a PWd internal representation in the Sylheti dialect.

3.3 Stimuli and tasks

A total of 64 regular verbs (only those requiring a '-t' or '-d' suffix) were presented in the test stimuli, with 33 LV verb stems (Appendix A) and 31 SV verb stems (Appendix B).[12] Of the LV verb stems, there were 14 stems with an unvoiced stem-final consonant and 14 with a voiced stem-final consonant and 5 vowel-final stems. Of the SV verb stems, there were 17 unvoiced and 14 voiced stem-final consonants. An anonymous reviewer pointed out that the voiced stem-final consonant test stimuli consist only of sonorants and that the results may reflect a sonority effect rather than a voicing one. In the case of Bengali, there is no indication that some, but not other types of consonants, are moraic in word-final position. In this respect, it is unlikely that the pattern of inflection seen here is driven by a sonority effect. However, opportunity for further testing would include a more controlled elicitation task to include voiced obstruent-final stems.

In keeping with Goad and White (2006), all test stimuli were phonetically monosyllabic, and the verbs were selected in terms of stem shape and vowel length. Verb frequency was controlled for; lemma frequency values were taken from the companion website for Leech et al. (2001) to create a log frequency. This was included to account for the extreme high or low frequency values of some tokens. Frequency was included as a factor in the statistical analysis.

A semi-spontaneous question-and-answer picture card task was used to provide a realistic indication of the representation of inflectional morphology within the confines of a controlled elicitation task. The target language was elicited within a pre- and post-target vocabulary task, diverting attention away from the test language, and differentiated to accommodate variation in proficiency levels. Figure 1 (adapted from Ingham (2019, p. 132) illustrates the elicitation task for the simple past form of 'live', where the expected response was the target verb followed by a vowel initial word, such as 'lived on the third floor'. The elicitation procedure followed three main stages: (a) vocabulary naming: the names of the buildings and related features, (b) elicitation of the target verb, (c) vocabulary and student focused mini-discussion. The elicitation procedure was conducted on a one-to-one basis over two sessions for the U.K.-based participants. This procedure was replicated online for Bangladesh-based participants, using face-to-face video.

The elicited production task was recorded on a portable Roland Edirol R-09 digital recorder at a standard 44.1 KHz. The audio files were exported and compressed as WAV files, and the sound files were analysed using version 5.4.16 of Praat (Boersma & Weenink, 2015). Elicited tokens were analysed only if the target form was followed by a vowel initial word or a pause to avoid effects of coarticulation or

12. Verbs with both tense vowels and diphthongs are grouped together in the LV verb category.

Figure 1. Sample elicitation card for present and post tense of 'live'

possible loss of inflection due to the complexity of the following segmental context. Items were classified as (i) produced with the target inflection, (ii) produced in the bare form, (iii) produced with a non-target inflection either accurately or (iv) inaccurately (i.e. grammatically ill-formed), as shown in Table 2.

Table 2. Coding of elicited responses

Target	(i) Suppliance target inflection	(ii) Omission target inflection	(iii) Other response e.g., present simple in past simple context	(iv) Other response e.g., present perfect in past simple context
picked	picked	pick	picks	*has pick

Utterances that fell into categories (iii) and (iv) were grouped together as 'other responses' and removed from the statistical analysis, along with other inappropriate tense forms or verbless utterances. From a potential 1,792 items, a total of 1,155 tokens were analysed across all proficiency levels, with 582 LV verb forms and 573 SV verb forms. Only simple past tense inflected and bare verb forms were included in the analysis.

As would be expected in a semi-spontaneous elicitation task, unsolicited tokens were also recorded, again either with target inflection, in a bare form, or in another non-target 'other responses' construction. For example, it is possible that in the elicitation of target form 'lived', the unsolicited past form 'stayed' could feasibly be produced. All contextually appropriate unsolicited tokens which were (a) produced with the target inflection or (b) produced in the bare form were, therefore, entered into the statistical analysis, and coded according to vowel length (LV or SV) and final consonant in the verb stem.

4. Results

The generalized estimating equations (GEE) approach with logistic regression mixed model estimated the associations between factors and the suppliance of simple past inflectional morphology according to verb stem type. The final model is presented in Table 3.

Table 3. Suppliance of regular simple past tense by verb stem type

	Predictor	QICC	Δ QICC
BASE MODEL	proficiency level, past tense verb type, proficiency level*past tense verb type	1270.867	–
FINAL MODEL	log frequency, proficiency level, past tense verb type, college education, past tense verb type* log frequency	1248.324	22.543

The estimated means for the suppliance of inflectional morphology in the context of LV and SV regular verbs are plotted in Figure 2.[13] Although the mean difference in the suppliance of inflectional morphology between LV and SV verbs is not statistically significant, a trend across proficiency levels can be seen to show a more consistent suppliance of inflection on SV regular simple past verbs compared to LV verb stems.

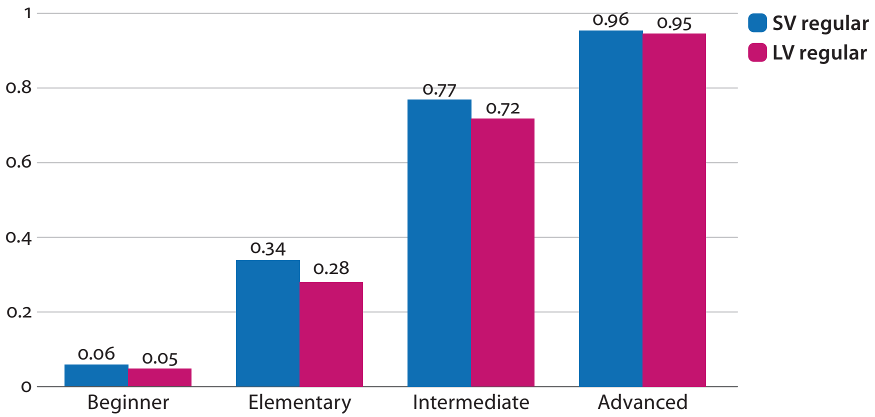


Figure 2. Estimated means: Suppliance of inflectional morphology on SV and LV regular verbs by proficiency

13. Figure 2 is adapted from Ingham (2019 p. 204).

The inference is that if learners have repaired the SV verb by forming a bimoraic minimal word according to L1 minimal word requirements, then it is more likely to prompt suffixation compared to that on LV verbs that are well-formed and will not draw attention to tense. Increased proficiency is expected to coincide with a levelling-up between suppliance of inflection on both SV and LV verbs alongside analysis of the moraic structures in the L1 and L2. At the Elementary and Intermediate proficiency levels, there is some indication of a preference for inflection on the SV verb forms. However, it should be noted that there is no evidence of epenthesis in the data set. If any repair with respect to minimality is taking place, it is not via epenthetic vowel insertion.

Because there was no statistical significance between the suppliance of inflection and verb type (LV or SV), a sub-model analysis was run to test whether learners omitted inflection across verb type with respect to stem-final consonant, to determine whether learners were able to access the target prosodic representation. However, because the interaction between proficiency and verb stem voice status was not significant, proficiency was not included in the final model and the levels were collapsed (Table 4).

Table 4. Sub-model for effect of stem-final voicing and vowel length

	Predictor	**QICC**	**Δ QICC**
BASE MODEL	proficiency level, SV & LV regular, college education, proficiency*SV & LV regular	1270.867	–
FINAL MODEL	log frequency, proficiency level, proficiency level, SV & LV regular, college education, proficiency level* SV & LV regular, verb stem voice status, SV & LV regular* verb stem voice status	923.510	6.508

The difference between unvoiced SV and unvoiced LV verbs was statistically significant (mean difference = .19, p = .011), and between the suppliance of inflection on voiced and unvoiced SV verbs (mean difference = .20, p = .0005). The estimated means for combined proficiency levels are presented in Figure 3.[14]

14. Figure 3 is adapted from Ingham (2019, p. 210).

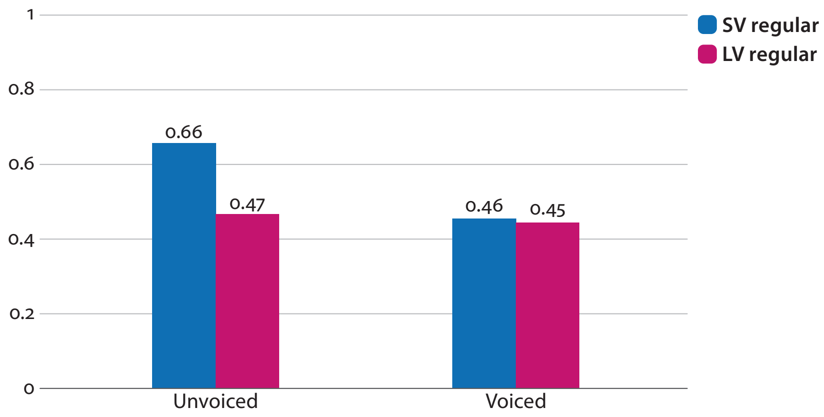


Figure 3. Estimated means: Suppliance of inflectional morphology according to stem-final voicing status for combined proficiency groups

The learners in this study appear to inflect SV verbs when the vowel precedes an unvoiced consonant more frequently than SV verbs that precede a voiced consonant or LV verbs of both types, as illustrated in Table 5, where the double check mark indicates a greater degree of suppliance of inflection.[15]

Table 5. Suppliance of inflection on voiced and unvoiced stem-final verbs

	Unvoiced stem-final C + [t]	Voiced stem-final C + [d]
SV Reg	packed √ √	grinned √
LV Reg	parked √	cleaned √

This seems to suggest that at least some learners have access to a relicensed L1 PWd-adjoined representation to form L2 English regular simple past tense. Otherwise, if L2 English regular verbs were produced in an L1 PWd-internal prosodic representation, the addition of inflection would inevitably result in a consonant cluster. Because the learners in this study do not break up clusters with an epenthetic vowel, it is possible that omission of inflection is a strategy to avoid clusters. However, as word-final clusters are avoided in Bengali and repaired with epenthesis in loanwords, it is difficult to account for this asymmetrical suppliance of inflection if learners apply the L1 PWd-internal representation in the production of L2 regular simple past verbs. In other words, it would be expected that consonant deletion would apply across all verb stems, regardless of the voicing status of the stem-final consonant (or SV/LV status). Conversely, no verb stem would be expected to retain a cluster in relation to the voicing status of the stem-final consonant.

15. Table 5 is adapted from Ingham (2019, p. 211).

A further exploratory sub-model analysis was run to determine whether the interaction between the voicing status of the stem-final segment and stem vowel length influenced the suppliance of inflectional morphology. A statistically significant mean difference was found between the means of suppliance of inflection on unvoiced SV regular stems compared to voiced SV regular stems (mean difference .20, $p = .0005$) and also between unvoiced SV regular stems and unvoiced LV regular stems (mean difference .19, $p = .011$). In order to establish whether this pattern was representative of suppliance ratios within proficiency levels, Beginner and Advanced proficiency levels were removed from the analysis. This was in response to a warning that the Hessian matrix is singular which was identified as being due to the extreme values of very low performance at Beginner level and very high performance at Advanced level. The suppliance of inflectional morphology according to stem-final voicing and vowel length for Elementary and Intermediate proficiency levels is illustrated in Figure 4.

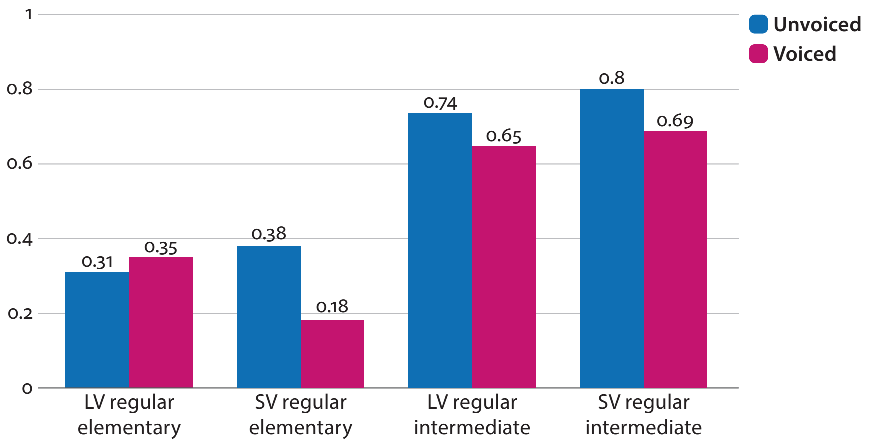


Figure 4. Estimated means: Suppliance of inflectional morphology according to stem-final voicing status for elementary and intermediate proficiency groups

Pairwise comparisons showed no statistical significance for LV regular stems in voiced or unvoiced conditions for either the Elementary proficiency level (mean difference = .04, $p = .364$) or Intermediate level (mean difference = .10, $p = .184$). However, whilst there was no statistical significance for SV regular stems in voiced or unvoiced conditions for Intermediate level learners (mean difference = .11, $p = .158$), there was a statistically significant difference for Elementary learners (mean difference = .20, $p = .0005$), indicating that inflection is supplied less consistently on verbs which are SV with voiced stem-final consonant.

4.1 Summary of results

Results indicated that in relation to prediction (i), at least some learners showed evidence of relicensing the L1 Bengali PWd-adjoined prosodic representation to a new position to produce L2 English simple past regular inflection in a target-like prosodic structure. This is because the omission of inflection was not consistent across verb types, as would be expected if regular verbs were not produced in a PWd-adjoined representation, especially considering word-final clusters are not attested in the L1 under any circumstances. With respect to prediction (ii), results initially showed a non-statistically significant trend indicating that SV verb stems were produced with inflection more frequently than LV verb stems. Further analysis of this trend revealed that when proficiency groups were combined, there was a statistically significant difference in the suppliance of inflection on SV verbs with a voiced stem compared to SV verbs with an unvoiced stem or LV verbs with an unvoiced stem. Although it was predicted that inflection would be supplied more consistently on SV verbs, it was not predicted that this would be restricted to SV verbs with unvoiced stem-final consonants. In relation to the repair of subminimal stems put forward in prediction (iii), there was no evidence of epenthesis in this data set (iii.a). Further analysis of Elementary and Intermediate proficiency levels found that at the Elementary proficiency level, SV verb stems with an unvoiced stem-final consonant were significantly more likely to be inflected than those with a voiced stem-final consonant.

5. Discussion

Once the interaction between the Elementary and Intermediate proficiency levels and verb stem voice status is taken into consideration, the picture that emerges is quite revealing. Rather than observing a spike in the suppliance of inflection on SV regular unvoiced stems, the outcome is more one of depressed suppliance of inflection on the SV regular voiced stems, a pattern that is not in evidence at the Intermediate proficiency level. The suppliance of more consistent inflection on SV verbs when the final consonant is unvoiced by Elementary level learners could be indicative of the influence of L1 minimality requirements; the shortest vowel and consonant combination is more likely to be identified as a violation of the L1 minimality requirements. This is perhaps illustrative of a sensitivity (by at least some L1 Bengali speakers of L2 English) to the influence of L2 English consonants on vowel length, with the shortest vowels generally perceived to be those before an unvoiced consonant, in turn prompting repair. That is, a transfer of the repair requirements in L1 Bengali to prevent a subminimal form in the L2. Rather than vowel insertion, which is applied to address word-final consonant clusters in loanwords, repair may arguably take the form of that in the L1 to address subminimal forms. An additional mora may therefore be invoked, and the default mechanism to fill a mora triggered, leading to subsequent lengthening of the vowel. The ensuing attention to the root form arguably instigates ‘noticing’, prompting the attachment of inflectional morphology.

By the Intermediate level, it is proposed that learners have identified the differences in moraic structure between the L1 and L2, and the SV unvoiced verb stems are no longer singled out for repair such that inflection is supplied more equivalently across regular verb types. If this intuition is on the right track, then L1 minimality requirements would be addressed according to the default setting in the L1 (i.e., with the addition of a mora and lengthening of the vowel to fill the newly made mora), as proposed in prediction (iii.b). This result may also support the claim, albeit indirectly, that inflection is attached in a target-like PWd-adjoined representation. However, because it was not possible to make per-participant comparative measurements of vowel length from this data set, this observation remains speculative and constitutes an area for further study, requiring a more controlled data set than is available through a semi-spontaneous elicitation procedure.

This explanation does not account, however, for the fact that Elementary level learners produce inflected LV verb stems in both stem-final voicing conditions at a similar rate to the SV unvoiced stem condition. A possible explanation for this result may again be found in a comparison of the written and spoken forms of Bengali. Whilst a long-short vowel form is still in evidence in Bengali orthography, marked with *hrashya* (‘short’) and *dirgha* (‘long’) symbols, this distinction is not phonemic (Boyle David, 2015). In terms of the pronunciation of the vowel, it has been suggested that phonetic vowel lengthening in Bengali is more accurately described as a ‘half-long’ vowel /ˑ/ in comparison to the Bengali short /i/ vowel, and compared to English tense /iː / and lax /ɪ/ vowels. If this is the case, then it could well be that the L2 English SV verb stem with a voiced stem-final consonant is a better example of L1 phonetic lengthening, and most closely satisfies the L1 minimality requirements; resulting in a greater proportion of ‘well-formed’ uninflected SV voiced verbs. An interpretation of this kind was also raised by an anonymous reviewer, suggesting that once the L1 minimality requirement has been achieved, an argument for lack of inflection (as in LV-SV verb stems) can still stand because the form is not subminimal and does not require extra attention.

Finally, it would appear, speculatively, at least, that it is possible that the effects of L1 transfer of prosodic structure below the level of the prosodic word, particularly with respect to word minimality requirements, may provide another layer of phonological complexity in the L2 acquisition process. For some learner backgrounds, this may require a readjustment of L1 requirements for well-formedness before inflectional morphology may be attached or incorporated into the prosodic word. It may well be that, at lower levels of proficiency, the L1 template for well-formedness must first be applied to L2 prosodic words, potentially generating an asymmetric suppliance of morphology towards the stems which require repair. With increased proficiency, the learner may be able to learn to apply L2 word minimality requirements (and override L1 word minimality requirements) and supply inflectional morphology more equally across verb stem conditions. Further investigation is required, both within and across L1 groups, in order to more fully explore the potential role of transfer of L1 prosodic constituents below the level of the prosodic word, as well as at the level of the word, in the suppliance of L2 inflectional morphology.

References

Bayer, J., & Lahiri, A. (1990). Bengali emphatic clitics in the lexicon-syntax interface. In W. U. Dressler, H. Luschützky, O. Pfeiffer, & J. Rennison (Eds.), *Contemporary morphology* (pp. 3–16). Mouton de Gruyter. https://doi.org/10.1515/9783110874082.3

Boersma, P., & Weenink, D. (2015). Praat: Doing phonetics by computer [computer program]. Version 5.4.16. Retrieved on 28 February 2022 from http://www.praat.org/

Boyle David, A. (2015). *Descriptive grammar of Bangla*. De Gruyter Mouton. https://doi.org/10.1515/9781614512295

Broselow, E., & Park, H.-B. (1995). Mora conservation in second language prosody. In J. Archibald (Ed.), *Phonological acquisition and phonological theory* (pp. 151–168). Lawrence Erlbaum Associates.

Bybee, J. (2001). *Phonology and language use*. Cambridge University Press. https://doi.org/10.1017/CBO9780511612886

Cabrelli Amaro, J., Campos-Dintrans, G., & Rothman, J. (2018). The role of L1 phonology in L2 morphological production. L2 English past tense production by L1 Spanish, Mandarin, and Japanese speakers. *Studies in Second Language Acquisition*, 40, 503–527. https://doi.org/10.1017/S0272263117000122

Ellis, N. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24, 143–188. https://doi.org/10.1017/S0272263102002024

Ellis, N. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 63–103). Blackwell. https://doi.org/10.1002/9780470756492.ch4

Ewen, C. J., & van der Hulst, H. (2001). *The phonological structure of words*. Cambridge University Press.

Ferguson, C. A., & Chowdhury, M. (1960). The phonemes of Bengali. *Linguistic Society of America*, 36 (1), 22–59. https://doi.org/10.2307/410622

Fitzpatrick-Cole, J. (1990). The minimal word in Bengali. In A. L. Halpern (Ed.), *The proceedings of the ninth west coast conference on formal linguistics* (pp. 157–170). CSLI.

Fitzpatrick-Cole, J. (1996). Reduplication meets the phonological phrase in Bengali. *The Linguistic Review*, 13, 305–356. https://doi.org/10.1515/tlir.1996.13.3-4.305

Goad, H., & White, L. (2004). Ultimate attainment of L2 inflection: Effects of L1 prosodic structure. In S. Foster-Cohen, M. Sharwood Smith, A. Sorace, & M. Ota (Eds.), *Eurosla Yearbook* (Vol. 4, pp. 119–145). John Benjamins. https://doi.org/10.1075/eurosla.4.07goa

Goad, H., & White, L. (2006). Ultimate attainment in interlanguage grammars: A prosodic approach. *Second Language Research*, 22(3), 243–268. https://doi.org/10.1191/0267658306sr268oa

Goad, H., & White, L. (2008). Prosodic structure and the representation of L2 functional morphology. A nativist approach. *Lingua*, 118 (4), 577–594. https://doi.org/10.1016/j.lingua.2007.01.008

Goad, H., & White, L. (2019). Prosodic effects on L2 grammars. *Linguistic Approaches to Bilingualism*, 9, 1–41. https://doi.org/10.1075/lab.19043.goa

Goad, H., White, L., & Steele, J. (2003). Missing inflection in L2 acquisition: Defective syntax or L1-constrained prosodic representations? *The Canadian Journal of Linguistics/La Revue Canadienne De Linguistique*, 48(3/4), 243–263. https://doi.org/10.1353/cjl.2004.0027

Gordon, M. K. (2013). *Syllable weight: Phonetics, phonology, typology*. Routledge.

Hawkins, R., & Chan, C. Y. (1997). The partial availability of universal grammar in second language acquisition: The 'failed functional features hypothesis'. *Second Language Research*, 13, 187–226. https://doi.org/10.1191/026765897671476153

Hawkins, R., & Hattori, H. (2006). Interpretation of English multiple wh-questions by Japanese speakers: A missing uninterpretable feature account. *Second Language Research*, 22, 269–301. https://doi.org/10.1191/0267658306sr269oa

Hawkin, R., & Liszka, S. (2003). Locating the source of defective past tense marking in advanced L2 English speakers. In R. van Hout, A. Hulk, F. Kuiken, & R. Towell (Eds.), *The interface between syntax and lexicon in second language acquisition* (pp. 21–44). John Benjamins. https://doi.org/10.1075/lald.30.03haw

Hayes, B. (1989). Compensatory lengthening in moraic phonology. *Linguistic Inquiry*, 20(2), 253–306.

Hyman, L. (1985). *A theory of phonological weight*. Foris. https://doi.org/10.1515/9783110854794

Ingham, J. (2019). L2 inflectional morphology and prosody: The case of L1 Bengali speakers of L2 English (Unpublished doctoral dissertation). University of Sheffield.

Inkelas, S. (1991). *Prosodic constituency in the lexicon*. Garland.

Kar, S. (2009). The syllable structure of Bangla in Optimality Theory and its application to the analysis of verbal inflectional paradigms in Distributed Morphology (Unpublished doctoral dissertation). University of Tübingen.

Klaiman, M. H. (1990). Bengali. In B. Comrie (Ed.), *The world's major languages* (pp. 490–513). Oxford University Press.

Lahiri, A. (2000). Hierarchical restructuring in the creation of verbal morphology in Bengali and Germanic: Evidence from phonology. In A. Lahiri (Ed.), *Analogy, levelling and markedness: Principles of change in phonology and morphology* (pp. 71–123). Mouton. https://doi.org/10.1515/9783110808933.71

Lardiere, D. (2008). Feature assembly in second language acquisition. In J. M. Liceras, H. Zobl, & H. Goodluck (Eds.), *The role of formal features in second language acquisition* (pp. 106–140). Lawrence Erlbaum Associates.

Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English: based on the British National Corpus*. Longman Available at: https://ucrel.lancs.ac.uk/bnc-freq/(April 2022).
Maddieson, I. (1997). Phonetic universals. In W. J. Hardcastle & J. Laver (Eds.), *The handbook of phonetic sciences* (pp. 619–639). Blackwell.
Mazumdar, B. C. (1920). *The history of the Bengali language*. University of Calcutta.
McCarthy, J. J., & Prince, A. (1996). Prosodic morphology 1986. *Linguistics Department Faculty Publication Series*, 13. Retrieved on 28 February 2022 from https://scholarworks.umass.edu/linguist_faculty_pubs/13
Nespor, M., & Vogel, I. (2007). *Prosodic phonology* (2nd ed.). Mouton de Gruyter. https://doi.org/10.1515/9783110977790
Prévost, P., & White, L. (2000). Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research*, 16, 103–133. https://doi.org/10.1191/026765800677556046
Ray, P. S., Hai, M. A., & Ray, L. (1966). *Bengali language handbook*. Center for Applied Linguistics.
Roberts, A., Kotzor, S., Wetterlin, A., & Lahiri, A. (2014). Cross-linguistic studies of allophonic durational differences. Retrieved on 28 September 2021 from http://brainlab.clp.ox.ac.uk/sites/default/files/RobertsSNL2014poster.pdf
Selkirk, E. (1997). The prosodic structure of function words. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 187–213). Lawrence Erlbaum Associates.
Tsimpli, I. M., & Dimitrakopoulou, M. (2007). The interpretability hypothesis: Evidence from wh-interrogatives in second language acquisition. *Second Language Research*, 23(2), 215–242. https://doi.org/10.1177/0267658307076546

Appendix A. LV verbs (including diphthong) *n* = 33

Token	**Voicing status/ lemma freq.**
	V = voiced U=unvoiced
beeped	U 0
blamed	V 44
boiled	V 11
chased	U 21
cleaned	V 40
faced	U 113
framed	V 11
gained	V 89
joined	V 174
leaked	U 0
marked	U 77
moved	V 627
parked	U 19

Token	Voicing status/ lemma freq.
peeped	U 0
phoned	V 31
poured	V (vowel) 37
raced	U 36
sailed	V 27
saved	V 118
sealed	V 15
shaped	U 21
showed	V (vowel) 598
smiled	V 112
smoked	U 29
stayed	V (vowel) 183
stirred	V (vowel) 25
stroked	U 13
tied	diph (vowel) 42
timed	V 13
warmed	V 15
wiped	U 24
worked	U 646

Appendix B. SV verbs (including diphthong) $n = 31$

Token	Voicing status/ lemma freq.
	V = voiced
	U=unvoiced
asked	U 610
banned	V 28
blocked	U 27
chopped	U 10
cooked	U 37
cracked	U 16
crossed	U 70
drilled	V 11
dropped	U 107
filled	V 110
grinned	V 24
hummed	V 0
jammed	V 0
kissed	U 36

Token	Voicing status/ lemma freq.
licked	U 0
lived	V 329
looked	U 1151
missed	U 108
packed	U 0
picked	U 150
pinned	V 13
planned	V 141
pressed	U 73
pulled	V 140
slammed	V 14
slipped	U 51
trimmed	V 0
washed	U 49
watched	U 202
wrapped	U 24
yelled	V 12

Impact of previously learned languages

The role of L1 Norwegian and L2 English in the acquisition of verb placement in L3 German

Anne Dahl[1], Kjersti Faldet Listhaug[1] and Guro Busterud[2]
[1] NTNU Norwegian University of Science and Technology /
[2] University of Oslo

Focusing on the much-debated question of transfer from previously learned languages in L3 acquisition, we investigate acquisition of finite verb placement in L3 German. Participants are L1 Norwegian high-school students with L2 English, in years 1, 2, 4, and 5 of German instruction. Norwegian and German have V2 word order, while English does not. Participants completed acceptability judgments in L3 German and L2 English. Results show no clear preference for either V2 or non-V2 in German in the earliest learners, but later development towards target-like intuitions. Target-like L2 English judgments do not seem to be associated with more transfer from L2 to L3 of a given structure, and higher L2 proficiency does not predict more L2 transfer to L3.

Keywords: word order, verb placement, V2, second language, third language, acquisition, transfer Norwegian, English, German

1. Introduction

In research on additional language (L2, L3, Ln) acquisition, a central question is that of the influence of previously learned language(s). While such influence is certainly well documented, there is no consensus on exactly how and to what extent previous linguistic experience influences the acquisition of subsequent languages. One previously studied phenomenon is the acquisition of verb placement when the languages in question differ in whether they are verb second (V2) or non-V2. The present study aims to shed more light on questions of transfer of verb placement using a new language combination, investigating the L3 acquisition of German by L1 Norwegian speakers with L2 English. In this specific language combination, transfer from L1 Norwegian would lead to target-like verb placement in L3 German, while transfer from L2 English would lead to non-target-like verb placement.

https://doi.org/10.1075/lald.67.08dah

In our study, all participants were tested in both L2 and L3, and proficiency in both languages was considered. Participants were in their first, second, fourth or fifth year of learning German and had learned English for 11–12 years. We investigate transfer, L3 developmental trajectories, and the influence of L2 proficiency in the acquisition of L3 verb placement.

2. Theoretical background

2.1 Transfer in L2 and L3 word order

The existence of L1 transfer in L2 acquisition is well established, and a number of studies have found transfer of non-target verb movement. A finding relevant to our study is that young Norwegians transfer the V2 word order of Norwegian main clauses to L2 English (Westergaard, 2002, 2003). Westergaard finds more prominent and lasting transfer for verb placement with adverbials than for topicalizations, and more transfer with auxiliaries than lexical verbs.

While questions of transfer processes in L2 address transfer from the L1, the acquisition of a third language introduces a number of new variables. The question is no longer limited to the extent to which the system of the L1 is transferred into the new language, but also which of the previously acquired language system(s), L1 and/or L2, may transfer. A general distinction can be made between models of L3 acquisition assuming transfer from only one of the previously acquired languages, and those assuming possible transfer from both L1 and L2. Among the former, some accounts assume that the L1 transfers in L3 acquisition, such as Hermas' (2010) study of verb raising in L3 French. Jin (2009) concluded similarly in her study of the acquisition of obligatory objects in L3 Norwegian. Other studies have found transfer from the L2 only, and when it comes to word order and V2, the opposite pattern compared to Westergaard (2002, 2003) has been observed. Some studies (Bohnacker, 2006; Håkansson, Pienemann, & Sayehli, 2002) have found that Swedish L1 speakers acquiring L3 German tended to use SVO word order without inversion rather than V2, which is the grammatical word order both in Swedish and German. Håkansson et al. (2002) do not attribute this to transfer from L2 English, but rather see it as a general feature of additional language acquisition. However, Bonacker (2006) failed to find non-V2 structures in a small group of L2 learners of German with no prior knowledge of English, indicating that L2 transfer may indeed be the cause of this observed phenomenon. Bardel and Falk (2007) similarly found evidence of transfer from L2 for placement of negation in learners with various language backgrounds acquiring an L3 with V2 (Dutch or Swedish, respectively), as did Falk and Bardel (2011) for placement of object pronouns in L3 acquirers of German with English or French as their L1 or L2, respectively. Their conclusion is that L2 has a privileged status for transfer, at least with relatively low L2 proficiency. Bardel and Falk (2012) do, however, hypothesize that this privileged status of the L2 may diminish when proficiency in L2 is high enough to approach that of the L1.

There are also studies whose findings indicate that language status (L1 vs L2) is not the determining factor for transfer in the initial stages, but rather similarity between languages, and that a previously acquired language that is (psycho-)typologically closer to the L3 is more likely to transfer. Many such studies have looked at Romance languages, often in combination with English (see Rothman, 2015), where at least one language pair is closely related, and defining which pair is most closely related is relatively straightforward. As such, they differ from the present study, where all three languages in question are Germanic.

Other studies have failed to find transfer from only one previously learned language, indicating that both L1 and L2 may transfer during the same acquisition process. Flynn, Foley, and Vinnitskaya (2004) interpret their results in a study of relative clause acquisition in speakers of L1 Kazakh/L2 Russian/L3 English as evidence that both previously learned languages can have a facilitative effect on the L3. Westergaard, Mitrofanova, Mykhaylyk, and Rodina (2017) found both facilitative and non-facilitative transfer from both languages for English word order in Russian-Norwegian bilinguals and argue for transfer on a property-by-property basis. Slabakova (2017) reviews a number of studies and similarly concludes that transfer cannot be wholesale, and that it can be facilitative or non-facilitative and come from either L1 or L2. Westergaard (2021) argues that rather than full transfer of a previously learned language, there might be full transfer potential of all prior languages in multilingual acquisition.

Highly relevant to the present study is a series of recent studies of word order in L3 French acquisition in speakers of L1 Dutch/L2 English (Stadt, Hulk, & Sleeman, 2016, 2018a, 2018b, 2020). They investigate on the one hand topicalized structures, which are similar in L2 English and L3 French, both non-V2 SVO languages, but different in L1 Dutch, which is V2. On the other hand, they study placement of short sentence adverbials, which is similar in Dutch and French, which both have verb raising, but different in L2 English where lexical verbs do not raise. They find that topicalizations are in place before sentence adverbial placement. Taken together, these studies find transfer from both L1 and L2, and that transfer from L1 is associated with the earliest stages of acquisition and is less pronounced later in the acquisition process. Exposure to L2 English, either in the form of immersion education or by virtue of more years of English in school, is associated with more L2 transfer. These findings do not immediately seem to support those studies arguing for wholesale transfer from one language. Furthermore, they indicate that the L2 does not necessarily have a privileged status for transfer in L3 word order acquisition. Finally, they do not support the suggestion in Bardel and Falk (2012) that high proficiency in L2 may lead to a weaker role for the L2 as a transfer source.

The above findings show no consensus on the source or extent of transfer from L1/L2 in L3 acquisition, nor of whether there is wholesale transfer from one language across all structures. It is unclear whether we can expect the same language to be most eligible for transfer at different stages of L3 development, since theories of wholesale transfer typically focus on the initial stages, while studies finding transfer from both languages may also take later stages into account. The role of L2 competence is also unresolved, i.e., whether specific structures must be "in place" to be eligible for transfer, and whether general L2 proficiency impacts transfer from L2. To address these unresolved questions, this study investigates developmental trajectories of lexical verb placement in L3 German. We test learners' competence of the same structures in L2 and L3, and investigate the relationship between L2 competence and judgments in L3.

2.2 Verb placement in Norwegian, English and German

The languages in question in the present study, i.e., Norwegian, English and German, vary with respect to both underlying word order and placement of finite verbs in main clauses. However, these differences are not visible in all sentence types. In the following, we discuss placement of the verb as the second (V2) or third (V3) constituent in the two sentence types included in our experiment. Our study focuses on two different types of declarative main clauses with lexical verbs: (1) main clauses with a topicalized adverbial, and (2) subject-initial main clauses with sentence adverbials of the type *often, always,* and *rarely*, see Examples (1) and (2) below.

(1) a. [$_{CP}$ Hver morgen *tar* [$_{IP}$ Marianne ~~tar~~ [$_{VP}$ ~~tar~~ bussen]]]. (Top,*V*,SU)
Every morning takes Marianne bus-the
b. [$_{CP}$ Every morning [$_{IP}$ Marianne [$_{VP}$ *takes* the bus]]]. (Top,SU,*V*)
c. [$_{CP}$ Jeden Morgen *nimmt* [$_{IP}$ Marianne ~~nimmt~~ [$_{VP}$
Every morning takes Marianne
den Bus ~~nimmt~~]]]. (Top,*V*,SU)
the bus.

(2) a. [$_{CP}$ Peter *spiser* [$_{IP}$ ~~spiser~~ [$_{VP}$ ofte ~~spiser~~ sushi.]]]. (*V*,SA)
Peter eats often sushi.
b. [$_{CP}$ Peter [$_{IP}$ [$_{VP}$ often *eats* sushi]]]. (SA,*V*)
c. [$_{CP}$ Peter *isst* [$_{IP}$ ~~isst~~ [$_{VP}$ oft Sushi ~~isst~~]]]. (*V*, SA)
Peter eats often sushi.

Norwegian has underlying SVO word order but obligatory movement of the finite verb via I/T to C in all main clauses (see Holmberg & Platzack, 1995; Roberts, 2001), resulting in V2 order both for topicalized constructions (1a), and for sentences with sentence adverbials (2a).[1] German is SOV[2] and has obligatory movement of finite verbs via I/T to C in main clauses, resulting in V2 in topicalized structures (1c) and with sentence adverbials (2c). English is also SVO; however, lexical verbs remain in situ, whereas auxiliaries move to I/T. English is often referred to as so-called residual V2 (Rizzi, 1996) due to subject-auxiliary inversion in *yes/no*-questions and *wh*-questions. For the structures in question in the present study, with lexical verbs only, English is clearly non-V2. In the topicalized construction in (1b), the verb follows the subject, whereas the verb follows the sentence adverbial in (2b).

3. Research questions and predictions

3.1 Research questions

For acquisition of verb placement in L3, there is contradictory evidence from previous research. For example, Håkansson et al. (2002), Bohnacker (2006), and Bardel and Falk (2007) failed to find L1 transfer of V2 from L1 to L3, when L1 and L3 have V2 word order. On the other hand, Stadt et al. (e.g., 2018a, 2020) found transfer of V2 form L1 Dutch to L3 French. It is thus unclear under which circumstances V2 may transfer, and whether L2 English impedes transfer of V2. The main purpose of this study is to further explore this issue. Furthermore, we ask whether there is evidence that one prior language is preferred as the source of transfer, which may indicate wholesale transfer at early stages of acquisition. Since previous research (e.g., Stadt et al., 2018a) has found that source language for transfer may change over the course of acquisition, we also look at developmental trajectories.

The present study sought to answer the following research questions, investigating verb placement in L1 Norwegian learners of L3 German with English as L2:

---

1. Sentence adverbials (SA) are commonly analyzed as adjoined to the specifier of the VP (Pollock, 1989). For Norwegian, Åfarli and Eide (2003) argue that SA are adjoined to IP/TP. Whether SA is adjoined to VP or IP/TP is not relevant for this study, and we do not discuss this further.

2. German SOV order is visible in main clauses with periphrastic verb constructions, where the object precedes the verb resulting in a separation of the finite and non-finite verb: *Peter hat oft Fisch gegessen*. SOV order is also visible in subordinate clauses, where the finite verb follows the non-finite verb: *Ich weiß, dass Peter Fisch gegessen hat.*

1. Is there evidence of L1 and/or L2 transfer to L3 in Norwegian L3 learners of German?
2. If so, does transfer occur to the same extent in sentences with topicalizations and in sentences with sentence adverbials?
3. Does mastery of the relevant structure in L2 matter for whether it is eligible for transfer?
4. Does proficiency in L2 influence transfer in L3; and if so, does higher proficiency predict more or less evidence of transfer from L2?
5. What do L3 developmental trajectories look like, and specifically: Does evidence of L2 transfer become more or less pronounced with increasing L3 proficiency?

3.2 Predictions

For our participants, transfer from L1 Norwegian in the sentences in question would lead to target-like judgments, while transfer from L2 English would mean non-target-like judgments.

Based on previous findings on similar language combinations (e.g., Bardel & Falk, 2007; Bohnacker, 2006; Falk & Bardel, 2011; Håkansson et al., 2002), we expect to find evidence of transfer of non-V2 from L2 English in L3 German (RQ1). If there is wholesale transfer from either L1 or L2 (see Bardel & Falk, 2007; Rothman, 2015), the prediction is that transfer should occur to the same extent in both sentence types in the earliest learners, to the extent that both structures are in place in L2 (RQ2).

The basis for transfer from L2 is the learner's mental grammar of L2 English, which may differ from the grammar of L1 English speakers. We would expect L2 transfer to be more robust for structures that are solidly in place in L2 compared to those that are not (RQ 3). Based on the performance for L2 English students up to year 7 in Westergaard (2002, 2003), we expect English word order generally to be in place among our participants. However, previous studies (Stadt et al., 2018b; Westergaard, 2002, 2003) indicate that we might see more consistent transfer of V3 in sentences with topicalizations than with sentence adverbials. Such findings would be in favor of a property-by-property view of transfer (see Slabakova, 2017; Westergaard et al., 2017; Westergaard, 2021).

RQ 4 asks whether L2 proficiency affects L2 transfer. According to Bardel & Falk (2012) high proficiency in L2 may entail a less privileged role for the L2 as a transfer source. However, this prediction is in conflict with findings from Stadt et al. (2016, 2018b), who found increased L2 transfer with increased L2 proficiency. Thus, we expect some variability in the learner group regarding what transfers, depending on proficiency in L2 English, although we do not have a clear prediction for the direction of this relationship.

At higher L3 proficiency levels, we expect any effect of L2 transfer to decrease (RQ 5). Although impossible to say for certain, this may be a result of learning/converging on the target rather than more pronounced L1 transfer.

4. Method

4.1 Participants

Participants were 154 students (56 female), age 16–17, in their first and second year of upper secondary school, corresponding to years 11 and 12 of school, respectively. They were in their first (*n* = 18), second (*n* = 15), fourth (*n* = 70), or fifth (*n* = 51) year of learning German.[3] All were native speakers of Norwegian, with English as their L2. All participants had started learning English in year 1, at age six. English is compulsory until year 11, and some continue studying it also in year 12. This means that the participants in this study all had English instruction in school for at least close to 11 years. In terms of hours of instruction in the L3, those in year 1 of German had roughly 80 hours, those in year 2 roughly 190 hours, and those in years 4 and 5 roughly 310 and 420 hours of German teaching, respectively. As an artefact of the Norwegian school system, many of those who had started learning German in year 11 (age 15/16) would have learned another foreign language from year 8 (age 13) and switched upon starting upper secondary school (see footnote 5).

Originally, 258 students from two larger cities in Norway participated.[4] We excluded participants indicating L1s other than Norwegian (both 2L1 and L1 = non-Norwegian), and participants reporting that they spoke another language regularly with close family members. We also excluded participants who indicated having learned a different L3 than German (at school) before learning German, provided they also indicated higher proficiency in the other foreign language than in German. Remaining participants reported very low competence in the other

3. In Norway, foreign language teaching starts in either year 8, i.e., the first year of lower secondary education, or year 11, i.e., the first year of upper secondary education. Foreign languages are not regularly taught in year 13. Since we tested students in upper secondary education (years 11 and 12), no participants in their third year of foreign language instruction (i.e., year 10) were included.

4. Because reports exist of optional V3 in Norwegian urban multi-ethnolects (Opsahl, 2009), schools with particularly multiethnic student populations were avoided.

language.[5] Furthermore, only participants who completed both the German and the English version of the experiment were included in the analyses. As a result, 104 participants were excluded from the study.

Participants reported their last term grades in L2 English and L3 German. Grading in Norway uses a scale of 1–6, where 2 is pass and 6 is excellent. Furthermore, participants self-rated their level in L2 English and L3 German on a scale of 1–6 with short descriptions of proficiency levels based on the Common European Framework of Reference for Languages (CEFR), where level 1 matched CEFR level A1, level 2 matched level A2, etc. Table 1 shows mean self-rated levels and mean grades per experimental group, as well as gender distribution.

Table 1. Self-rated level and grade in L2 English and L3 German per experimental group

	Year 1	**Year 2**	**Year 4**	**Year 5**
n (female/male)	**18 (10/8)**	**15 (2/13)**	**70 (25/44)**	**51 (18/32)**
L2 English self-rated level	4.7	4.8	5.2	5.1
L2 English grade	4.1	4.2	4.5	4.7
L3 German self-rated level	2.4	2.2	2.9	2.9
L3 German grade	4.7	3.0	4.0	3.8

The study was registered with and recommended by the Norwegian Centre for Research Data (NSD).

4.2 Materials and procedure

The experimental method consisted of acceptability judgment tasks (AJT) with a 1–4-point Likert scale. Points in the scale were not labelled; rather, each end point was indicated with an emoji (smiley with thumbs up/down). Participants were instructed to rate how 'good' they thought the sentence was based on their own intuitions. The participants were tested both in L3 German and in L2 English.

Each test contained a total of 48 items, 24 of which were targets: Six sentences with topicalizations (see Example (1) above) where the verb appeared in second position (V2) and six where it appeared in third position (V3), and six sentences with a sentence adverbial such as *immer/always* (see Example (2) above) where

5. Having started another L3 in year 8 does not necessarily entail three years of instruction before switching to German, as some may have given up the previous language after a short time. Too many of our participants reported rudimentary knowledge of another foreign language, instructed or not, for us to be able to exclude all participants with any such exposure.

the verb preceded the adverbial and six where it followed it. Additionally, the tests contained 12 grammatical and 12 ungrammatical filler sentences.

The German and the English versions of the test were not identical, but sentences in the two versions were compared across all three languages (German, English, Norwegian) during test development to ensure differences were limited to vocabulary and did not involve structural complexity or other potentially relevant factors.

On the lexical level, sentences were kept simple and used basic vocabulary which, based on consultations with textbooks and teachers, could be expected to be familiar to beginners. Structurally, only main clauses were included, and thus differences in underlying word order between Norwegian and German (i.e., SVO vs SOV) were not overt. All sentences contained lexical verbs in the present tense with habitual meaning in order to avoid aspectual differences between languages, and to avoid complex verb forms which may be beyond the participants' level of German. Topicalized elements were always adverbials compatible with habitual meaning, avoiding unnatural topicalizations in either language.[6] The distribution of different subject types (NP, pronominal, singular/plural etc.) was the same across conditions, and in both tests.

The fillers consisted of two types; grammatical and ungrammatical subject-initial declaratives without sentence adverbials, and grammatical and ungrammatical wh-questions (expressions typically taught early, such as *Wie alt bist du? 'How old are you?'*). Ungrammatical declaratives typically lacked an argument or the verb (e.g., *Ole und seine Freunde Playstation – 'Ole and his friends Playstation'*), whereas ungrammatical wh-sentences had deviant word order (e.g., *Wie du alt bist? 'How you old are?'*). Consequently, some filler sentences were probably more obviously grammatical/ ungrammatical than the target sentences.

Both tasks were administered on paper during school hours, the German AJT approximately two weeks before the English version. Instructions were given in Norwegian, and participants were told that they could ask for the meaning of individual words if necessary. The German version of the test took approximately 20 minutes, the English version a little less. Since test sentences in the two tests were very similar, the English test was relatively easier for participants, and ceiling effects could be expected. Participants also provided information about language background, proficiency, and any diagnoses that may affect language acquisition.

6. To avoid heavy topicalizations, topicalized elements were phrases of no more than three words.

5. Results

Results from the L3 German test, particularly relevant to RQs 1,2 and 5 are presented first. Subsequently, results for L2 English are presented with RQs 3 and 4 in mind. Data were analyzed using Tibco Statistica 13.5.0.

5.1 L3 German

We first look at the L3 German filler sentences in order to establish whether participants distinguished between grammatical and ungrammatical sentences in L3 in the task. Table 2 shows mean scores and standard deviations for L3 German fillers per experimental group.

Table 2. Mean scores and standard deviations for filler sentences in L3 German per experimental group

Years of German	n	Grammatical				Ungrammatical			
		Declaratives		Wh-questions		Declaratives		Wh-questions	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	18	3.10	0.40	3.19	0.37	2.25	0.48	2.25	0.61
2	15	2.99	0.63	3.06	0.61	2.27	0.56	2.21	0.60
4	70	3.38	0.45	3.42	0.43	1.88	0.42	1.59	0.49
5	51	3.26	0.49	3.41	0.46	1.98	0.48	1.56	0.52

Fillers were compared using a 2 (grammaticality: grammatical vs ungrammatical) x 2 (sentence type; declarative vs Wh- sentence) repeated measures ANOVA with years of German instruction as the grouping variable. Results showed a main effect of grammaticality (F(1,150) = 349.16, $p < .0001$, $\eta_p^2 = .7$) where grammatical sentences received higher scores (mean score 3.23, $SE = 0.06$) than ungrammatical sentences (mean score 1.99, $SE = 0.06$). This effect was modulated by group, where students in years 4 and 5 discriminated more clearly between grammatical and ungrammatical sentences than students in years 1 and 2; see Figure 1. There was also an interaction between grammaticality and sentence type, where ungrammatical wh-questions received lower scores than ungrammatical declaratives. This effect was in turn modulated by group (F(3,150) = 3.47, $p = .018$, $\eta_p^2 = .7$), as only students in years 4 and 5 made this distinction.

The results showed development towards target-like judgments of filler sentences. While students in years 4 and 5 most clearly distinguished between grammatical and ungrammatical fillers, also students in years 1 and 2 showed a significant preference for grammatical over ungrammatical fillers. We take this

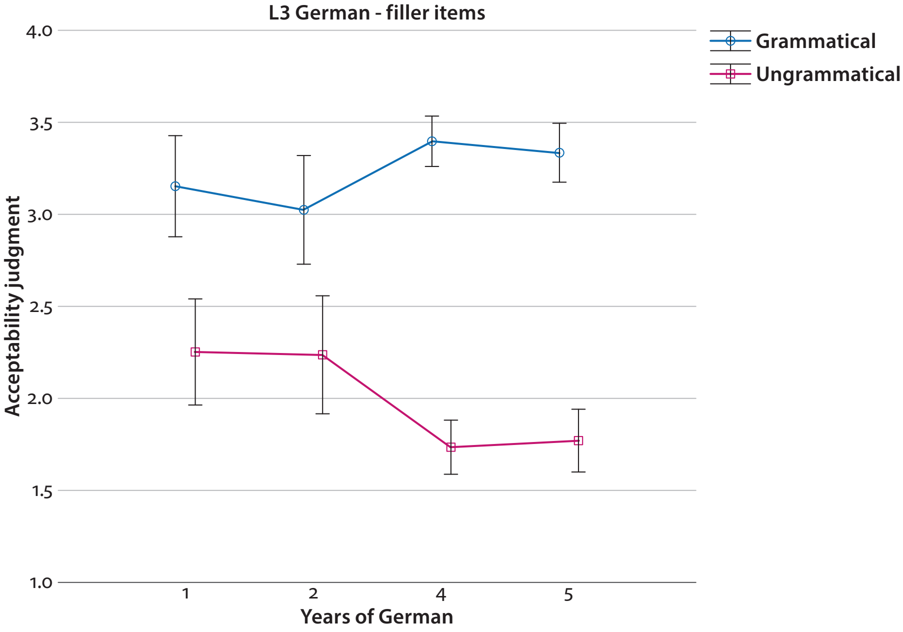


Figure 1. Two-way interaction of grammaticality and year of L3 German for filler sentences
Note: Vertical bars denote 95% confidence intervals.

to indicate that the acceptability judgment task validly measured the participants' linguistic intuitions in L3.

Next, we present judgments on the German target sentences. Table 3 shows mean scores and standard deviations for L3 German target sentences per experimental group. We find fewer clear judgments than with the fillers, with means generally hovering around 2.5.

Table 3. Mean scores and standard deviations for target items in L3 German per experimental group

Years of German	n	V2				V3			
		Topicalization		Sentence adverbial		Topicalization		Sentence adverbial	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	18	2.48	0.41	2.67	0.39	2.63	0.53	2.69	0.25
2	15	2.62	0.70	2.77	0.46	2.77	0.67	2.63	0.38
4	70	3.07	0.56	2.95	0.53	2.55	0.50	2.34	0.50
5	51	2.90	0.46	2.91	0.48	2.59	0.59	2.39	0.54

Results from the L3 German test were compared in a 2 (verb placement; V2 vs V3) x 2 (Sentence type; topicalization vs Sentence adverbial) repeated measures ANOVA with years of German instruction as the grouping variable. Results showed a main effect of verb placement ($F(1, 150) = 16.59$, $p < .0001$, $\eta_p^2 = .10$), where sentences with V2 received higher acceptability scores (mean score = 2.8, *SE* = 0.6) than V3 sentences (mean score 2.6, *SE* = 0.6). This effect was modulated by group ($F(3, 150) = 9.26$, $p < .0001$, $\eta_p^2 = .16$), see Figure 2.

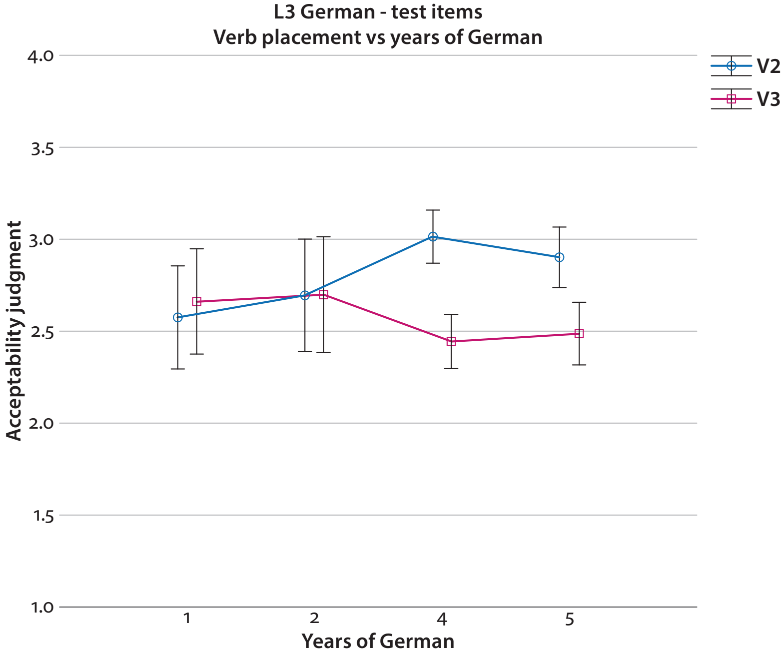


Figure 2. Interaction effect of verb placement and participant group in test items in L3 German

Note: Vertical bars denote 95% confidence intervals.

To break down the interaction between verb placement and group, Tukey HSD post-hocs were conducted. They revealed no significant difference in judgments on V2 and V3 for participants in years 1 and 2, with mean scores on both structures around the middle of the scale, i.e., no indication of a preference for one over the other. For participants in years 4 and 5, however, sentences with V2 received significantly higher scores (mean score year 4 3.05, *SE* = 0.07 and year 5 2.90, *SE* = 0.08) than sentences withV3 (mean score year 4 2.45, *SE* = 0.07, $p < .0001$, and year 5 2.49, *SE* = 0.09, $p < .0001$). Thus, we see no evidence of full transfer of either L1 or L2 in the early learners. In years 4 and 5, the results indicate a preference for grammatical V2 sentences over ungrammatical V3 sentences.

Furthermore, the results showed an interaction effect of verb placement and sentence type, i.e., V2 vs V3 and topicalization vs sentence adverbials, ($F(1, 150) = 6.21$, $p = .014$, $\eta_p^2 = .04$), see Figure 3. There was no three-way interaction between verb placement, sentence type and group. Consequently, results below are for all groups combined.

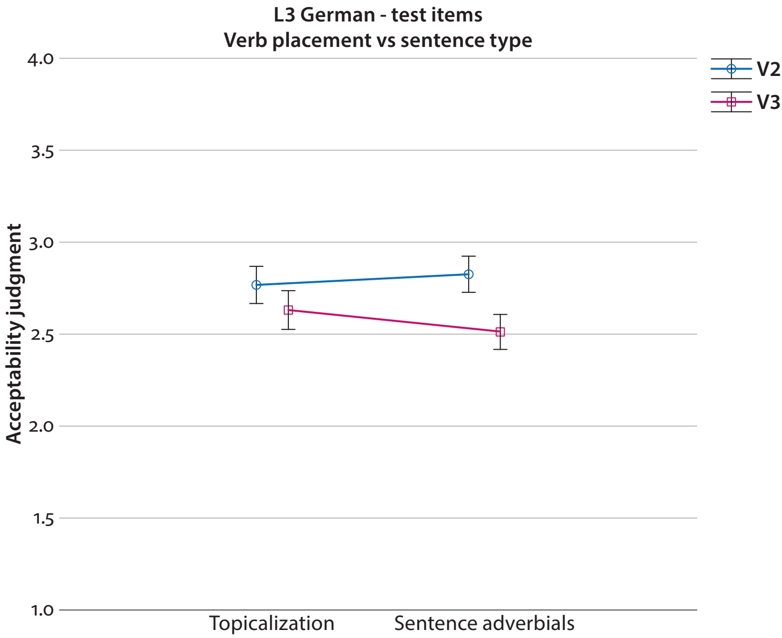


Figure 3. Interaction of verb placement and sentence type in L3 German
Note: Vertical bars denote 95% confidence intervals.

To investigate the interaction between verb placement and sentence type, Tukey HSD post-hocs were conducted. They revealed a significant difference between judgments for V2 and V3 word orders both for topicalizations ($p < .0001$) and sentences with a sentence adverbial ($p < .0001$), where V2 received higher acceptability scores than V3. Importantly, with V2, there was no difference in acceptability judgments for topicalizations (mean score 2.77, $SE = 0.05$) and sentences with a sentence adverbial (mean score 2.83, $SE = 0.05$). With V3, however, topicalizations received significantly higher acceptability scores (mean score 2.63, $SE = 0.05$) than sentences with a sentence adverbial (mean score 2.51, $SE = 0.05$, $p = .0003$). To the extent that participants accepted V3, this was not equal for both sentence types. Although mean scores for ungrammatical word order were lower than for grammatical word order in both structures, V3 word order cannot be said to be generally rejected, as mean scores are still above 2.5.

For the question of whether we see remaining evidence of wholesale transfer at the initial stages, inspection of individual data from participants in year 1 of German are particularly interesting, although we do not claim that they are still in the initial stages. For topicalizations, only 4 participants gave a mean score of 3 or above for one word order and 2 or below for the other. For sentence adverbials, no participants rated one word order on average 3 or above while rating the other 2 or below. This means that no participant consistently rejects one word order and accepts the other.

The overall conclusion is that we have no indication of wholesale transfer from either L1 or L2, and that there is a slight difference in judgments between the two sentence types. To the extent that there is any influence from the L2, it appears to become less pronounced with more L3 exposure as learners become more target-like in later years.

5.2 L2 English

We now discuss the results for L2 English in order to relate these to L3 German (RQs 3 and 4).

Performance on L2 English fillers was at ceiling across all groups; grammatical fillers were generally accepted (mean = 3.71, *SD* = 0.32) and ungrammatical fillers generally rejected (mean = 1.39, *SD* = 0.37). The similar performance across groups is not surprising, given that they all had 11–12 years of English instruction.

Table 4 shows mean scores and standard deviations for target sentences in L2 English per experimental group. Generally, performance is close to ceiling for all groups. Note, however, that mean ratings for ungrammatical V2 with sentence adverbials are higher than for topicalizations.

Table 4. Mean score and standard deviation for test items in L2 English per experimental group

Years of German	n	V2				V3			
		Topicalization		Sentence adverbial		Topicalization		Sentence adverbial	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	18	1.64	0.52	2.04	0.62	3.28	0.61	3.19	0.44
2	15	1.53	0.46	1.94	0.65	3.14	0.51	3.08	0.55
4	70	1.67	0.48	2.04	0.62	3.29	0.41	3.44	0.44
5	51	1.68	0.41	1.98	0.52	3.19	0.43	3.31	0.36

Sentence types were compared using a 2 (Verb placement: V2 vs V3) x 2 (Sentence type: Topicalization vs Sentence adverbial) repeated measures ANOVA with years of German instruction as the grouping variable. There were no significant differences between experimental groups. Thus, in the following, L2 English results are collapsed over all groups.

There was a main effect of verb placement ($F(1, 150) = 630.25$, $p < .0001$, $\eta_p^2 = .81$). Overall, grammatical V3 was accepted, whereas ungrammatical V2 was rejected (mean score = 3.24, *SE* = 0.05 and mean score = 1.82, *SE* = 0.08, respectively). However, there was a main effect of sentence type ($F(1, 150) = 37.53$, $p < .0001$, $\eta_p^2 = .20$), where sentences with a topicalized element received lower acceptability scores than sentences with sentence adverbials (mean score 2.43, *SE* = 0.05 and 2.63, *SE* = 0.05 respectively). Importantly, there was an interaction effect of verb placement and sentence type ($F(1,150) = 27.52$, $p < .0001$, $\eta_p^2 = .16$). This is illustrated in Figure 4 below.

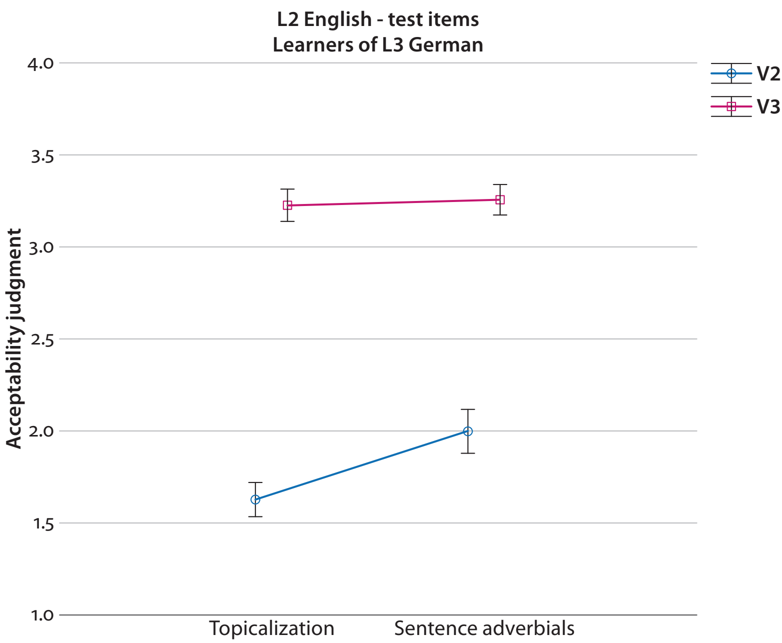


Figure 4. Interaction effect of verb placement and sentence type in L2 English
Note: Vertical bars denote 95% confidence intervals.

To break down the interaction, Tukey HSD post-hocs were conducted. The tests showed no significant differences in acceptance of grammatical (V3) sentences with topicalizations vs sentences with a sentence adverbial, mean scores 3.22 ($SE = 0.4$) and 3.24 ($SE = 0.4$), respectively. However, for the ungrammatical (V2) sentences, there was a significant difference ($p < .0001$) between topicalizations (mean score 1.63, $SE = 0.5$) and sentences with a sentence adverbial (mean score 2.0, $SE = 0.6$). Thus, ungrammatical V2 sentences with a sentence adverbial were not rejected as consistently as ungrammatical V2 sentences with topicalizations.

5.3 Relationship between L2 English and L3 German

In order to investigate the relationship between the two structure types in each language, and between each structure across L2 English and L3 German, we calculated the difference between each participant's mean score on the grammatical versus the ungrammatical condition of each sentence type by subtracting scores on the ungrammatical word order from scores on the grammatical word order (V2 and V3, respectively). This gives us a measure of the degree to which participants discriminated between the two, referred to as the discrimination score in the following. Discrimination scores constitute one variable for sentences with topicalizations and one variable for sentences with sentence adverbials in English and German, respectively. The maximum possible score for each variable is 3, indicating that the participant gave an overall score of 4 to grammatical structures and 1 to ungrammatical structures. A participant who gave a higher score in the ungrammatical condition than in the grammatical condition would receive a negative value. The higher the score, the greater the participants' discrimination between grammatical and ungrammatical word orders, and, if the score is positive, the higher their accuracy in judgments.[7]

To investigate the relationship between judgments in L2 and L3, correlation analyses were performed with Pearson's r across languages and structure types. Since L2 vs L1 transfer can be expected to differ at different stages of L3 acquisition (RQ5), we look both at all groups combined and at each experimental group separately.

Results for all participant groups combined showed a significant correlation between discrimination scores on both sentence types in L2 English ($r = .503$, $p < .0001$). This also held for each of the four experimental groups separately. Similarly, there was a significant correlation between discrimination scores on the

7. Note, however, that a discrimination score of for Example (1) does not tell us whether participants generally accepted (e.g., mean scores 3 and 4) or generally rejected (e.g., mean scores 2 and 1) sentences.

two sentence types in L3 German ($r = .479$, $p < .0001$) for all groups combined. However, the relationship between the two structures in L3 differed across the experimental groups. For participants in years 1 and 2 of German instruction, there was no significant correlation between the two structures ($r = -.062$, $p = .806$ and $r = .150$, $p = .593$, respectively). For participants in years 4 and 5, however, this correlation was significant ($r = .547$, $p < .0001$ and $r = .351$, $p = .012$, respectively). Thus, in years 1 and 2 preferences for V2 vs. V3 word order in L3 do not seem to be related in the two structures, whereas in years 4 and 5, participants show similar preferences across structures. This corroborates findings from Section 5.1 that we have no indication of wholesale transfer from either L1 or L2 across these structures in the earliest learners (RQ2).

As for relationships between judgments across languages (RQ3), there was no significant correlation between discrimination scores on sentences with topicalizations for L2 English and L3 German, and this holds for all four groups combined and across each experimental group separately. This may have to do with ceiling effects in L2 English for these structures. With sentence adverbials, however, there was a significant correlation between discrimination scores in L2 English and L3 German across all groups combined ($r = .381$, $p < .0001$). For this structure, the correlation was significant in year 1 ($r = .530$, $p = .024$), year 4 ($r = .395$, $p = .001$), and year 5 ($r = .298$, $p = .034$), but not in year 2 ($r = .246$, $p = .377$). These were the structures where V2 was not consistently rejected in L2 English. The results thus indicate that participants who did reject ungrammatical V2 in English preferred grammatical V2 in German.

To investigate the importance of L2 proficiency, we ran correlation analyses for English proficiency, as measured by self-rating scale and grade obtained, with discrimination scores in L3 German (RQ4). We found a significant positive correlation between both self-rated proficiency and grade obtained in English with discrimination scores on German sentences with sentence adverbials ($r = .322$, $p < .0001$ and $r = .366$, $p < .0001$, respectively), but not with topicalizations. Thus, higher proficiency in English seemed to predict more target-like performance with sentence adverbials in German.

As for proficiency in L3, we ran correlation analyses for the same measures of L3 proficiency with discrimination scores in L3 German (RQ5). For self-rated proficiency, there was a positive correlation across all groups with discrimination on judgments of German sentences with sentence adverbials ($r = .311$, $p < .0001$) and with topicalizations ($r = .161$, $p = .046$), indicating more target-like performance with higher self-rated proficiency.

For grades obtained in German, correlations were run separately per experimental group, since grades are relative depending on what can be expected in the given year; a high grade in year 1 does not reflect proficiency comparable to the same grade in year 5. There were no significant correlations with discrimination scores on either structure in years 1 and 2. In year 4, however, there was a significant positive correlation with both sentence adverbials (r = .348, p = .003) and topicalized structures (r = 390, p = .001). In year 5, there was a positive correlation with sentences with sentence adverbials only (r = .503, p < .0001). We return to these results in the discussion.

6. Discussion

We now consider what our results mean for our five research questions. We first asked whether there is evidence of L1 and/or L2 transfer to L3 in Norwegian L3 learners of German. Our results make it difficult to argue for clear evidence of wholesale transfer from either L1 or L2 at the earliest stages (Rothman, 2015), with no participant consistently accepting one word order and rejecting the other. Arguably, this most clearly indicates no wholesale transfer from L1 (Hermas, 2010), as we would expect this to lead to target-like judgments right from the start. On the other hand, it is possible that judgments do reflect residual transfer from L2, as was our prediction (in line with Bardel & Falk 2007; Bohnacker, 2006; Falk & Bardel, 2011), which could have been more evident at the very initial stages of acquisition, as our participants in year 1 had been learning the L3 for some months. However, our data most clearly show insecurity abut verb placement in German, indicating unstable representations for the structures. This interpretation is supported by the lack of correlation between the two sentence types in L3 German in years 1 and 2, since residual transfer would be expected to be related for the two structures (RQ 2). We do, however, see some evidence in years 4 and 5 that development towards target-like word order with sentence adverbials happens earlier than with topicalized structures. This may indicate property-by-property transfer (Westergaard et al., 2017; Slabakova, 2017). The observed insecurity in the earliest learners could be due to the potential availability for transfer of structures from two previously learned languages, e.g., along the lines of full transfer potential (Westergaard, 2021). The earliest learners may not yet have had enough evidence to establish which language is more similar to German in the relevant respects, indicating that both structures are equally available for transfer.

Our third question was whether mastery of the relevant structure in L2 matters for transfer into L3. It seems obvious that learners can only transfer the L2 grammar such that it is, meaning that transfer of English V3 must depend on the relevant structure having been acquired in L2 English. However, in the present study, we do not see evidence of a relationship between target-like judgments in L2 and possible transfer to L3.

In L2 English, we saw evidence of ceiling performance, but less clear rejection of sentence adverbials with V2, consistent with our prediction based on Westergaard (2002, 2003). In German, ungrammatical V3 with sentence adverbials receives lower scores than V3 with topicalizations, possibly implying that V3 word order with sentence adverbials, which may not be entirely in place in L2, is less eligible for transfer to L3 German. However, correlation analyses do not support this line of argument; those participants with most target-like performance on English verb placement also score more accurately on German verb placement for these structures. To the extent that there is evidence of transfer from English, it is less pronounced for those for whom these structures are firmly in place in English.

This finding relates to our fourth research question, about the role of proficiency in L2 and L3. Based on Bardel and Falk (2012) and Stadt et al. (2016, 2018b), we expected L2 proficiency to play a role, but had no prediction for the direction of the relationship. We find a positive correlation between our English proficiency measures and correct judgments on German sentences with sentence adverbials, but not with judgments on German topicalized structures. This indicates more accurate performance in German when English proficiency is higher, which could imply thatL2 is less eligible for transfer at higher proficiency levels. Alternatively, however, individual factors that have facilitated the acquisition of L2 English may have had the same effect on L3 German.

We do not see clear evidence of a privileged status of the L2 as the L2 Status Factor (e.g., Bardel & Falk, 2007) would predict. This could be due to generally high English proficiency in our participants, such that it approaches L1 competence in the relevant respects (Bardel & Falk, 2012). An explanation for why Stadt et al. (2016, 2018b) found different results may lie in how English proficiency was attained: In Stadt et al.'s studies, those with higher English proficiency had either studied English for longer, or they were in an immersion English class setting. In other words, their higher English proficiency was due to external circumstances.[8] In our study, participants had been learning English for longer, and importantly, external circumstances for English were relatively constant across groups, with no important differences between groups in years of English instruction or learning environments. Thus, the main differences in English proficiency in our participants are likely to have been internal or at least due to individual factors, which may be common to L2 and L3 learning. This would explain the general correlation between target-like performance in the two languages.

For L3 proficiency (RQ 5), those with self-perceived higher German proficiency were also more accurate in judgments. However, there was no relationship between

8. It should be noted that our participants were older than those in Stadt et al.'s studies. This could potentially entail cognitive differences in the learning process.

grades in German and discrimination scores in years 1 and 2. This may reflect a focus on for instance vocabulary learning, fixed expressions and rote learning of morphology in the first years of formal L3 instruction, such that grades do not reflect syntactic competence. The positive correlation between grades and discrimination scores in years 4 and 5 is to be expected, as syntactic competence is more likely to influence grades in later years of formal language learning.

RQ 5 also asked what developmental trajectories look like. From year 1 to year 5, we see a general development from judgments around the middle of the scale toward more target-like judgments. However, even in years 4 and 5, there is a relatively high degree of insecurity, especially for ungrammatical structures. Still, these participants do clearly distinguish between grammatical and ungrammatical. Moreover, variance decreases in years 4 and 5, which may indicate less insecurity, although it should be noted that group sizes differ. Overall performance on sentences with sentence adverbials was more accurate than on topicalizations. Still, the lack of interaction effect between groups and sentence types means that we cannot make claims about developmental trajectories for each structure individually.

Of course, the less accurate performance in years 1 and 2 could indicate development from more L2 transfer at the earlier stages to more L1 transfer at later stages, since the former would lead to non-target-like judgments and the latter to target-like judgments. However, a more likely explanation is, as argued above, that learners start out being insecure about German verb placement, and then acquire it with more exposure. In the earliest learners, both structures seem to be readily available for transfer. In later stages, the target-like structure (compatible with L1) seems to become more stable, while the L2 structure is clearly still available. This is compatible with a full transfer potential account of L3 acquisition (Westergaard, 2021).

7. Conclusion

Our results did not show evidence which would indicate that wholesale transfer had taken place from either L1 or L2 at the earliest stages of L3 acquisition. Rather, our data seem to indicate insecurity indicative of unstable representations, and are compatible with property-by-property transfer and full transfer potential (Slabakova, 2017; Westergaard et al., 2017; Westergaard, 2021). We saw no indication that structures which are 'in place' (i.e., acquired or internalized) in L2 are more eligible to transfer into L3. Furthermore, higher proficiency in L2 English predicted more target-like performance on L3 German structures, despite the structures in question being different in the two languages. With additional years of instruction in L3 German, participants more clearly distinguished between grammatical and ungrammatical German sentences, indicating development towards increased stability in target-like representations.

The study used a simple methodology of acceptability judgments. While we might have gained more insight into the participants' L3 competence by asking them to explain why they rejected certain sentences, this might have introduced an undesired element of explicit metalinguistic knowledge. However, methods forcing participants to indicate a preference for either V2 or V3 word order might have provided clearer results. Still, it is not obvious that this would have given us clearer insight into transfer processes in L3 acquisition in this learning context. Production tasks, on the other hand, could have provided further insight into possible transfer processes. Furthermore, both structures in the test were generally in place in L2. Testing other types of structures which are known to be more problematic in L2 English, e.g., V2/non-V2 with auxiliaries (see Rankin, 2011), might have provided further understanding of the influence of the developing L2 grammar in L3. Moreover, measures of proficiency more objective than self-reports and grades might have allowed a more precise understanding of the role of L2/L3 proficiency. Finally, comparison to L3 learners of languages with word orders similar to L2 English rather than to L1 Norwegian would be useful to further gauge the role of L1 vs L2 as candidates for transfer in instructed L3 acquisition.

References

Bardel, C., & Falk, Y. (2007). The role of the second language in third language acquisition: The case of Germanic syntax. *Second Language Research*, 23(4), 459–484. https://doi.org/10.1177/0267658307080557

Bardel, C., & Falk, Y. (2012). The L2 status factor and the declarative/procedural distinction. In J. Cabrelli Amaro, S. Flynn, & J. Rothman (Eds.), *Third language acquisition in adulthood* (pp. 61–78). John Benjamins. https://doi.org/10.1075/sibil.46.06bar

Bohnacker, U. (2006). When Swedes begin to learn German: From V2 to V2. *Second Language Research*, 22(4), 443–486. https://doi.org/10.1191/0267658306sr275oa

Falk, Y., & Bardel, C. (2011). Object pronouns in German L3 syntax: Evidence for the L2 status factor. *Second Language Research*, 27(1), 59–82. https://doi.org/10.1177/0267658310386647

Flynn, S., Foley, C., & Vinnitskaya, I. (2004). The Cumulative-Enhancement Model for language acquisition: Comparing adults' and children's patterns of development in first, second and third language acquisition of relative clauses. *International Journal of Multilingualism*, 1(1), 3–16. https://doi.org/10.1080/14790710408668175

Hermas, A. (2010). Language acquisition as computational resetting: Verb movement in L3 initial state. *International Journal of Multilingualism*, 7(4), 343–362. https://doi.org/10.1080/14790718.2010.487941

Holmberg, A., & Platzack, C. (1995). *The role of inflection in Scandinavian syntax*. Oxford University Press.

Håkansson, G., Pienemann, M., & Sayehli, S. (2002). Transfer and typological proximity in the context of second language processing. *Second Language Research*, 18(3), 250–273. https://doi.org/10.1191/0267658302sr206oa

Jin, F. (2009). Third language acquisition of Norwegian objects: Interlanguage transfer or L1 influence? In Y.-k. I. Leung (Ed.) *Third language acquisition and Universal Grammar* (pp. 144–161). Multilingual Matters. https://doi.org/10.21832/9781847691323-010

Opsahl, T. (2009). "Egentlig alle kan bidra!": En samling sosiolingvistiske studier av strukturelle trekk ved norsk i multietniske ungdomsmiljøer i Oslo (Unpublished doctoral dissertation). University of Oslo.

Pollock, J.-Y. (1989). Verb movement, universal grammar, and the structure of IP. *Linguistic Inquiry*, 20(3), 365–424.

Rankin, T. (2011). The transfer of V2: Inversion and negation in German and Dutch learners of English. *International Journal of Bilingualism*, 16(1), 139–158. https://doi.org/10.1177/1367006911405578

Rizzi, L. (1996). Residual verb second and the wh-criterion. In A. Belleti & L. Rizzi (Eds.), *Parameters and functional heads: Essays in comparative syntax* (pp. 63–90). Oxford University Press.

Roberts, I. (2001). Head movement. In M. R. Baltin & C. Collins (Eds.), *The handbook of contemporary syntactic theory* (pp. 112–147). Blackwell. https://doi.org/10.1002/9780470756416.ch5

Rothman, J. (2015). Linguistic and cognitive motivations for the Typological Primacy Model (TPM) of third language (L3) transfer: Timing of acquisition and proficiency considered. *Bilingualism: Language and Cognition*, 18(2), 179–190. https://doi.org/10.1017/S136672891300059X

Slabakova, R. (2017). The scalpel model of third language acquisition. *International Journal of Bilingualism*, 21(6), 651–665. https://doi.org/10.1177/1367006916655413

Stadt, R., Hulk, A., & Sleeman, P. (2016). The influence of L2 English and immersion education on L3 French in the Netherlands. *Linguistics in the Netherlands*, 33(1), 152–165. https://doi.org/10.1075/avt.33.11sta

Stadt, R., Hulk, A., & Sleeman, P. (2018a). The influence of L1 Dutch and L2 English on L3 French: A longitudinal study. *Journal of the European Second Language Association*, 2(1), 63–71. https://doi.org/10.22599/jesla.42

Stadt, R., Hulk, A., & Sleeman, P. (2018b). The role of L2 exposure in L3A: A comparative study between third and fourth year secondary school students in the Netherlands. In J. Berns, H. Jacobs, & D. Nouveau (Eds.), *Romance languages and linguistic theory: Selected papers from 'Going Romance' 29, Nijmegen* (Vol. 13, pp. 279–296). John Benjamins. https://doi.org/10.1075/rllt.13.17sta

Stadt, R., Hulk, A., & Sleeman, P. (2020). L1 Dutch vs L2 English and the initial stages of L3 French acquisition. In C. Bardel & L. Sánchez (Eds.), *Third language acquisition: Age, proficiency and multilingualism* (pp. 237–261). Language Science Press.

Westergaard, M. (2002). Minimal input in L2 acquisition: How Norwegian children learn English word order. *Acta Didactica*, 3, 199–218.

Westergaard, M. (2003). Unlearning V2: Transfer, markedness, and the importance of input cues in the acquisition of word order in English by Norwegian children. *EUROSLA Yearbook*, 3, 77–101. https://doi.org/10.1075/eurosla.3.07wes

Westergaard, M. (2021). Microvariation in multilingual situations: The importance of property-by-property acquisition. *Second Language Research*, 37(3), 379–407. https://doi.org/10.1177/0267658319884116

Westergaard, M., Mitrofanova, N., Mykhaylyk, R., & Rodina, Y. (2017). Crosslinguistic influence in the acquisition of a third language: The Linguistic Proximity Model. *International Journal of Bilingualism*, 21(6), 666–682. https://doi.org/10.1177/1367006916648859

Feature reconfiguration at the syntax-discourse interface

L2 acquisition of Italian CLLD

Liz Smeets
York University

This study examines L1 transfer in the L2 acquisition of feature mappings associated with Italian Clitic Left Dislocation (CLLD) by native speakers of English and Romanian. In Italian, insertion of a clitic after dislocating a direct object is restricted to [+anaphor] objects (López, 2009) and in Romanian to [+specific] objects. Extending the Feature Reassembly Hypothesis (Lardiere, 2009) to the syntax-discourse interface, successful acquisition is arguably more difficult for the L1 Romanian group than the L1 English group due to the need for feature reassembly. The findings show convergence with the target language only for the L1 English group in the near-native levels of L2 proficiency. The results suggest that reconfiguration is prone to fossilization when reassembly is required.

Keywords: feature reconfiguration, syntax-discourse interface, clitic left dislocation, Romance languages, Italian-Romanian, near-native grammars

1. Introduction

Second language (L2) acquisition research examines whether L2 learners can reach native-like ultimate attainment and, if not, which factors contribute to persistent non-convergence with the target language. One contributing factor is the impact of the L1 grammar on L2 performance; while positive transfer can be helpful, negative transfer can impede acquisition. This paper focuses on L1 transfer in the acquisition of discourse constraints on the use of clitics after left dislocating an object in Italian, a construction called Clitic Left Dislocation (CLLD).[1] An example of

1. The term CLLD does not refer to left dislocation of discourse topics per se, as is typically the case in the Romance literature. To accommodate data from Romanian, where CLLD is not restricted to discourse topics, the term CLLD is used to describe a syntactic construction in which an object has moved to left periphery and is doubled by an agreeing clitic, independently of the discourse properties of the dislocated constituent.

https://doi.org/10.1075/lald.67.09sme

CLLD is shown in (1), where the object *il divano* is moved to the left periphery and doubled by an agreeing clitic (l').

(1)	[**Il**	**divano**]$_i$ **li'**	ho	messo	t_i	in	soggiorno.
	the	couch	cl.acc.m.sg-have	put	t_i	in	living room

'The couch, I put in the living room.'

For an overt clitic to be felicitous in Italian, speakers need to integrate discourse information with the syntax. This study investigates whether L2 learners can associate CLLD with different discourse contexts in the L2 (Italian) than those in which CLLD is used in the L1 (Romanian). Romanian learners of Italian will be compared to English learners of Italian, as English does not use resumptive pronouns after left dislocating an object, and, thus, L1 transfer of features associated with clitic use should not be an issue.

In the recent literature, the syntax-discourse interface has received considerable attention as one of the areas of the grammar in which L2 speaker performance falls short of native speaker performance (Sorace & Filiaci, 2006; Sorace & Serratrice, 2009; Tsimpli & Sorace, 2006). Persistent non-convergence with the target language has been argued to be due to processing difficulties in integrating discourse information into the syntax, and, crucially, this is argued to be the case independently of the L1 of the learner (Sorace, 2011). A generalization based on earlier studies testing second language learners' acquisition of discourse constraints on syntax is that while higher demands on processing resources can contribute to performance differences between L2 learners and native speakers (e.g., Hahne, 2001; Hopp, 2006; McDonald, 2006), evidence for a generalized interface-related deficit is not reported. Slabakova, Kempchinsky, and Rothman (2012); Slabakova and Ivanov (2011); Hopp (2009); Smeets (2018); and Laleko and Polinksy (2016), for example, did not find evidence for persistent difficulties with the phenomena at the syntax-discourse interface that they tested. To gain further insights into the factors that facilitate or hinder L2 acquisition of discourse-syntax phenomena, other aspects should be studied, such as how negative transfer of L1 discourse properties affects learnability, a factor currently understudied in the domain of syntax-discourse (with the exception of Bohnacker & Rosen (2008); Hopp (2009); Slabakova & García Mayo (2015); and Valenzuela (2005)).

The view on L1 transfer I adopt in this paper is based on the Feature Reassembly Hypothesis (FRH) (Lardiere, 2009), which assumes that L2 learners transfer formal linguistic features (following the Minimalist view (Chomsky, 2000)) from the L1 to the L2. In cases where the one-to-one initial mapping of features is unsuccessful, as is the case for the Italian-Romanian combination investigated here, L2 learners will have to reorganize the grammatical system, i.e., they need to reassemble the features involved.

The paper is structured as follows: Section 2 discusses the object movement constructions under consideration and the relevant properties of CLLD in Italian and Romanian. Section 3 summarizes the FRH and discusses the predictions related to L1 transfer within the framework of the FRH. Section 4 presents the research questions and Section 5 focuses on the experimental tasks and results and is followed by a discussion and conclusion in Section 6 and 7.

2. Syntax and properties of left dislocated objects

2.1 Object left dislocation constructions

This paper considers two types of object left dislocation constructions which are typically known as Topic Fronting/ Topicalization and Focus Fronting/ Focalization, examples of which are shown in (2) and (3), respectively.

(2) **Topic Fronting**
Q: What did you do with the couch and with the table?
A: [The couch]$_i$ I put t_i in the living room, but the table broke during transportation.

(3) **Focus Fronting**
Q: You put the table in the living room, right?
A: [THE COUCH]$_i$ I put t_i in the living room, not the table.

The three languages under consideration (Italian, Romanian, and English) all allow Topic Fronting and Focus Fronting.[2] The crucial difference between these languages, however, is whether they insert a resumptive clitic after dislocating an object and, if so, in which contexts a clitic is used. English does not use CLLD (Note: *The couch I put (*it) in the living room.*), but the construction is used in both Italian and Romanian. To describe the featural differences between Italian and Romanian in their use of CLLD, I elaborate on the feature [±anaphor], as developed by López (2009), and the feature [±specific].

López (2009) explains the difference between sentences like (2) and (3) using the discourse notion [±anaphor]. The dislocated constituent is discourse anaphoric when it stands in a relation of structural asymmetry to a sentence external antecedent from within the immediate discourse. In (2), the object is discourse anaphoric;

2. All examples of Focus Fronting used in this paper and in the experiment are instances of corrective focus where a constituent in the previous utterance is being corrected, as this is the context in which Focus Fronting is most felicitous (based on experimental evidence from Bianchi et al. (2013) on Italian).

the dislocate *the couch* has a discourse antecedent in the question (*the couch)*, and the answer elaborates on the previous sentence by contributing new information about what happened to the couch. In discourse structures like (3), the dislocated object has no antecedent in the immediate discourse. Hence, the dislocate is not anaphoric. According to Lopez (2009), the relevant discourse feature is [±anaphor], which is assigned to the syntactic derivation by an external system called *pragmatics*, as will be elaborated on in Section 2.3.[3]

With respect to specificity, it will be assumed that a noun phrase is interpreted as specific when the speaker has a particular referent in mind. In the examples in (2) and (3), we are referring to a specific couch. Compare this to the sentence *Shall we go shopping for a couch?,* where the speaker does not have a specific couch in mind. Hence, *a couch* here is non-specific. Let us look at how the properties of *discourse anaphoricity* and *specificity* determine the use of clitics in object left dislocation construction in Italian and Romanian.

2.2 The interpretative properties of CLLD in Italian and Romanian

The CLLD sentences in (4a) and (4b) are embedded in a context that renders this construction felicitous in both languages.

(4) **[+specific, +anaphor]**
Luca and Michaela recently moved into their new house. Luca worked all day while Michaela stayed home to organize the furniture. Luca calls her to ask how it went and says:
Q: What did you do with the couch and the table?

Italian

a. [Il divano]$_i$ l$_i$' ho messo t_i in soggiorno, ma il tavolo
the couch cl.acc.m.sg have put t_i in living room but the table
si è rotto durante il trasporto.
REFL is broken during the transportation
'The couch I put in the living room, but the table broke during transportation.'

3. López (2009) argues for the existence of two discourse notions [±contrast] and [±anaphor], which account for the differences between various dislocation constructions. [±Contrast] is not relevant for the purpose of the current study. In all languages under consideration (Italian, Romanian and English), objects in the left periphery are typically associated with a contrastive interpretation of the fronted object. Similar claims have been made for other languages, for example by Frey (2006) for German or Kiss (1998) for Hungarian.

Romanian

b. [Canapeaua]$_i$ am pus-o$_i$ t_i în sufragerie, dar masa
couch.def have put-cl.acc.m.sg t_i in living room but table.def
s-a rupt în timpul transportului.
REFL-is broken in time transportation
'The couch I put in the living room, but the table broke during transportation.'

However, as summarized in Table 1, the use of the clitic is driven by different features in the grammar; while Italian restricts CLLD to constructions where the fronted object is a *discourse anaphor* (see the first line in Table 1), this restriction does not hold in Romanian. Romanian uses clitics only when the fronted object has a *specific referent* (see line 2 in Table 1).

Table 1. Distribution of resumptive clitics in Italian and Romanian

		[+anaphor]		[−anaphor]		Property
		[+specific]	[−specific]	[+specific]	[−specific]	
1	Italian	✓	✓	✗	✗	anaphoricity
2	Romanian	✓	✗	✓	✗	specificity

To illustrate the role of discourse anaphoricity in Italian, compare the Italian and Romanian examples in (5) to the examples in (4) with respect to the status of the clitic. In the Italian sentence in (4a), where the dislocate has an antecedent in the immediate discourse, a clitic is obligatory, while the clitic is not allowed in (5a) where the dislocate is not a discourse anaphor. Romanian, on the other hand, requires a clitic in both discourse contexts. In other words, Italian CLLD is restricted to constructions where the dislocate stands in an anaphoric relation to an antecedent in the discourse. For Romanian, however, discourse anaphoricity is irrelevant as to whether a clitic is included.

(5) [**+specific, −anaphor**]
Q: You put the table in the living room, right?

Italian

a. Il DIVANO (*l') ho messo in soggiorno, non il tavolo.
the couch cl.m.3sg have.1sg put in living room not the table
Il tavolo si è rotto durante il trasporto.
the table REFL is broken during the transportation
'The couch I put in the living room, not the table. The table broke during the transportation.'

Romanian

b. CANAPEAUA am pus-*(o) în sufragerie, nu masa.
couch-the have.1sg put-cl.f.3sg in living room not table-the
Masa s-a rupt în timpul transportului.
table-the REFL-has broken in time transportation
'The couch I put in the living room, not the table. The table broke during the transportation.'

In Romanian, but not in Italian, clitics can *only* be used with specific objects (see also Avram & Martine, 1999; and Cornilescu & Dobrovie-Sorin, 2008). In the examples in (4) and (5), the speaker has a specific couch in mind (there is one particular referent) and therefore the clitic is obligatory in the Romanian sentences. However, if we change the discourse context such that the dislocate is non-specific, as is the case for a *(glass of) wine* in (6), resumptive clitics are not allowed in the Romanian sentences (see (6b) and (7b)). For Italian, specificity does not affect the use of clitics; both specific and non-specific objects are used in CLLD constructions, as can be seen with the obligatory status of the clitic in (6a).

(6) **[−specific, +anaphor]**
Emma and Elio are at a restaurant with Nicolò and Susanna, Elio's brother and sister. Nicolò and Susanna excuse themselves to go to the bathroom and ask Elio to order for them. When the waiter arrives, Elio isn't sure what Nicolò and Susanna requested and he asks Emma:
Q: 'Who wanted a glass of wine and who a beer?'

Italian

a. Un vino *(lo) ha ordinato tuo fratello e tua sorella
a wine cl.m.3sg has ordered your brother and your sister
vorrebbe una birra.
want.cond.3sg a beer

Romanian

b. Un vin (*l-)a comandat fratele tău iar sora ta ar
a wine cl.m.3sg-has ordered brother your and sister your would
vrea o bere.
want a beer
'Your brother ordered a glass of wine, and your sister would like a beer.'

In the context in (7) neither Italian nor Romanian allows the use of a clitic. This holds for Italian because the fronted object is [−anaphor] and for Romanian because it is [−specific].

(7) **[–specific, –anaphor]**
Q: Nicolò ordered a beer, right?

Italian

a. UN VINO (*lo) ha ordinato tuo fratello, non una birra. È
a wine cl.m.3sg has ordered your brother not a beer is
tua sorella che vorrebbe una birra.
your sister who want.cond.3sg a beer

Romanian

b. Un VIN (*l-)a comandat fratele tău, nu o bere. Sora ta ta
a wine cl.m.3sg-has ordered brother your not a beer sister you
e cea care ar vrea o bere.
is one who would want a beer
'Your brother ordered a glass of wine, not a beer. It's your sister who would like a beer.'

In sum, two features are relevant, namely [±anaphor] for Italian and [±specific] for Romanian. The anaphoricity requirement for CLLD in Italian reflects the fact that the dislocate is obligatorily linked to a discourse antecedent. Romanian has a specificity requirement, as left dislocated object DPs are doubled by a clitic when they are referentially anchored to one particular entity.

2.3 Assignment of [+anaphor] and [+specific]

This section discusses how Italian and Romanian compare in the assignment of the features [+anaphor] and [+specific]. Let us first look at the feature [+anaphor] that triggers the use of clitics in Italian, for which I adopt the analysis proposed by López (2009). He argues that discourse features like [±anaphor] are assigned by an external pragmatic module of computation onto the syntactic derivation. Crucial to his analysis is the idea that information structure features can only be assigned to dislocated constituents which have moved to a specific phase edge, as this is the location where the syntactic structure is visible to the pragmatic module (following Chomsky, 2000). For [+anaphor], in particular, López (2009) argues that this feature is assigned to elements that have moved to the edge of vP and who stand in an agreement relation with a clitic.

To illustrate with the example in (4a), here repeated as (8), in the syntactic structure in Figure 1, the dislocate *il divano* in Italian has moved from its VP internal base position to the specifier of *v*P. In this position, the dislocated constituent can enter an Agree relation with the clitic that is adjoined to the *v* head of the vP (here the auxiliary *ho*). The pragmatic module inspects the syntactic object in

Spec, v and assigns the pragmatic feature [+anaphor] to the clitic in the phrasal head position. This feature is then automatically shared with the object in Spec, *v*P by agreement. When there is no clitic, a dislocated object is automatically assigned [−anaphor].

(8) [Il divano]$_i$ l$_i$' ho messo t_i in soggiorno.
the couch cl.acc.m.sg have put t$_i$ in living room
'The couch I put in the living room, […]'

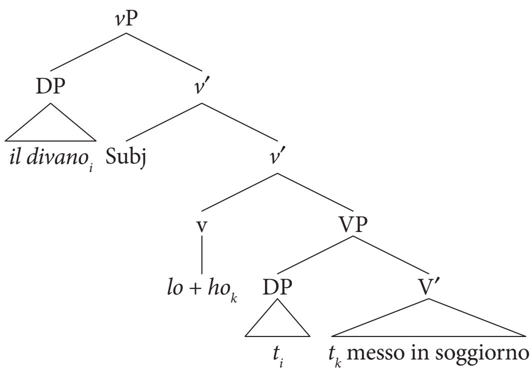


Figure 1. Syntactic representation of the assignment of [+anaphor]

Because anaphoricity does not predict syntactic behavior in Romanian (and this language is not discussed in López (2009)), I assume that [±anaphor] is not assigned to clitics in Romanian, where the use of clitics is subject to a specificity requirement. One crucial difference between [±anaphor] for Italian and [±specific] for Romanian is that, while the assignment of [±anaphor] is dependent on movement, specificity is not. In fact, specificity is an intrinsic property of clitics in Romanian and not a characteristic of clitics in CLLD constructions alone. Evidence for this comes from the observation that specificity is also a requirement on Romanian clitics in constructions where the associated object does not move, such as clitic doubling with Differential Object Marking (e.g., Dobrovie-Sorin, 1990).

Based on these observations, I assume that in Romanian [+specific] is a property of the clitic which shares this feature with the dislocated constituent when this has moved to Spec, *v*P and creates a spec-head relation with the clitic. When there is no clitic, the dislocated object is automatically [−specific]. [+specific] is not a property of Italian clitics since they can agree both with specific and non-specific dislocated objects.

In sum, I assume that, in both languages, the relevant features are shared by the clitic with the dislocated object in Spec, *v*P through agreement. The languages differ in how the clitic receives the relevant feature: [+anaphor] is a property of CLLD in Italian, but not in Romanian, and is assigned to the clitic by the pragmatic module (following López, 2009). [+specific] is an intrinsic property of Romanian clitics but not of Italian clitics. Section 3 focuses on the acquisition task of English and Romanian learners of Italian in light of the Feature Reassembly Hypothesis (FRH).

3. Feature acquisition and feature reconfiguration

The predictions for the Romanian learners of Italian are developed within the ideas of the FRH) as proposed by Lardiere (2008, 2009). The FRH argues that the acquisition task consists of mapping or assembling universal features onto new formal configurations. Specifically, the FRH assumes that the L2 acquisition process occurs in two stages. There is a mapping stage where L2 learners apply their L1 feature specifications onto the L2 grammar. This means that the L1 instantiates the starting point of L2 acquisition (following Schwartz & Sprouse's (1996) Full Transfer Full Access model). The mapping stage is followed by a reassembly stage where L2 learners reconfigure their feature organizations. In the mapping stage, second language learners "look for morpholexical correspondences in the L2 to those in their L1, presumably on the basis of semantic meaning or grammatical function (the phonetic matrices will obviously differ)" (Lardiere, 2009, p. 91). In the reassembly stage, L2 learners can modify their L1 feature representations by adding or deleting features on the basis of the L2 input. This reassembly stage, which is required if the L1 specifications differ from those of the L2, is predicted to be hard for L2 learners. In consequence, successful L2 acquisition may be delayed or can even lead to fossilization. To date, the FRH has been applied to the acquisition of semantic and morphosyntactic features. The current study extends this approach to the acquisition of discourse features.

In this study, the performance of Romanian learners of Italian is compared to learners of the same L2 who are native speakers of English. Table 2 summarizes if and when clitics are used in the three languages under investigation.

Table 2. Distribution of clitics in Italian, Romanian and English

	Clitic used?	**Clitic distribution based on:**
Italian (L2)	YES	Anaphoricity / [+anaphor]
Romanian (L1)	YES	Specificity / [+specific]
English (L1)	NO	NA

The FRH predicts that the L2 acquisition process differs for the two L1 groups due to differences between Romanian and English; while the Romanian learners may experience L1 transfer effects and require reassembly of features for successful L2 acquisition, this does not hold for English. Comparing learners in both groups allows us to distinguish possible general difficulties with acquisition of discourse dependent syntactic constructions from persistent L1 transfer effects. Let us look at the predictions for each L1 group in more detail.

Since CLLD does not exist in English, English learners of Italian have to acquire this construction from scratch. Earlier studies investigating discourse constraints on Clitic Left Dislocation show successful L2 acquisition of discourse constraints on syntax. Convergence with the target language is found in studies testing different L2s (where the L1 is always English), including Donaldson (2011) for L2 French, Ivanov (2009, 2012) for L2 Bulgarian and Slabakova et al. (2012) for L2 Spanish. These findings suggest that L2 learners can acquire new syntactic constructions and use them in the right discourse contexts. The L1 English group is therefore also predicted to successfully acquire the discourse constraints on Italian CLLD.

Following the FRH, the Romanian learners are predicted to initially map their L1 features onto Italian CLLD. That is, they associate CLLD with fronted specific objects. In the reassembly stage, the [±specific] feature on the clitic needs to be removed and replaced with the [±anaphor] feature. To do so, the Romanian learners of Italian need to acquire the fact that clitics are used in Italian with fronted specific and non-specific topics, thus removing the [±specific] feature as not relevant for Italian clitics; they also have to learn that the [±anaphor] feature determines clitic distribution in L2 Italian. Unlike Romanian, in Italian clitics can only appear when they (and the dislocated object) are specified as [+anaphor] in the derivation.

An additional question is how the Romanian learner of Italian alters the L1 grammar into a target-like grammar on the basis of the input alone, without explicit instruction. Learners can use the presence of clitics with fronted non-specific topics to realize that the [±specific] feature is not relevant in the L2 and thus remove the [+specific] specification from the clitic that they may have transferred from their L1 Romanian. Acquisition of the [±anaphor] feature may be more difficult, as learners need to observe the absence of clitics with [−anaphor] objects in order to retreat from the use of a clitic in such contexts.

Feature Reassembly may not be achieved on the basis of positive evidence alone (see also Slabakova & García Mayo (2017) for data on Spanish learners of English) and explicit instruction may be needed to alter the grammar. Crucially, however, negative evidence is generally not reliably provided to all learners. In fact, Leal and Slabakova (2019) showed that CLLD is typically not taught in the L2 classroom and that language teachers often do not have metalinguistic awareness of its discourse use.

The next section presents the research question in more detail, focusing on L1 transfer and the role of L2 input in overcoming L1 transfer effects.

4. Research question and feature reconfiguration

Based on the predictions discussed in Section 3, the following research question is investigated:

> *Is reassembling discourse features possible for the Romanian learners of Italian and if so, is it more difficult than acquiring discourse restrictions on CLLD for the English learners of Italian, who have no prior L1 knowledge of this construction?*

To answer this question, we tested knowledge of discourse constraints on Italian CLLD by two groups of L2 learners (L1 English and L1 Romanian) at two different stages of development (Intermediate/Advanced and Near Native). The next section focuses on the methodologies used and the experimental results.

5. The experiment

5.1 Participants

In total, 18 Italian monolinguals, 17 Romanian monolinguals, 25 L1 Romanian and 30 L1 English L2 learners of Italian were tested on an Acceptability Judgment task and a Written Elicitation task. The Italian monolinguals and the L2 learners took the tests in Italian and the Romanian monolinguals in Romanian. Participants were found through word of mouth, Facebook pages on the internet targeting English or Romanian native speakers living in Italy, the British school in Rome and the Romanian cultural institute in Rome. Prior to participating, it was ensured that they were at least at the high-intermediate level (a B2 level in the European system) and did not speak another language with CLLD (such as Spanish). All learners lived or had lived in Italy and used Italian on a regular basis (see Table 3 for details).

The L2 learners were divided into two proficiency levels (Intermediate/Advanced and Near-Native) based on their scores on an Italian proficiency tasks and a C-test (see Section 5.3.1. and Smeets (2019) for details). Participants who scored between 115–134 points were categorized as near native (they all had a (near-)perfect score on the proficiency task and scored within the same range as native speakers), 100–114 as advanced, and 70–99 as Intermediate. In the L1 Romanian group, there were 10 Int/Adv and 15 Near-Native speakers, and the L1 English group consisted of 15 Int/Adv and 13 Near-Native speakers (two speakers were excluded due to low proficiency in the L2). In the analysis, the data from the Intermediate and Advanced groups were collapsed because there was no observable difference in behavior between those two groups.

Table 3. Background information, showing mean and range

	#	Age	Age at onset	Years of learning	Hours per week
Italian monolinguals	18	33.4 (24–53)			
Romanian monolinguals	17	33.4 (24–51)			
L1 Romanian Int/Adv	10	32.9 (21–48)	22.3 (16–30)	10.6 (2–27)	10 (0–50)
L1 Romanian Near Native	15	30.1 (21–43)	19.4 (16–24)	10.7 (6–19)	50.2 (0–140)
L1 English Int/Adv	15	36.1 (21–62)	22.14 (16–33)	13.9 (3–34)	24.8 (1–82)
L1 English Near Native	13	43.1 (22–68)	22.6 (16–40)	20.3 (4–50)	44.1 (4–130)

5.2 Task 1: Acceptability judgments

5.2.1 *Materials and procedure*

In this experiment participants judged the acceptability of sentences embedded in *wh*-question contexts identical to the examples illustrated in (4) to (7) on a 6-point rating scale (where 1 indicated highly unacceptable and 6 highly acceptable). Please note that in the experiment the contexts and questions as shown in (4) to (7) were presented in Italian or Romanian, not in English. The target items differed by three factors: *Discourse* and *Clitic*, which are within item factors and *Specificity*, which is a between item factor. The factor Discourse had two levels: in the Topic context (as in (4) and (6)) the left dislocated object was a contrastive topic which was discourse anaphoric and in the Focus context (as in (5) and (7)) the fronted object was a contrastive focus which was not discourse anaphoric. The factor Specificity also had two levels: the fronted object was either specific (as in (4) and (5)) or non-specific (as in (6) and (7)). The experiment was presented to the participants as follows (both in written and auditory form): the context automatically appeared after which participants clicked "next", followed by the question-and-answer pair. Auditory stimuli were used to ensure that participants processed the sentences with the intended intonation.[4] Two native speakers of Italian (one female and

4. A difference in discourse meaning is typically associated with a difference in intonation contour. The corrective focus construction considered here carried a high tone (H*) on the fronted constituent, followed by a default low tone (Jackendoff, 1972; Pierrehumbert, 1980). Contrastive topic configurations were associated with a 'rise-fall-rise' intonation, where the fronted contrastive topic was realized as H or L+H∗ followed by a L – H% boundary sequence.

one male) recorded the stimuli for the Italian version and two native speakers of Romanian (one female and one male) those for the Romanian version. Questions were pronounced by the male and contexts and answers by the female speaker. The Romanian version of the experiment was identical to the Italian version.

5.2.2 *Results and analysis*

The results from the Italian and Romanian monolinguals are discussed first, followed by the results from the L2 speakers, divided by L1 group.

5.2.2.1 *Italian and Romanian monolinguals*

To examine the effects of discourse function (anaphoricity) and specificity on the use of CLLD in Italian and in Romanian, the results were plotted for each language group separately. Figure 2 shows the mean acceptability judgments and standard errors for each condition and each L1 group. As predicted, the Italian monolinguals judged the clitic sentences as more acceptable and the no-clitic sentences as less acceptable when the object was a topic ([+anaphor]), independently of specificity. When the fronted object was a focus ([−anaphor]), no-clitic sentences were preferred over clitic sentences. Romanian monolinguals preferred clitic sentences and disfavoured no-clitic sentences when the fronted object was specific and preferred no-clitic sentences over clitic sentences when the fronted object was non-specific, independently of anaphoricity.

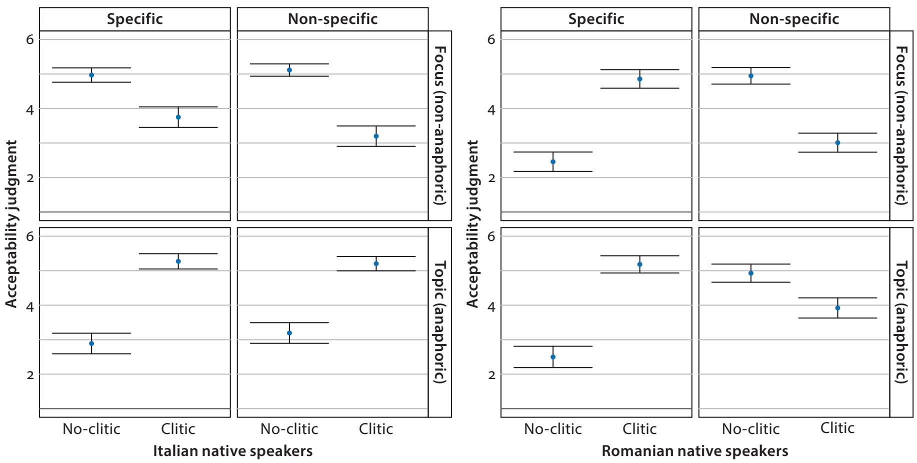


Figure 2. Acceptability judgments from native speakers of Italian (left) and native speakers of Romanian (right)

Table 4 shows the results of the models for each L1 group individually to investigate the effect of specificity and anaphoricity on acceptability ratings for clitic and no-clitic sentences. The third model includes the data from the two groups together to further examine whether the two languages are significantly different. All felicity ratings were analyzed using cumulative link mixed effects models (Christensen, 2015). All models include fixed effects for Clitic, Specificity, Anaphoricity and their interactions and random effects for Participant and Item. Adding the slopes for predictors to random effects for Participant and Item was based on a maximal model, following Barr, Levy, Scheepers, and Tilly (2013), using the most maximal model that allows convergence. Recall that Specificity cannot be included in the item slopes because Specificity was not a within-item factor. The fixed effects for the combined model included Clitic, Specificity, Anaphoricity, Group (Italian or Romanian), and all possible two-way interactions as well as three-way interactions with Clitic, Specificity and Group and Clitic, Anaphoricity and Group.

Table 4. Acceptability judgments from Italian and Romanian monolinguals and their comparison

Effects on acceptability judgment			
	L1 Italian	**L1 Romanian**	**Combined**
NoClitic.vs.Clitic	**0.67** (0.24)**	**1.33** (0.41)**	**0.84*** (0.21)**
Specific.vs.NonSpecific	−0.10 (0.24)	0.89 (0.50)	0.22 (0.26)
Topic.vs.Focus	−0.04 (0.19)	−0.49 (0.28)	−0.25 (0.16)
NoCl.vs.Cl:Spec.vs.NonSpec	**−0.92* (0.47)**	**−7.48*** (1.36)**	**−3.85*** (0.61)**
NoCl.vs.Cl:Top.vs.Foc	**−6.63*** (1.16)**	**−1.29* (0.51)**	**−3.77 (0.63)*****
Spec.vs.NonSpec:Top.vs.Foc	−0.48 (0.36)	−0.23 (0.90)	
NoCl.vs.Cl:Sp.vs.NonSp:T.vs.F	−0.00 (0.56)	−0.03 (0.11)	
L1Italian.vs.L1Romanian			−0.39 (0.41)
NoCl.vs.Cl:L1Ita.vs.L1Rom			0.35 (0.31)
Sp.vs.NonSp:L1Ita.vs.L1Rom			**0.79** (0.28)**
Top.vs.Foc:L1Ita.vs.L1Rom			**−0.46* (0.22)**
NoCl.vs.Cl:Sp.vs.NonSp:L1I.vs.L1R			**−5.5*** (1.06)**
NoCl.vs.Cl:T.vs.F:L1I.vs.L1R			**5.4*** (1.24)**

Note. All predictors are categorical and each line shows the levels that were contrasted. For example, NoClitic.vs.Clitic estimates how the acceptability of a sentence changes when a clitic is used as opposed to no clitic (the baseline).
*$p < .1$; **$p < .05$; ***$p < .01$.

For both languages, the ratings for clitic sentences are significantly higher than those for no-clitic sentences. As can be observed from Figure 2, the presence of a clitic in a context where it would typically not be used was more acceptable than omitting a clitic in a context where one was expected. That is, for Italian, it was more acceptable to have a clitic in the [–anaphor] condition than it was not to have a clitic in the [+anaphor] condition. Similarly for Romanian, it was more acceptable to have a clitic with non-specific objects than it was to omit a clitic with specific objects. This caused the overall higher rating of clitic sentences over no-clitic sentences.

As expected for Italian, there was a significant effect of Anaphoricity (Topic. vs.Focus) on the acceptability of clitic versus no-clitic sentences: as NoCl.vs.Cl:Top. vs.Foc shows, the strongest interaction was between Clitic and Anaphoricity ($p < .0001$). The table also shows a small interaction between Clitic and Specificity ($p < .05$). As can be seen in Figure 1, clitics in sentences with fronted specific foci were rated slightly higher than with non-specific foci. The interaction between Clitic and Anaphoricity was, however, much stronger than that between Clitic and Specificity. In fact, clitic sentences were rated as more acceptable than no-clitic sentences only in the [+anaphor] conditions. Data from Romanian monolinguals show a strong interaction between Clitic and Specificity ($p < .0001$), suggesting that the rating for clitic and clitic-less sentences was mainly driven by whether the fronted object was specific. There was also a small interaction between Clitic and Anaphoricity, most likely because clitic sentences with non-specific objects were rated as more acceptable when this object was a topic (see Figure 2). Thus, as predicted, the acceptability rating of clitic and no-clitic sentences was most strongly affected by Anaphoricity in Italian and by Specificity in Romanian.

To confirm that the two languages are significantly different, data from the two groups were compared directly. The model output shown in the last column of Table 4 provides information about whether the interaction between Clitic and Specificity and the interaction between Clitic and Anaphoricity differs between the two monolingual groups. The last two rows show that this was indeed the case. Specifically, there was a significant 3-way interaction ($p < .01$) between Clitic, Specificity and Group on line "NoCl.vs.Cl:Sp.vs.NonSp:L1I.vs.L1R", suggesting that the effect of Specificity on the acceptability of clitic versus no-clitic sentences differed strongly between Italian and Romanian. Similarly, the significant interaction ($p < .01$) between Clitic, Anaphoricity and Group shown with "NoCl.vs.Cl:T. vs.F:L1I.vs.L1R" confirms that Italian and Romanian differ in whether discourse has an effect on the acceptability of clitic and no-clitic sentences.

5.2.2.2 *L1 Romanian, L2 Italian*

Figure 3 shows the results of the acceptability judgment task by L1 Romanian speakers, divided in two groups based on proficiency level. Note that we are interested in whether L2 learners can distinguish between acceptable and non-acceptable sentences in a given discourse context. The Int/Adv speakers rated clitic sentences higher when the object was specific and no-clitic sentences higher when the object was non-specific, independently of anaphoricity. Their pattern of behavior was

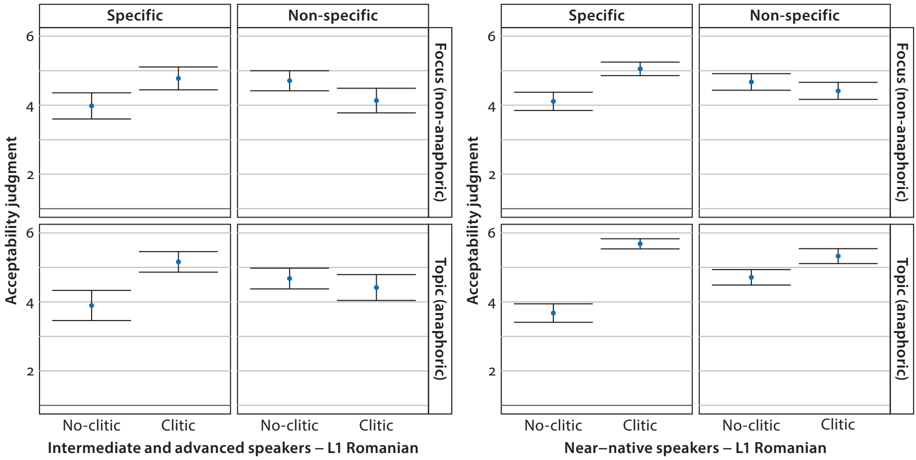


Figure 3. Acceptability judgments from L1 Romanian intermediate and advanced learners of Italian (left) and L1 Romanian near-native speakers of Italian (right)

similar to the Romanian monolinguals. Learners in the near-native group showed a target-like preference for sentences with clitics when the object was a discourse anaphor but, unlike Italian native speakers, they also preferred clitic sentences when the object was a specific focus. In this context, clitic sentences are preferred by the Romanian monolinguals in Romanian as well.

Table 5 shows the results of the cumulative link mixed effects models for the Int/Adv and the Near-Native group separately. For the Int/Adv group there was a significant interaction between Clitic and Specificity, suggesting that clitic sentences were preferred over no-clitic sentences when the fronted object was specific, regardless of Anaphoricity. No other interactions are significant for this group. Interestingly, for the near-native speakers, there is an almost equal effect of Specificity and

Table 5. Acceptability judgments from the L1 Romanian L2 Italian group

Effect on acceptability judgment		
	Int/adv	**Near native**
NoClitic.vs.Clitic	0.82 (0.52)	**2.03*** (0.41)**
Specific.vs.NonSpecific	−0.11 (0.25)	0.16 (0.29)
Topic.vs.Focus	−0.37 (0.32)	**−0.80** (0.29)**
NoClitic.vs.Clitic:Spec.vs.NonSpec	**−2.78* (1.23)**	**−2.52*** (0.51)**
NoClitic.vs.Clitic:Topic.vs.Focus	−0.94 (0.65)	**−2.65*** (0.54)**
Spec.vs.NonSpec:Topic.vs.Focus	0.14 (0.51)	−0.68 (0.46)
NoClitic.vs.Clitic:Spec.vs.NonSpec:Topic.vs.Focus	0.42 (0.86)	−0.07 (0.78)

Note. *$p < .1$; **$p < .05$; ***$p < .01$.

Anaphoricity, as the strong interactions between Clitic and Anaphoricity and Clitic and Specificity show. In addition, for this group there is a main effect of Clitic, because speakers in this group rated clitic sentences as more acceptable than no-clitic sentence in three out of four of the possible conditions.

5.2.2.3 *L1 English, L2 Italian*

Figure 4 shows the results of the acceptability judgment task by L1 English speakers. The left plot, representing data from the Int/Adv group, shows a high level of acceptance of target sentences overall and does not show a clear preference for clitic or no-clitic sentences in any of the conditions. The Near-Native speakers (right plot) also rated all sentences as highly acceptable but rated sentences with fronted topics ([+anaphor]) as more acceptable when there was a clitic than when there was no clitic, regardless of Specificity. For target trials where the fronted object was a contrastive focus ([–anaphor]), near-native speakers had no preference between clitic and no-clitic sentences.

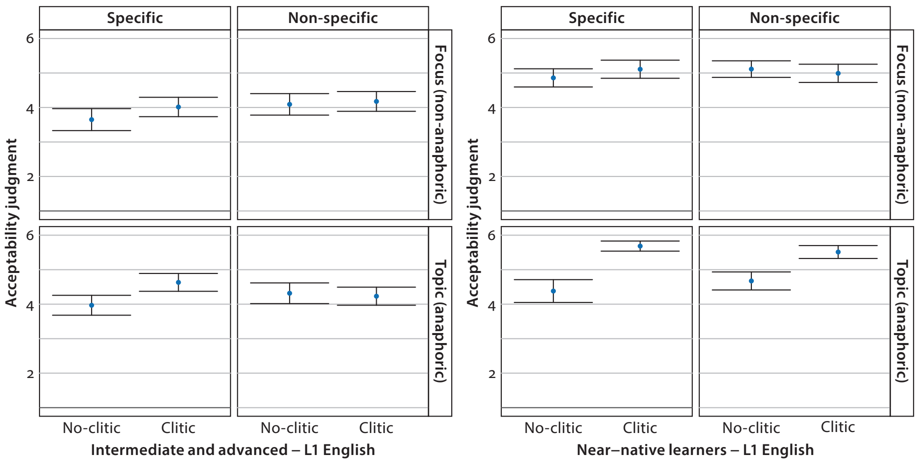


Figure 4. Acceptability judgments from L1 English intermediate and advanced learners of Italian (left) and L1 English near-native speakers of Italian (right)

Table 6 shows the results of the cumulative link mixed effects models for the Int/Adv and the Near-Native group separately. The model for the Int/Adv group shows a small but significant interaction between Clitic and Specificity, but not between Clitic and Anaphoricity, suggesting that clitics are rated as slightly more acceptable when the fronted object was specific than when it was non-specific (but note that this effect is not as strong as with the L1 Romanian speakers at the same proficiency level). Furthermore, the near-native speakers, but not the participants in the Int/Adv group, distinguished between clitic and no-clitic sentences with topical

Table 6. Acceptability judgments from the L1 English L2 Italian group

Effect on acceptability judgment		
	Int/adv	**Near native**
NoClitic.vs.Clitic	0.43 (0.68)	**1.47*** (0.40)**
Specific.vs.NonSpecific	0.80 (0.43)	−0.16 (0.29)
Topic.vs.Focus	0.43 (0.39)	−0.23 (0.27)
NoClitic.vs.Clitic:Spec.vs.NonSpec	**0.48* (0.50)**	−1.08 (0.62)
NoClitic.vs.Clitic:Topic.vs.Focus	0.62 (0.39)	**−2.68*** (0.73)**
Spec.vs.NonSpec:Topic.vs.Focus	0.14 (0.39)	0.37 (0.43)
NoClitic.vs.Clitic:Spec.vs.NonSpec:Topic.vs.Focus	−0.81 (0.56)	0.39 (0.69)

Note. *$p < .1$; **$p < .05$; ***$p < .01$.

dislocated objects. This can be observed from a significant interaction between Clitic and Anaphoricity, clitic sentences being preferred when the fronted object is a Topic ([+anaphor]).

5.3 Task 2: Written elicitation

5.3.1 *Materials and procedure*

In the written elicitation task, participants were asked to complete sentences that were partially left blank. The written elicitation task consisted of 20 target items, five of each in the four possible conditions as illustrated in (9)–(12), varying the factors [±specific] and [±anaphor]. The parts left blank aimed at eliciting a verb alone or a clitic and a verb. The example sentences show the context and question in the English translation and the target sentence in Italian. In the actual experiment, all contexts, questions, and answers were provided in Italian for the Italian monolinguals and the Italian L2 learners and in Romanian for the Romanian monolinguals.

(9) **[+specific, +anaphor]**
Livio is looking for someone who can take his granny's cat and dog as she can't take care of them anymore. Livio asks Silvia:
Q: Would you maybe want to adopt the cat or the dog?
A: Il gatto ___*lo adotterei/lo prenderei*___ volentieri, ma non abbiamo spazio per un cane.
The cat ___*CL adopt.1SG/CL take.1SG*___ happily, but I do not have space for a dog.

(10) [**+specific, −anaphor**]
Anna and Beatrice are talking about Lea and Gianni who recently got married. Anna says to Beatrice:
Q: They have visited the Virgin Islands if I remember correctly.
A: Le Maledive ___*hanno*___ visitator per il viaggio di nozze, non le isole Vergini.
The Maldives ___*have.3pl*___ visited for the honeymoon, not the Virgin Islands.

(11) [**−specific, +anaphor**]
Alessandra is in the library, but she isn't sure what she wants to read and she goes to the librarian to ask for recommendations. The librarian says:
Q: Would you like to read a book about airplanes or one about cars?
A: Un libro sugli aeroplani ___*lo leggerei*___ con piacere, ma le macchine non mi interessano.
A book about airplanes ___*CL.read.1sg*___ with pleasure, but I am not interested in cars.

(12) [**−specific, −anaphor**]
Elena will go shopping this weekend because she has a date. Her friend tries to be helpful and says:
Q: Weren't you looking for a black shirt? I saw some cute ones at Zara.
A: Un VESTITO nero ___*cerco*___, non una maglietta nera.
A black dress ___*search.1sg*___ not a black shirt.

The experimental items were interspersed with 2 control items and 32 filler items. The control items consisted of partitive indefinites[5] and the filler items doubled as a proficiency test for the second language learners and included various verb tenses, moods and aspects, prepositions, number and gender agreement on adjectives, direct object clitics, articles, and possessive adjectives.

5. Partitive indefinites such as "one of the books" are indefinites with a specific interpretation. They were included to ensure that the intended specificity effect was not a definiteness effect, as I used indefinites in the non-specific conditions and definites in the specific conditions. As predicted, results from the partitive indefinite control items were identical to those of the specific target items.

5.3.2 *Results and analysis*

In the Written Elicitation task, participants were expected to write a word to complete each sentence, which allowed for examining whether and in which conditions they would attach a clitic to the verb. Answers with a clitic were assigned the value '1' and answers without a clitic were assigned the value '0'. Responses where participants provided answers that were not verbs were removed from the analysis (4% for the monolingual Italian group, 5% for the monolingual Romanian group, 10% for the English learners of Italian and 7% for the Romanian learners of Italian).

5.3.2.1 *Italian and Romanian monolinguals*

As shown in Figure 5, the Italian monolinguals inserted a verb with a clitic in the object-topic sentences regardless of specificity and the Romanian native speakers chose a verb with a clitic in the object-specific sentences regardless of Anaphoricity. The results are therefore in line with those from the Acceptability Judgment task.

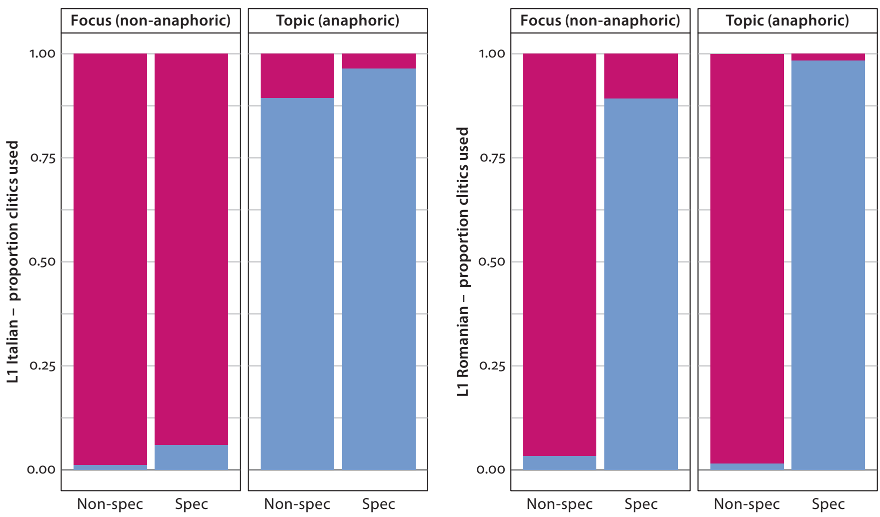


Figure 5. Proportion of clitics used in written elicitation task by Italian and Romanian monolinguals

Table 7 shows the results of binary logistic regression models predicting how the factors Anaphoricity and Specificity affected the use of clitics in Italian and in Romanian. The first model includes the data from the Italian monolingual group, the second from the Romanian monolingual group and the last model includes the combined data.

Table 7. Written elicitation from Italian monolinguals, Romanian monolinguals, and their comparison

Effects of specificity and discourse on use of clitic			
	L1 Italian	**L1 Romanian**	**Combined**
Intercept	0.65 (0.66)	−0.27 (3.43)	0.49 (0.47)
Topic.vs.Focus	9.08*** (1.95)	−0.53 (6.82)	5.08*** (1.15)
Specific.vs.NonSpecific	2.25* (1.05)	48.35*** (12.03)	5.96*** (1.21)
Topic.vs.Focus:Spec.vs.NonSpec	−0.18 (1.91)	−2.88 (18.75)	−1.08 (1.48)
L1Ita.vs.L1Rom	−0.17 (0.91)		
Topic.vs.Focus:L1Ita.vs.L1Rom	−8.49*** (2.21)		
L1Ita.vs.L1Rom:Spec.vs.NonSpec	7.66*** (1.93)		

Note. **p* < .1; ***p* < .05; ****p* < .01.

In line with what can be observed in Figure 4, the model shows a significant effect of Anaphoricity for the monolingual Italian group and a significant effect of Specificity for the monolingual Romanian group. For the L1 Italian speakers, there is also a small effect of Specificity. Crucially however, if we convert the estimates into odds we can get an idea of what the model predicts the odds to be that Italian monolinguals use a clitic depending on Specificity or depending on Anaphoricity. We can convert the estimate in an odds ratio ($e^{9.08} = 8777.96$) and see that the model estimates that the odds of using a clitic when the object is a Topic ([+anaphor]) is about 8778 times higher than when the object is a Focus ([−anaphor]). The odds of using a clitic when the object is specific is only about 9 times higher ($e^{2.25} = 9.48$) than when it is non-specific. Compare this to the monolingual Romanian group, where the odds of choosing a clitic is much higher ($e^{43.35} = 6.71^{e+18}$) when the object is specific compared to when it is non-specific. The significant interactions with Anaphoricity and Group and Group and Specificity shown in the last two lines of the third model confirm again that the two languages are significantly different.

5.3.2.2 *L1 Romanian, L2 Italian*

Figure 6 shows the proportion of Italian clitics used by L1 Romanian speakers. Speakers in the Int/Adv proficiency level use clitics mainly when the fronted object is specific. The pattern of behavior for the near-native speakers looks comparable to that of the speakers in the Int/Adv group in that specificity plays a role. One crucial difference is the increased use of clitics with non-specific topics, suggesting that these learners know that clitics are obligatory in Italian in this context, in contrast to Romanian.

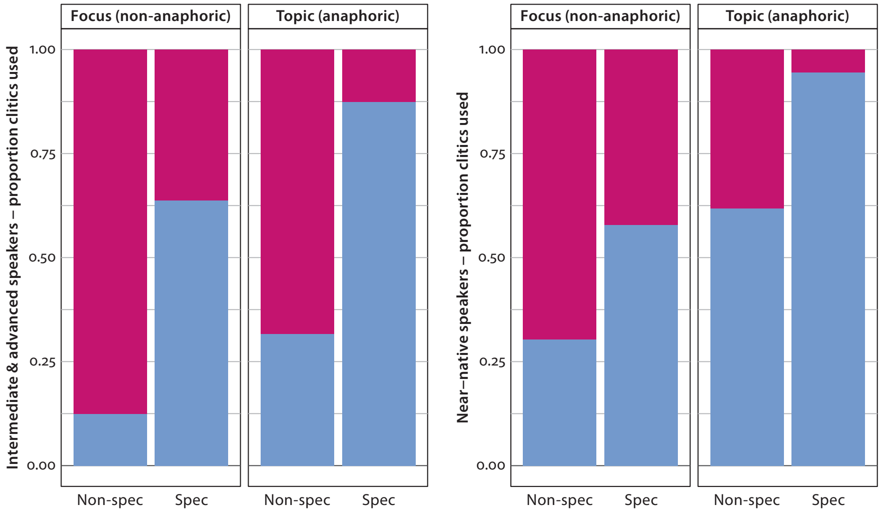


Figure 6. Proportion of clitics used in written elicitation task by L1 Romanian, L2 Italian speakers

Table 8 shows the results from binary logistic regression models reporting the data from the Romanian Int/Adv and Near-Native learners of Italian, respectively. The model outcomes show that Specificity, but not Anaphoricity, had a significant effect on the use of clitics for speakers in the Int/Adv group. This pattern is similar to the one observed for the Romanian monolinguals. For the near-native speakers, however, both Specificity and Anaphoricity had a significant effect. In fact, as Figure 6 clearly shows, clitics were used more when the fronted object was a topic, whether specific or non-specific, but they also used clitics when the object was a specific focus.

Table 8. Written elicitation results from L1 Romanian L2 Italian speakers, divided by proficiency level

Effects of specificity and discourse on use of clitic		
	L1 Rom – Int/Adv	**L1 Rom – Nn**
Intercept	0.36 (0.37)	−0.71 (0.49)
Topic.vs.Focus	1.10 (0.58)	**2.97*** (0.71)**
Specific.vs.NonSpecific	**3.11*** (0.66)**	**2.92*** (0.63)**
Topic.vs.Focus:Spec.vs.NonSpec	−0.67 (1.10)	0.26 (1.07)

Note. **p* < .1; ***p* < .05; ****p* < .01.

5.3.2.3 *L1 English, L2 Italian*

Figure 7 shows the proportion of Italian clitics used by L1 English speakers in the Written Elicitation task. The left graph shows that participants in the Int/Adv group did not consistently use clitics in any condition. In particular, Anaphoricity had no effect. Rather, they tended to use clitics more when the fronted object was specific regardless of Anaphoricity. The near-native speakers used clitics more when the object was [+anaphor] than when it was [–anaphor], suggesting sensitivity to the discourse status of the fronted object, unlike the Int/Adv group.

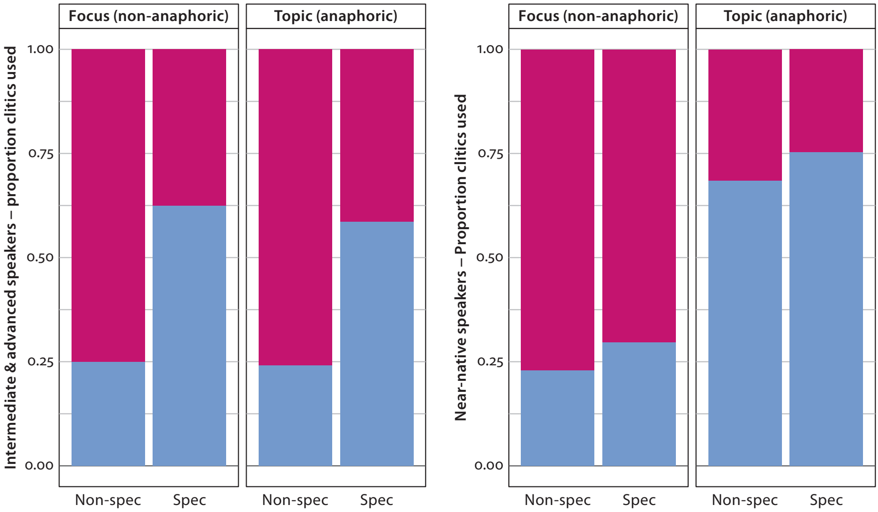


Figure 7. Proportion of clitics used in written elicitation task by L1 English, L2 Italian speakers

Table 9 shows the results of binary logistic regression models predicting how the factors Anaphoricity and Specificity affected the use of clitics for each of the proficiency levels. As can also be observed from the plot, there was a significant effect

Table 9. Written elicitation results from L1 English L2 Italian speakers, divided by proficiency level

Effects of specificity and discourse on use of clitic		
	L1 Eng – Int/Adv	**L1 Eng – NN**
Intercept	0.76 (0.69)	0.38 (0.60)
Topic.vs.Focus	–0.08 (0.51)	**3.56*** (0.85)**
Specific.vs.NonSpecific	**2.95*** (0.69)**	0.60 (0.65)
Topic.vs.Focus:Spec.vs.NonSpec	0.82 (1.09)	0.26 (1.07)

Note. **p* < .1; ***p* < .05; ****p* < .01.

of specificity for the Int/Adv group: clitics were used more when the fronted object was specific than when it was non-specific. The near native speakers showed an effect of Anaphoricity, choosing a clitic significantly more often when the fronted object was a Topic than when it was a Focus.

6. Discussion

This experimental study looked at the role of the L1 in the acquisition of discourse constraints on Italian CLLD. The experiment was constructed to examine whether reassembling discourse features (L1 Romanian, L2 Italian) is more difficult than acquiring discourse restrictions on CLLD without prior L1 knowledge of this construction (L1 English, L2 Italian). English learners of Italian have to acquire the syntax of CLLD and restrict its use to the correct discourse contexts, namely those with fronted [+anaphor] objects. Following the FRH, it was predicted that Romanian learners of Italian have to reassemble the relevant features associated with CLLD by removing the [±specific] feature that was observed for Romanian monolinguals and adding a [±anaphor] feature as was observed for Italian monolinguals.

The results for the L1 English Intermediate/Advanced group indicate that while they allowed sentences with clitics, they had not yet figured out the discourse restrictions on its use as they did not restrict clitics to dislocated [+anaphor] objects. The near-native speakers, on the other hand, had acquired the target restriction on CLLD, as predicted; in the Acceptability Judgment task they rated sentences with clitics higher than sentences without clitics when the fronted object was [+anaphor] and in the Written Elicitation task they used them significantly more with fronted [+anaphor] objects.

The results from the Romanian learners of Italian were in line with the FRH, as they showed convincingly that these learners initially mapped their L1 features onto Italian CLLD. Specifically, the Romanian learners of Italian in the Int/Adv group accepted clitics and used clitics significantly more when the fronted object was specific, independently of whether the fronted object was discourse anaphoric. The Romanian learners did not completely overcome this L1 specificity effect; even the near-native speakers showed a preference for clitic sentences over no-clitic sentences when the fronted object was a specific focus. This is the preferred pattern for Romanian, but not for Italian. Unlike the Int/Adv group, the near-native speakers correctly accepted and used clitics with non-specific topics. In other words, the L1 Romanian group eventually learned that clitics can be used with non-specific topics, but they continued to accept and use clitics when the fronted object was a specific focus.

Focusing now only on the results from the near-native speakers of Italian, we can answer the research question by concluding that acquiring the [+anaphor] value associated with Italian CLLD is more complicated when this involves reassembly of L1 features, as is the case for Romanian learners of Italian, than for speakers without prior L1 knowledge, as is the case for English learners of Italian. Specifically, the results suggest that the Romanian learners of Italian correctly removed the [+specific] in the conditions with Topic Fronting and correctly rejected clitics with Focus Fronting ([−anaphor]) when the object was [−specific]. These learners did not, however, acquire the [±anaphor] feature of Italian CLLD, as they allowed clitics in [+specific, −anaphor] conditions. In other words, they allowed and used clitics in those situations where either L1 or L2 uses the clitic.

In the remaining part of this section, I aim to discuss some findings that extend beyond the research questions of this paper but are nevertheless interesting for further research. Recall that for the English Int/Adv group, there was a small effect of specificity in both tasks (Note: this effect was not as strong as for the Romanian Int/Adv group). This is possibly a result of their L1, because English pronouns, which are the closest equivalent to clitics, seem to be restricted by specificity similarly to Romanian clitics, as discussed in Section 2.3. To illustrate, the example in (14) shows that it is infelicitous to use the pronoun *it* to refer to the non-specific NP *a red skirt*. It is possible that the English learners of Italian carried over this property to L2 Italian.

(14) I am looking for a red skirt. Do you know where I can find *it/ one.

It should furthermore be mentioned that although the English near-native speakers of Italian showed a native-like knowledge of the use of CLLD, their behavior was less categorical than that of the native Italian speakers. Namely, clitics were not consistently used in Topic Fronting and omitted in Focus Fronting constructions in the Written Elicitation task and there was also a higher rating of infelicitous sentences in the Acceptability Judgment task. Similarly, the responses from the Romanian learners of Italian were less categorical than those from Italian monolinguals. In other words, we observe less consistent use of clitics by second language learners compared to native speakers for both L1 groups. These results are in line with the idea proposed by Sorace (2011) that even very experienced L2 learners exhibit optionality due to the fact that the integration of information across different interfaces places higher demands on the processing mechanism. In consequence, values of the feature [±anaphor] may not always be appropriately mapped to the syntax.

Another remarkable observation is that in the Acceptability Judgment task (but not in the Written Elicitation task), the English near-native speakers of Italian did not show a significant preference for no-clitic sentences over clitic sentences in contexts with non-anaphoric objects (Focus Fronting). These results differ from those reported for the English near-native speakers of Spanish in Slabakova et al. (2012), who also tested knowledge of discourse constraints on the use of Spanish CLLD using an Acceptability Judgment task. Their L2 learners in the near-native stage rated no-clitic sentences as significantly more acceptable than clitic sentences with Focus Fronting. A slight difference in methodology can potentially explain these results. Namely, in Slabakova et al. (2012), the participants rated a sentence with and one without a clitic within the same trial, while in the current study participants judged the sentence with the clitic and the one without in the same discourse condition at different moments of the experiment. Slabakova et al.'s (2012) methodology allowed for direct comparison of clitic and no-clitic sentences, making it explicit to the participant what the relevant syntactic factor was that the researchers were interested in. This may have made the presence versus absence of the clitic more salient, facilitating the experimental task.

7. Conclusion

The findings presented in this paper shed light on the role of L1 transfer in the L2 acquisition of discourse-syntax mappings. Persistent L1 effects in near-native grammars were found only for the L1 Romanian speakers: the English near-native speakers of Italian correctly associated CLLD with fronting of [+anaphor] objects, while the results for the Romanian near-native speakers of Italian showed an L1 specificity effect. In particular, it appears to be difficult for Romanian L2 learners of Italian to unlearn that a clitic is not allowed with fronted [+specific, −anaphor] objects on the basis of absence of clitics in such contexts in the Italian input. In consequence, they were unable to reassemble the features associated with CLLD from the L1 [+specific] feature to L2 [+anaphoric] feature. Therefore, feature reassembly, which is required for the Romanian group, seems to be more difficult than acquiring the discourse features associated with CLLD from scratch.

References

Avram, L., & Martine C. (1999). Null objects and accusative clitics in Romanian. *Bucharest Working Papers in Linguistics* IV (1), 88–106.

Barr, D., Roger, L., Scheepers, C., & Tilly, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bianchi, V., Bocci, G., & Cruschina, S. (2013). Focus fronting and its implicatures. In E. O. Aboh, J. Schaeffer, & P. Sleeman (Eds.), *Romance languages and linguistic theory 2013: Selected papers from Going Romance, Amsterdam* 2013. John Benjamins. https://doi.org/10.1075/rllt.8.01bia

Bohnacker, U., & Rosen, C. (2008). The clause-initial position in L2 German declaratives: Transfer of information structure. *Studies in Second Language Acquisition*, 30(4), 511–538. https://doi.org/10.1017/S0272263108080741

Chomsky, N. (2000). Minimalist inquiries: The framework. In R. Martin, D. Michaels, & J. Uriagereka (Eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik* (pp. 89–156). The MIT Press.

Christensen, R. H. B. (2015). Analysis of ordinal data with cumulative link models – Estimation with the r-package ordinal. Retrieved on 1 March 2022 from http://people.vcu.edu/~dbandyop/BIOS625/CLM_R.pdf

Cornilescu, A., & Dobrovie-Sorin, C. (2008). Clitic doubling, complex heads and interarboreal operations. In L. Tasmowski & D. Kallulli (Eds.), *Clitic doubling in the balkan languages* (pp. 289–319). John Benjamins. https://doi.org/10.1075/la.130.16cor

Dobrovie-Sorin, C. (1990). Clitic doubling, Wh-movement, and quantification in Romanian. *Linguistic Inquiry*, 21(3), 351–398.

Donaldson, B. (2011). Left-dislocation in near-native French. *Studies in Second Language Acquisition*, 33, 399–432. https://doi.org/10.1017/S0272263111000039

Frey, W. (2006). Contrast and movement to the German prefield. In V. M. S. Winkler (Ed.), *Architecture of Focus*. Mouton de Gruyter.

Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research*, 30(3), 251–266. https://doi.org/10.1023/A:1010490917575

Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research*, 22(3), 369–397. https://doi.org/10.1191/0267658306sr272oa

Hopp, H. (2009). The syntax-discourse interface in near-native L2 acquisition: Off-line and on-line performance. *Bilingualism: Language and Cognition*, 12(4), 463–483. https://doi.org/10.1017/S1366728909990253

Ivanov, I. (2009). Second language acquisition of Bulgarian object clitics: A test case for the interface hypothesis (Unpublished doctoral dissertation). University of Iowa. https://doi.org/10.17077/etd.077u404q

Ivanov, I. (2012). L2 acquisition of Bulgarian clitic-doubling: A test case for the interface hypothesis. *Second Language Research*, 28, 345–368. https://doi.org/10.1177/0267658312452066

Jackendoff, R. (1972). *Semantic interpretation in generative grammar*. The MIT Press.

Kiss, K. E. (1998). Identical focus versus information focus. *Language*, 74, 245–273. https://doi.org/10.1353/lan.1998.0211

Laleko, O., & Polinsky, M. (2016). Between syntax and discourse. *Linguistic Approaches to Bilingualism*, 6(4), 396–439. https://doi.org/10.1075/lab.14018.lal

Lardiere, D. (2008). Feature assembly in second language acquisition. In H. Zobl, J. M. Liceras, & H. Goodluck (Eds.), *The role of formal features in second language acquisition* (pp. 106–140). Lawrence Erlbaum Associates.

Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research*, 25, 173–227. https://doi.org/10.1177/0267658308100283

Leal, T., & Slabakova, R. (2019). The relationship between L2 instruction, exposure and the acquisition of a syntax-discourse property in L2 Spanish. *Language Teaching Research*, 23(2), 237–258. https://doi.org/10.1177/1362168817745714

López, L. (2009). *A derivational syntax for information structure*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199557400.001.0001

McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, 55, 381–401. https://doi.org/10.1016/j.jml.2006.06.006

Pierrehumbert, J. (1980). The phonetics and phonology of English intonation (Unpublished doctoral dissertation). MIT.

Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the full transfer/full access model. *Second Language Research*, 12, 40–72. https://doi.org/10.1177/026765839601200103

Slabakova, R., & García Mayo, M. P. (2015). The L3 syntax-discourse interface. *Bilingualism: Language and Cognition*, 18(2), 206–226. https://doi.org/10.1017/S1366728913000369

Slabakova, R., & García Mayo, M. P. (2017). Testing the current models of third language acquisition. In T. Angelovska & A. Hahn (Eds.), *L3 syntactic transfer models, new developments and implications* (pp. 63–84). John Benjamins. https://doi.org/10.1075/bpa.5.04sla

Slabakova, R., & Ivanov, I. (2011). A more careful look at the syntax-discourse interface. *Lingua*, 121, 637–651. https://doi.org/10.1016/j.lingua.2010.05.003

Slabakova, R., Kempchinsky, P., & Rothman, J. (2012). Clitic-doubled left dislocation and focus fronting in L2 Spanish: A case of successful acquisition at the syntax-discourse interface. *Second Language Research*, 28(3), 319–343. https://doi.org/10.1177/0267658312447612

Smeets, L. (2018). The acquisition of object movement in Dutch: L1 transfer and ultimate attainment at the syntax-discourse interface. *Second Language Research*, 35(4), 479–504. https://doi.org/10.1177/0267658318782357

Smeets, L. (2019). Conditions on L1 transfer in L2 discourse-syntax mappings: The case of Clitic Left Dislocation in Italian and Romanian (Unpublished doctoral dissertation). McGill University.

Sorace, A. (2011). Pinning down the concept of 'interface' in bilingualism. *Linguistic Approaches to Bilingualism*, 1, 1–33. https://doi.org/10.1075/lab.1.1.01sor

Sorace, A., & Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Language Research*, 22, 339–368. https://doi.org/10.1191/0267658306sr2710a

Sorace, A., & Serratrice, L. (2009). Internal and external interfaces in bilingual language development: Beyond structural overlap. *International Journal of Bilingualism*, 13(2), 195–210. https://doi.org/10.1177/1367006909339810

Tsimpli, I. M., & Sorace, A. (2006). Differentiating Interfaces: L2 performance in syntax-semantics and syntax-discourse phenomena. *BUCLD Proceedings*, 30, 653–664.

Valenzuela, E. (2005). L2 ultimate attainment and the syntax-discourse interface: the acquisition of topic constructions in non-native Spanish and English (Unpublished doctoral dissertation). McGill University.

Exploring extra-linguistic factors and their impact on L2 acquisition

Interference-based and capacity-based approaches to working memory in second language sentence processing

Ian Cunnings
University of Reading

The role that working memory may play in explaining potential differences between native and non-native sentence processing has been increasingly debated. In this chapter, I discuss how the conceptualisation of working memory is crucial to our understanding of its role in second language processing. In particular, I compare capacity-based approaches that focus on working memory resources and interference-based approaches that focus on memory encoding and retrieval. After reviewing evidence for both approaches to working memory, I argue that interference-based accounts provide a promising approach for examining the role of working memory in second language processing. Although I focus on non-native sentence processing, I also touch on related issues in second language acquisition.

Keywords: second language processing, second language acquisition, working memory, similarity-based interference, memory capacity

1. Introduction

There has long been consideration of the link between working memory, broadly defined as the ability to store and manipulate a limited amount of task-relevant information at one time (Baddeley, 2007; Cowan, 2017), and research in linguistics. Indeed, how grammaticality interacts with memory limitations has been debated since the early days of generative syntax (Chomsky & Miller, 1963; Hofmeister et al., 2013; Hofmeister & Sag, 2010; Sprouse, Wagers, & Phillips, 2012). One often cited example of how memory limitations interact with sentence comprehension is the decrease in acceptability observed as the number of centre-embeddings within a sentence increases, as in (1a)–(c).

https://doi.org/10.1075/lald.67.10cun

(1) a. The cat was scared.
b. The cat the dog chased was scared.
c. The cat the dog the boy saw chased was scared.

The role that working memory plays in second language acquisition (SLA) and processing has also been widely debated (Cunnings, 2017a, 2017b; Harrington & Sawyer, 1992; Juffs & Harrington, 2011; Linck, Osthus, Koeth, & Bunting, 2014; McDonald, 2006; Wen, Mota, & McNeill, 2015). In research on native (L1) and non-native (L2) sentence comprehension, the similarities and differences between L1 and L2 processing are contested (e.g., Clahsen & Felser, 2006, 2018; Cunnings, 2017a, 2017b; Hopp, 2018; McDonald, 2006). An important question in this field has been whether putative differences between L1 and L2 processing can be neutralised when individual differences in working memory are considered (e.g., McDonald, 2006).

Crucial in examining the link between working memory and L2 sentence processing is the conceptualisation of working memory itself. In this chapter, I discuss how different ways of conceptualising working memory lead to very different predictions about how it may modulate L2 sentence processing. I will contrast, in particular, "capacity-based" accounts of working memory during sentence processing with other "interference-based" models. Capacity-based models (Caplan & Waters, 1999; Daneman & Carpenter, 1980; Just & Carpenter, 1992) typically describe memory resource limitations during sentence processing in terms of the *amount* of information that can be stored in memory at one time. Alternatively, interference-based models (e.g., Lewis, Vasishth, & Van Dyke, 2006; McElree, 2000; McElree, Foraker, & Dyer, 2003; Van Dyke & Johns, 2012; Vasishth et al., 2019) focus instead on the *content* of information in memory, rather than the quantity per se. I argue that interference-based accounts, which focus on how information is encoded and retrieved from memory during real-time language processing, provide a well-motivated framework for comparing L1 and L2 sentence processing. Although I focus on L1 and L2 sentence processing, I also touch upon how this characterisation of working memory can help bridge research in L2 acquisition and L2 processing.

This chapter is outlined as follows. In Section 2, I discuss different accounts of working memory, with a focus on how capacity-based models and interference-based models can be applied to L2 sentence processing. Section 3 provides discussion of studies that have tested the capacity-based view, focusing on work that has used complex span tasks to measure individual differences in memory capacity, along with a discussion of how the interference-based account can be applied to research in L2 sentence processing. Section 4 provides a summary and conclusion.

2. Working memory in L2 sentence processing

Although broadly defined as the ability to store and manipulate a limited amount of information, different theories conceptualise working memory in different ways. Some accounts assume a separate working memory component distinct from long-term memory. The multi-component model of working memory (e.g., Baddeley, 2007, 2015) is a highly influential model in this regard. Other accounts do not assume a separate working memory component, and instead posit that working memory is instead the activated portion of memory (e.g., Cowan, 1988, 2017). Others describe working memory in terms of executive attention, and the ability to allocate attentional resources to task-relevant information (e.g., Engle, 2002, 2018).

In research on sentence processing, these different conceptualisations of working memory make different predictions about how working memory should influence successful comprehension. Below, I discuss how these different approaches attempt to account for differences between L1 and L2 processing.

2.1 Capacity-based approaches to L2 sentence processing

Perhaps the dominant model of working memory in L2 sentence processing research is the capacity-based view (e.g., Harrington & Sawyer, 1992; McDonald, 2006). According to capacity-based accounts, individuals have a limited pool of cognitive resources that can be used to maintain task-relevant information. This view is most consistent with models that assume a separate working memory component that is distinct from long-term memory. From this perspective, the capacity of this working memory component is hypothesised to vary between individuals.

Daneman and Carpenter (1980) were amongst the first to propose that individual differences in working memory capacity are related to individual differences in language comprehension in L1 readers. Applying this to L2 learners, Harrington and Sawyer (1992) argued that individual differences in working memory capacity predict individual differences in L2 reading. This approach predicts that working memory capacity is taxed to a greater extent during L2 processing than L1 processing, because L2 processing is slower, more effortful and generally more cognitively demanding. For example, Dekydtspotter and Renaud (2014) argued that generally slower processing in the L2 may lead to syntactic representations fading more quickly in a capacity-limited working memory. Such accounts would predict that L1 and L2 processing should be similar for L2 learners with high enough memory capacity or processing speed (e.g., McDonald, 2006).

Different approaches have described individual differences in working memory capacity in different ways. For example, Daneman and Carpenter (1980) hypothesised that sentence comprehension and other cognitive abilities utilise a shared domain-general working memory capacity, while others have argued that sentence processing has a dedicated, domain-specific pool of memory resources (e.g., Caplan & Waters, 1999). In addition to the question of whether sentence processing utilises a domain-general or domain-specific set of memory resources, another issue relates to precisely what types of information tax memory resources. Capacity-based approaches could be described in terms of the raw amount of information (i.e., words or sentences) that can be actively maintained at once, in terms of the number of different syntactic parses that may be maintained at once, or in terms of the number of different types of information sources that a reader can utilise at once (e.g., Daneman & Carpenter, 1980; Just & Carpenter, 1992; Just, Carpenter, & Keller, 1996). For example, Just and Carpenter (1992) claimed that high-capacity readers take multiple sources of syntactic and non-syntactic information into consideration when resolving syntactic ambiguity and are able to consider multiple possible parses of an ambiguous input. The hypothesis that high-capacity readers can take into account multiple sources of information is reminiscent of research in SLA, where the Interface Hypothesis (Sorace, 2011; Sorace & Filiaci, 2006) proposes that L2 learners have difficulty integrating information between narrow syntax and other cognitive domains. In the most recent instantiation of the Interface Hypothesis, this difficulty is proposed to result from L2 difficulty in integrating multiple information sources during processing (Sorace, 2011). While L2 difficulties with anaphora resolution are well documented as evidence for this hypothesis (see Section 3.2), note that some studies have suggested that L2 learners may be *more* likely to take into consideration discourse-level information when resolving syntactic ambiguities than L1 readers (Pan & Felser, 2011; Pan, Schimke, & Felser, 2015). This is unexpected under a capacity-based view.

The discussion above highlights how different conceptualisations of memory 'capacity' lead to different predictions for L2 processing. In observing any L1/L2 differences, it is of course trivially true that L2 speakers can be said to lack some type of processing 'capacity' (in the broadest sense of the term) in comparison to L1 speakers. Precise characterisation of what is thought to tax memory 'resources' is required here to adequately test the capacity-based view of L2 processing (for critical discussion of the notion of resources, see Navon, 1984).

2.2 Interference-based approaches to L2 sentence processing

Other models of working memory, which typically do not posit a dedicated working memory component, focus on how attention is allocated to the encoding and retrieval of task-relevant information (e.g., Engle, 2002, 2018). Applying this type of model to sentence processing, successful language comprehension will involve encoding and retrieving information from memory that is relevant to the task of language comprehension.

The cue-based parsing framework has been developed as an account of these memory encoding and retrieval operations required for language comprehension (Lewis & Vasishth, 2005; Lewis et al., 2006; McElree, Foraker, & Dyer, 2003; Van Dyke & Johns, 2012; Vasishth et al., 2019). As an example of how memory encoding and retrieval is required during sentence processing, consider a filler-gap dependency as in (2a) (adapted from Cunnings & Sturt, 2018). Here, the displaced filler ('the book') needs to be interpreted as the direct object of the verb 'read,' even though they are non-adjacent (compare to 'The boy with the coffee very happily *read the book* before lunch'). According to cue-based parsing, a representation of 'the book' is encoded in memory when it is first encountered and then stored while the other constituents are processed (and themselves encoded in memory). At 'read', a representation of 'the book' needs to be retrieved from memory to correctly interpret the sentence.

(2) a. It was the book that the boy with the coffee very happily read before lunch.
b. It was the book that the boy with the magazine very happily read before lunch.

Memory retrieval within this framework is achieved via a direct-access memory retrieval mechanism that compares a set of retrieval cues against all items in memory in parallel, with the best matching item being retrieved (e.g., McElree, Foraker, & Dyer, 2003). In sentences like (2a), this will likely consist of a series of syntactic and semantic cues. For example, a syntactic cue, or set of cues, will be required to guide retrieval to the correct constituent in the syntactic structure. For simplicity sake, I refer to this here as [+OBJECT]. A semantic cue, or cues, will also guide retrieval to a constituent that can plausibly act as the direct object of the verb. In this case, and again simplifying somewhat, this might be [+READABLE]. Upon encountering the verb, items in the sentence that match these retrieval cues become activated, and the item that becomes most activated is retrieved. In (2a), these combined cues uniquely identify 'the book' as the target for memory retrieval, which will become most activated. However, in (2b), although 'the book' provides the best match to this set of retrieval cues, 'the magazine' provides a partial match, as it matches the semantic but not syntactic cue and as such will become partially activated. Cue-based parsing predicts that this *distractor* constituent may be retrieved a proportion of the time, as a result of its partial match. This is known as similarity-based interference (Lewis, Vasishth, & Van Dyke, 2006; Van Dyke & Johns, 2012; Vasishth et al., 2019), as successful comprehension is dependent on the number of items that match a set of retrieval cues.

Different types of sentences will pose different challenges from this interference-based perspective. While syntactic and semantic retrieval cues may combine to resolve filler-gap dependencies, other types of retrieval cues will be required elsewhere. For example, anaphora resolution may involve cues related to a pronoun's number and gender or related to discourse prominence (see Section 3.2). Although different types of retrieval cues will thus be required for successful comprehension in different circumstances, the underlying prediction of this interference-based approach is that successful comprehension is dependent on how a reader or listener is able to encode and retrieve task-relevant information.

As there are different instantiations of the capacity-based view of sentence processing, there is also debate regarding the precise characterisation of interference-based accounts (for discussion, see Parker, Shvartsman, and Van Dyke, 2017; Vasishth et al., 2019). I do not attempt to examine this issue here. For present purposes, the main point is that these accounts all predict similarity-based interference to be a function of the similarity between a set of retrieval cues and the number of constituents that match, or partially match, these cues. Thus, unlike the capacity-based view, which focuses on the raw *quantity* of information that needs to be maintained, interference-based accounts focus on the *quality* and *content* of information in memory as the primary determinant of successful comprehension (Van Dyke & Johns, 2012).

L1/L2 differences from this perspective would be described in terms of how L1 speakers and L2 learners encode and retrieve information from memory during processing (Cunnings, 2017a, 2017b). Cunnings proposed that L1/L2 differences can at least in part be explained in terms of how native and non-native speakers utilise different cues to guide memory retrieval. For example, if an L2 learner were to weight the semantic cue over the syntactic cue more heavily than L1 readers in (2b), L2 readers may be more prone to interference from the distractor ('the magazine').

In teasing apart capacity-based and interference-based accounts of working memory in sentence processing, it is important to consider that differences in capacity and interference are often confounded (Van Dyke & Johns, 2012). For example, returning to (1), in (1a) the predicate 'was scared' is adjacent to the sentence subject ('the cat'), while in (1c) they are separated by several other constituents. The capacity-based view would explain the comparative difficulty of (1c) in terms of a reader's limited memory capacity being exceeded. However, as there are more potentially interfering constituents in (1c) than (1a), such that 'the dog' and 'the boy' could be subjects of 'was scared', cue-based retrieval can also explain the difficulty in (1c) in terms of interference. Thus, the two accounts both predict difficulty in (1c), but for very different reasons, and teasing apart the capacity-based and interference-based view can be difficult, as increasing memory load, like in longer dependencies, also increases interference. However, it is less clear how a capacity-based view could explain differences between sentences like (2a) and (2b), as the number of words in each sentence is the same; it is the *content* of the sentences that differs.

Similar issues arise in disentangling capacity-based and interference-based approaches to working memory in L2 processing. Although some studies have been taken as evidence of capacity-based L2 limitations, whether these studies provide veridical evidence for capacity-based accounts of L2 processing over interference-based models is less clear. I argue that the interference-based account provides a well-specified framework that explicitly ties working memory operations with real-time sentence processing, and a fruitful way of conceptualising individual differences in L1 and L2 sentence processing.

3. Characterising and measuring L2 individual differences

Important to the capacity-based view is the measurement of individual differences in memory capacity. Measuring individual differences in working memory is often achieved using one or more complex span task, which involve a participant remembering a list of items while completing a secondary task (see e.g., Conway et al., 2005; Mathy, Chekaf, & Cowan, 2018). A commonly used complex span task in L2 processing research is reading span, first developed by Daneman and Carpenter (1980), and adapted for L2 learners by Harrington and Sawyer (1992). In this task, participants must remember a series of words for subsequent recall while reading a series of sentences (which are often judged for grammaticality or plausibility). From a capacity-based view, this task is taken to measure the amount of information an individual is able to maintain at one time, with higher span readers presumably being able to hold more information in memory at once than lower span readers. From the perspective of L1/L2 differences during sentence processing, a higher span would suggest increased memory capacity and more cognitive resources that presumably can be utilised during L2 processing. As such, higher span L2 learners may behave more nativelike. Below, I discuss how complex span tasks have been used to test this capacity-based view.

3.1 Memory capacity in L2 processing

Studies have examined the correlation between performance on complex span tasks and L2 sentence processing. Here, however, the results have been mixed. Some studies have tested the relationship between span scores and the processing of filler-gap dependencies (Dallas, DeDe, & Nicol, 2013; Dussias & Piñar, 2010; Felser & Roberts, 2007; Juffs, 2005; Miller, 2014). Dallas, DeDe, and Nicol (2013) examined sentences like (3), manipulating a displaced filler to be either a plausible ('which player') or implausible ('which football') direct object of the critical verb ('threatened'), using an event-related potential (ERP) paradigm. They expected implausible sentences to yield an increased N400 component (Kutas & Hillyard, 1980) compared to plausible sentences.

(3) The umpire asked which player/football the coach threatened before the game.

While L1 readers displayed the expected N400 effect, a group of Chinese L2 English learners did not. However, an individual differences analysis suggested that L2 learners with high English proficiency did show an N400. A composite score of memory span did not however correlate with the L2 learners' ERP responses. Other studies testing filler-gap dependencies using different paradigms also failed to observe an effect of memory span (Felser & Roberts, 2007; Juffs, 2005; Miller, 2014). One counterexample is Dussias and Piñar (2010), who examined filler-gap dependencies in a reading study. Here, high span L2 learners behaved nativelike, while low span L2 learners did not. The effects that Dussias and Piñar observed, however, were related to how filler-gap dependencies may be revised during processing, rather than how they are initially resolved. For example, in (4), readers may initially interpret 'who' to be the direct object of 'know' when it is in fact the subject of 'killed' and reading span was found to correlate with this aspect of processing in the L2 group. This might suggest a particular role of memory capacity in reanalysis during L2 processing.

(4) Who did the policeman know killed the pedestrian?

This potential interpretation of these findings is complicated by the fact that studies examining other types of syntactic ambiguity have not found significant correlations between memory capacity and reanalysis during L2 processing (Juffs, 2004). Other studies have reported effects of memory span in some measures of ambiguity resolution but not others. For example, Hopp (2014) found significant correlations between reading span and L2 performance in an offline task on ambiguous relative clause attachment, but did not find significant correlations between reading span and relative clause attachment in an eye-tracking during reading task that examined online sentence processing.

Correlations between span scores and sentence processing have also been investigated in work examining morphosyntactic agreement. There is a long line of research in L2 processing of agreement phenomena, with some studies suggesting nativelike performance and others not (for review, see Cunnings, 2017a). Some researchers have examined the extent to which working memory may explain this discrepancy. Sagarra and Herschensohn (2010) tested Spanish sentences like (5), manipulating the extent to which an adjective ('*famoso*') agreed in either number or gender with a preceding noun ('*prototipo*').

(5) *El ingeniero presenta el prototipo famoso/famosa/famosos en la conferencia.*
'The engineer presented the prototype$_{\text{-MASC-SING}}$ famous$_{\text{-MASC-SING}}$ / famous$_{\text{-FEM-SING}}$ / famous $_{\text{-MASC-PLURAL}}$ at the conference.'

Spanish L1 speakers and intermediate but not beginner L2 learners showed longer reading times for ungrammatical sentences in which the adjective mismatched in either gender or number with the preceding noun. Additionally, the size of the grammaticality effect was positively correlated with reading span scores in the intermediate L2 speakers for gender but not number agreement.

Coughlin and Tremblay (2013) investigated shorter and longer agreement dependencies with French clitics, to test the hypothesis that longer dependencies may pose a larger burden on working memory capacity. They tested sentences as in (6). In (6a)/(b), the clitic ('*le*'/'*les*') must agree in number with the sentence subject ('*Ce fruit*'/'*Ces fruits*'). In (6a) the distance between the clitic and subject is short, while in (6b) it is longer, due to an intervening constituent ('*avant l'entretien*').

(6) a. *Ce fruit/Ces fruits Marie le(s) mangera pour sa collation avant l'entretien.*
'The fruit/The fruits Marie it (them) will eat for her snack before the interview.'
b. *Ce fruit/Ces fruits avant l'entretien Marie le(s) mangera pour sa collation.*
'The fruit/The fruits before the interview Marie it (them) will eat for her snack.'

In a self-paced reading experiment, high proficiency but not intermediate proficiency L2 learners exhibited longer reading times following ungrammatical clitics in both short and long conditions. Numerically, the effect was larger in the shorter conditions. There was also a numerical tendency for higher reading span scores to correlate with larger grammaticality effects for plural clitics in the short condition, and for singular clitics in the long condition, though the relevant effects here were not statistically significant (p = .083 and .068 respectively). While this might be suggestive of an effect of reading span on L2 processing, these effects are not in the predicted direction to be consistent with the hypothesis that longer dependencies impose a greater burden on memory capacity.

Foote (2011) investigated subject-verb number agreement and noun-adjective gender agreement in L2 learners of Spanish. The length of the dependency was also manipulated by including conditions with adjacent constituents, and a long condition in which another constituent intervened, as exemplified in (7a) and (7b) respectively. Both Spanish L1 speakers and L2 speakers exhibited longer reading times for sentences with ungrammatical compared to grammatical agreement, and the size of the grammaticality effect was smaller in the long conditions for both types of agreement. Reading span scores were however not found to correlate with grammaticality effects in this study.

(7) a. *Dicen que el libro blanco/blanca está en esa mesa.*
'They say that the book$_{\text{-MASC}}$ white$_{\text{-MASC}}$ / white$_{\text{-FEM}}$ is on that table.'
b. *El pollo del taco está rico/rica pero picante.*
'The chicken$_{\text{-MASC}}$ of the taco is tasty$_{\text{-MASC}}$ / tasty$_{\text{-FEM}}$ but spicy.'

As such, while some studies have reported effects of memory span on the processing of gender agreement, these have not been consistently observed. Other studies have also reported similar effects of distance, however. For example, Keating (2010) examined noun-adjective gender agreement in Spanish, and found that while L1 readers showed grammaticality effects irrespective of distance, L2 learners showed grammaticality effects in short conditions only. Such results might be taken as supporting a capacity-based view of L2 processing, with L2 learners showing less sensitivity to agreement violations as more information needs to be maintained in memory. Here, however, dependency length and interference are confounded. In the studies reported by Coughlin and Tremblay (2013), Foote (2011) and Keating (2010), the intervening constituents in the long conditions could all cause interference in computing the agreement dependency. For example, in (7b) the gender of the distractor constituent ('*taco*') may interfere when it matches the gender of the manipulated adjective ('*rico*'). To fully tease apart the capacity-based view from interference, the morphological properties of the intervening constituent would need to be systematically manipulated (for further discussion, see Cunnings, 2017a: 667–668). Interference in agreement in L1 processing is well attested (e.g., Lago et al., 2015; Wagers, Lau, & Phillips, 2009), and further research is required here in L2 processing (see Tanner, Herschensohn, & Osterhout, 2012, for some preliminary evidence).

In sum, the evidence that individual differences in L2 sentence processing correlate with performance on complex span tasks is mixed. Although some studies have reported correlations between memory span and L2 processing, others have not. Additionally, some of the existing evidence is not able to adjudicate between capacity-based and interference-based accounts of working memory. Further research is needed to not only test the extent to which reading span scores correlate with L2 sentence processing, but to also tease apart effects of L2 memory capacity from interference.

3.2 Memory interference in L2 processing

As discussed above, from the perspective of cue-based parsing and interference-based accounts of working memory, individual differences in sentence processing relate to how information is encoded and retrieved from memory. This perspective, thus, focuses on the *quality* and *content* of representations in memory, rather than the *quantity* per se.

That the quality of memory representations may explain individual variation even in L1 sentence processing is exemplified by Van Dyke, Johns, and Kukona (2014). Van Dyke et al. examined filler-gap dependencies like (8), where '*the boat*' needs to be retrieved from memory and interpreted as the direct object of the main verb ('fixed'/'sailed'). Participants were tested in a 'no memory load' condition in which they simply read the sentences, and a 'memory load' condition in which they also had to remember a list of words ('table', 'sink', 'truck'). The crucial point here being that the list of words in the memory load condition may interfere in resolving the filler-gap dependency in (8b) but not (8a), as in (8b) only the word list contains words that are a possible theme of the main verb ('fixed').

(8) a. It was the boat that the guy who lived by the sea sailed in two sunny days.
b. It was the boat that the guy who lived by the sea fixed in two sunny days.

Participants also completed a battery of individual differences measures, including IQ, reading span and measures of reading/listening comprehension. After partialling out shared variance with IQ, Van Dyke et al. found that reading times at the critical verb in (8b), in the memory load condition, correlated with measures of receptive vocabulary but not reading span. They took this as indicating that the quality of memory representations influences sentence processing, rather than memory capacity. Specifically, they argued that individuals with lower scores of receptive vocabulary may have lower quality memory representations that are more prone to activating irrelevant information. These lower quality representations then lead to higher amounts of interference during memory retrieval during sentence processing.

Although not couched within this framework of memory encoding and retrieval, some existing research in the L2 processing literature is also potentially consistent with the broad idea that the quality of memory representations is important for successful comprehension. Hopp (2018) proposed the Lexical Bottleneck Hypothesis. This predicts that L1/L2 differences are related to lexical knowledge, and that L2 processing can become nativelike if L2 learners have robust lexical entries. Consider Hopp (2014), who conducted an eye-tracking during reading study on relative clause attachment in German learners of L2 English. Hopp tested sentences like (9), where the reflexive disambiguates interpretation of the relative clause to either a *local* or *non-local* noun phrase ('the boy' and 'the mother' respectively).

(9) The doctor examined the mother of the boy who had badly injured herself/himself with the knife.

L1 readers had longer reading times when the reflexive disambiguated towards the non-local noun phrase, replicating the long-standing observation of a local attachment preference in English (e.g., Cuetos & Mitchell, 1988). As a group, the L2 learners did not show any differences between conditions, replicating previous results showing no clear L2 attachment preferences during processing (Felser et al., 2003; but see Witzel, Witzel, & Nicol, 2012). However, a measure of lexical automaticity interacted with L2 processing, such that L2 learners with high levels of lexical automaticity showed a nativelike local attachment preference (see also Hopp, 2013, and Miller, 2014, for evidence of the role of lexical processing during L2 sentence processing).

Although suggestive of the importance of lexical representations in L2 processing, more research is required to test the extent to which L1/L2 differences can be reduced to differences in the quality of lexical representations. A key prediction here would be to test whether L1 processing is influenced by the same lexical factors as L2 processing. Such a finding would provide strong evidence of similarities between the relevant aspects of L1 and L2 processing once differences in lexical processing are accounted for. Direct L1/L2 comparisons in this regard are currently scant, however. Irrespective of this need for further research, the hypothesis that the quality of lexical representations is important for L2 sentence processing may be broadly in-line with models of working memory that focus on the quality, rather than quantity, of information that needs to be encoded and retrieved from memory during comprehension.[1]

---

1. The role of L2 lexical representations could potentially be described from a capacity-based view if greater lexical automaticity frees up processing resources for other aspects of sentence processing. Here, however, we require precise characterisations about how different resources are hypothesised to be allocated during processing (e.g., how are resources allocated to lexical access and syntactic parsing?). Precisely how efficient lexical processing 'frees up' resources for other aspects of sentence processing is currently not well specified, but precise characterisation of this issue is required to adequately test the capacity-based view.

Given the importance of memory retrieval in cue-based parsing, another key predicted source of individual variation from the perspective of interference-based accounts is the set of retrieval cues that are utilised during sentence processing. Individual variation in the relative weightings of retrieval cues could give rise to individual differences in interference (Vasishth et al., 2019). For L2 processing, Cunnings (2017a) proposed that L2 learners may weight retrieval cues differently to L1 speakers, and argued in particular that L2 learners may weight discourse-based cues comparatively more highly than L1 speakers (see also Felser, 2016). It is beyond the scope of this chapter to provide an exhaustive review of how to apply this approach to different phenomena in the L2 sentence processing literature. Instead, I discuss below some recent studies on anaphora resolution to exemplify how this interference-based approach can be utilised to compare L1 and L2 sentence processing.

Felser and Cunnings (2012) investigated sentences like (10) in an eye-tracking during reading experiment with L1 English speakers and German learners of L2 English. They manipulated the extent to which a reflexive matched or mismatched in gender with an 'accessible' antecedent ('the soldier') and an 'inaccessible' antecedent ('James/Helen'). According to binding theory (Chomsky, 1981), only the accessible antecedent can be coindexed with the reflexive. From the perspective of cue-based parsing, anaphora resolution involves retrieving an antecedent from memory when the reflexive is encountered. Retrieval is likely to be guided by structural cues, that favour a syntactically local antecedent (e.g., [+LOCAL]), along with features of the reflexive (e.g., [+SINGULAR], [+MASCULINE] etc.), and potentially discourse-related factors (e.g., [+TOPIC]).

(10) a. James/Helen has worked at the army hospital for years. She/He noticed that the soldier had wounded himself while on duty in the Far East.
b. James/Helen has worked at the army hospital for years. She/He noticed that the soldier had wounded herself while on duty in the Far East.

Similar to some L1 studies (Dillon et al., 2013; Sturt, 2003), Felser and Cunnings (2012) found that L1 speaker's reading times were longer in (10b) than (10a), when the accessible antecedent mismatched in gender with the reflexive, suggesting rapid application of binding constraints. During first-pass processing of the reflexive, the L2 learners showed a different pattern and had longer reading times when the inaccessible antecedent mismatched the gender of the reflexive. L2 learners did show effects of the gender of the accessible antecedent in later measures of processing, such as rereading time, and evidenced knowledge of binding constraints in an offline task. The results however suggest that during initial processing, L2 learners weighted discourse-based cues over the syntactic requirement for a local antecedent. This finding might be in particular unexpected from a capacity-based view of L2 processing. If L2 processing was capacity limited, we might expect an even greater focus on the local antecedent in L2 processing than L1 processing, but this was not found (see Felser, 2019, for further discussion).

It should be noted that the extent to which anaphora resolution is susceptible to interference in L1 processing is widely debated. While some studies suggest L1 readers weight structural cues more highly than non-structural cues (Dillon et al., 2013), it is too strong to claim that L1 anaphora resolution is impervious to interference (Jäger et al., 2020). Further L1/L2 comparisons are required here to tease apart the relative weightings of different retrieval cues by L1 and L2 readers.

The claim that L2 learners weight discourse-based cues more heavily than L1 speakers can also potentially provide an alternative account of the long-debated controversy in SLA research on null and overt pronouns. In null subject languages, null pronouns typically refer to the current discourse topic, while overt pronouns index a topic-shift to a less prominent antecedent. Studies have shown that L2 speakers interpret null pronouns in a similar way to L1 speakers, but additionally allow overt pronouns to refer to topic antecedents, rather than indexing a topic shift (e.g., Belletti, Bennati, & Sorace, 2007; Sorace & Filiaci, 2006). These results motivated the Interface Hypothesis (Sorace, 2011; Sorace & Filiaci, 2006), which states that L2 learners have difficulty integrating information between syntax and other cognitive domains. However, these results are also consistent with the claim that L2 speakers weight discourse-based cues that favour prominent antecedents more heavily than L1 speakers.

Although overt pronouns may cause difficulty for L2 learners in a null subject L2, they do not always appear to cause difficulty for learners of a non-null subject L2, where overt pronouns refer to the current discourse topic. Cunnings, Fotiadou, and Tsimpli (2017) investigated the processing of overt pronouns as in (11) using the visual world paradigm. They found that upon hearing the pronoun, both L1 and L2 listeners' gaze across a visual display quickly focused on the discourse prominent subject antecedent ('Mr Smith') rather than the object ('Peter'). This was despite the fact that the L2 learners' L1 was a null subject language in which overt pronouns index a topic shift. This finding would be unexpected under a capacity-based view, which may favour the linearly closer object antecedent ('Peter'). However, it is compatible with the hypothesis that L2 learners strongly weight discourse-based cues to anaphora resolution.

(11) After Mr Smith spoke to Peter by the till in the shop, he paid for the expensive ice cream that looked tasty.

Although Cunnings (2017a) emphasised the relative weighting of discourse-based cues in L2 processing, further research is required to test the extent to which this claim generalises to other linguistic phenomena. A focus on discourse-based cues may be pertinent for anaphora, where discourse prominence is a key factor in determining coreference, but whether this is the case for other aspects of sentence processing is a matter for future research. For example, it might be that discourse prominence plays less of a role in the processing of morphosyntactic agreement, but L1/L2 differences may still arise if L1 speakers and L2 learners weight morphological and syntactic cues differently to guide retrieval. What is required here is careful consideration and hypotheses about what retrieval cues a particular group of L2 learners may utilise. The hypothesis that L2 learners may weight discourse-based cues more highly than L1 speakers is one starting point to examine this issue, but further research is required to elucidate how L1 speakers and L2 learners weight different retrieval cues during processing.

3.3 Retrieval cues in processing and acquisition

This account of working memory, that focuses on the quality of representations in memory and the cues utilised to guide memory retrieval, would predict that instead of describing individual variation in L2 processing in terms of memory capacity, the primary predictor of individual differences in both L1 and L2 processing should be related to linguistic experience. For example, robust lexical representations require adequate linguistic experience of the relevant vocabulary. Additionally, the relative weightings of different retrieval cues by L1 speakers and L2 learners would result from the differing degrees of linguistic exposure to the relevant cues. From this perspective, L2 acquisition can in part be considered a process of learning to weight the relevant retrieval cues in an appropriate way to guide successful memory retrieval during sentence processing.

This focus on a learners' linguistic experience echoes long-debated questions in the L2 acquisition literature on the role of proficiency and L1 transfer. For example, one might hypothesise that for reflexives in sentences like (10), L2 learners whose L1 allows long-distance reflexives (e.g., Japanese or Chinese) might weight retrieval cues differently to L2 learners whose first language requires local binding as in English. Note that the first language of the L2 leaners tested by Felser and Cunnings (2012), German, has binding constraints similar to English. However, these learners still seemed to weight retrieval cues to initially favour the discourse prominent non-local antecedent. Felser, Sato, and Bertenshaw (2009) reported that Japanese L2 learners also temporarily violate binding constraints on reflexives during processing, though they tested slightly different stimuli to Felser and Cunnings. Direct comparisons between L2 learners whose L1 allows local binding compared to those that allow non-local binding are required here.

Transfer might also play a role in resolving pronouns as in (11). Although the results of Cunnings et al. (2017) are not suggestive of transfer, which would have favoured less prominent 'Peter' as the antecedent of the pronoun over 'Mr Smith', other studies have suggested a role of L1 transfer in processing and interpreting overt pronouns in a non-null subject L2 (Contemori, Asiri, & Perea Irigoyen, 2019; Roberts, Gullberg, & Indefrey, 2008). Further research is required to examine the extent to which an L2 learners' first language may influence how they weight different cues to memory retrieval during L2 processing.

3.4 Task performance and characterising individual differences

Although measures such as reading span are often assumed to be measures of working memory capacity, the precise characterisation of what such tasks measure is debated. For example, it has long been questioned whether reading span is a measure of memory capacity or linguistic experience (e.g., Kidd, Donnelly, & Christiansen, 2018; MacDonald & Christiansen, 2002). Indeed, Farmer et al. (2017) recently demonstrated that linguistic experience influences reading span performance in L1 speakers. The extent to which L2 performance on the reading span test is related to memory capacity or L2 experience has also been debated (e.g., Juffs & Harrington, 2011).

One way to potentially overcome this confound with L2 proficiency is to test L2 learners on the reading span test in both their L1 and L2 and calculate a composite reading span score from both tests (Juffs & Harrington, 2011). Another possibility is to use a complex span task, such as operation span (Turner & Engle, 1989), which is less dependent on language proficiency. More broadly, a number of researchers have advocated computing composite scores across several complex span tasks rather than testing a single measure (e.g., Waters & Caplan, 2003). Additionally, the way that span tasks are administered and scored can also influence the extent to which they correlate with other tasks (Leeser & Sunderman, 2016). To date, these issues have not been systematically examined in L2 sentence processing. Further research is required here to more stringently test the role that span tasks and memory capacity may play in L2 processing.

It is unlikely that a single construct related to working memory and/or language experience can explain all L1/L2 differences in sentence processing, and different aspects of language comprehension may correlate with different individual differences. For example, although individual differences in memory encoding and retrieval may be relevant when examining linguistic dependencies, a different set of memory operations are likely required for other aspects of comprehension, such as the reanalysis processes involved in resolving temporary syntactic ambiguity.

Consider (12), where the reader may initially interpret 'the baby' as the direct object of the subordinate clause verb ('dressed') when it is in fact the subject of the main verb. Even L1 readers are known to have difficulty with such sentences, sometimes maintaining the initially assigned interpretation ('Anna dressed the baby') after reanalysis (Christianson et al., 2001). Recent studies have indicated that reanalysis causes particular difficulty for L2 learners (Jacob & Felser, 2016; Pozzan & Trueswell, 2016; for in-depth discussion, see Fujita, 2019).

(12) After Anna dressed the baby played in the cot.

Recent research in L1 processing has eschewed memory capacity as a predictor of reanalysis success in sentences like (12) in favour of cognitive control, the ability to switch attention during cases of information conflict (Kan et al., 2013; Novick et al., 2014). Woodard, Pozzan, and Trueswell (2016) recently provided evidence that individual differences on cognitive control tasks correlate with reanalysis ability during L1 language development, but how cognitive control may influence reanalysis in L2 processing has not yet been systematically examined.

In sum, further research is required to tease apart the capacity-based and interference-based accounts of L2 processing. In testing purported measures of memory capacity, it is important to tease apart capacity from language experience, and to consider how different aspects of sentence processing rely on different memory operations.

4. Conclusion

While the interaction between sentence processing and working memory has long been debated, recent approaches to working memory in language comprehension that focus on memory encoding and retrieval allow for greater integration between work in second language acquisition and non-native sentence processing. While capacity-based views of working memory may have been dominant in research in L2 processing, I argue that a clearer focus on the memory encoding and retrieval operations required during real-time sentence processing would allow for new insight into key debate in L2 acquisition and processing, and a better understanding of the similarities and differences between L1 and L2 processing.

References

Baddeley, A. (2007). *Working memory, thought, and action*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198528012.001.0001

Baddeley, A. (2015). Working memory in second language learning. In W. Zhisheng, M. Morta, & A. McNeil (Eds). *Working memory and second language acquisition and processing: Theories, research and commentaries* (pp. 17–28). Multilingual Matters. https://doi.org/10.21832/9781783093595-005

Belletti, A., Bennati, E., & Sorace, A. (2007). Theoretical and developmental issues in the syntax of subjects: Evidence from near-native Italian. *Natural Language & Linguistic Theory*, 25(4), 657–689. https://doi.org/10.1007/s11049-007-9026-9

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, 22(1), 77–94. https://doi.org/10.1017/S0140525X99001788

Chomsky, N. (1981). *Lectures on government and binding: The Pisa lectures* (7th ed). Mouton de Gruyter.

Chomsky, N., & Miller, G. A. (1963). Introduction to the formal analysis of natural languages. In R. Luce, R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 2, pp. 269–323). Wiley.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, 42(4), 368–407. https://doi.org/10.1006/cogp.2001.0752

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, 27, 3–42. https://doi.org/10.1017/S0142716406060024

Clahsen, H., & Felser, C. (2018). Some notes on the Shallow Structure Hypothesis. *Studies in Second Language Acquisition*, 40, 693–706. https://doi.org/10.1017/S0272263117000250

Contemori, C., Asiri, O., & Perea Irigoyen, E. D. (2019). Anaphora resolution in L2 English: An analysis of discourse complexity and cross-linguistic interference. *Studies in Second Language Acquisition*, 41(5), 971–998. https://doi.org/10.1017/S0272263119000111

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12(5), 769–786. https://doi.org/10.3758/BF03196772

Coughlin, C. E., & Tremblay, A. (2013). Proficiency and working memory based explanations for nonnative speakers' sensitivity to agreement in sentence processing. *Applied Psycholinguistics*, 34(3), 615–646. https://doi.org/10.1017/S0142716411000890

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin*, 104, 163–191. https://doi.org/10.1037/0033-2909.104.2.163

Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review*, 24(4), 1158–1170. https://doi.org/10.3758/s13423-016-1191-6

Cuetos, F., & Mitchell, D. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the late closure strategy in Spanish. *Cognition*, 30(1), 73–105. https://doi.org/10.1016/0010-0277(88)90004-2

Cunnings, I. (2017a). Parsing and working memory in bilingual sentence processing. *Bilingualism: Language and Cognition*, 20(4), 659–678. https://doi.org/10.1017/S1366728916000675

Cunnings, I. (2017b). Interference in native and non-native sentence processing. *Bilingualism: Language and Cognition*, 20(4), 712–721. https://doi.org/10.1017/S1366728916001243

Cunnings, I., Fotiadou, G., & Tsimpli, I. (2017). Anaphora resolution and reanalysis during L2 sentence processing: Evidence from the visual world paradigm. *Studies in Second Language Acquisition*, 39(4), 621–652. https://doi.org/10.1017/S0272263116000292

Cunnings, I., & Sturt, P. (2018). Retrieval interference and semantic interpretation. *Journal of Memory and Language*, 102, 16–27. https://doi.org/10.1016/j.jml.2018.05.001

Dallas, A., DeDe, G., & Nicol, J. (2013). An event-related potential (ERP) investigation of filler-gap processing in native and second language speakers. *Language Learning*, 63(4), 766–799. https://doi.org/10.1111/lang.12026

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6

Dekydtspotter, L., & Renaud, C. (2014). On second language processing and grammatical development: The parser in second language acquisition. *Linguistic Approaches to Bilingualism*, 4, 131–165. https://doi.org/10.1075/lab.4.2.01dek

Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language*, 69(2), 85–103. https://doi.org/10.1016/j.jml.2013.04.003

Dussias, P. E., & Piñar, P. (2010). Effects of reading span and plausibility in the reanalysis of wh-gaps by Chinese-English second language speakers. *Second Language Research*, 26(4), 443–472. https://doi.org/10.1177/0267658310373326

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, 11(1), 19–23. https://doi.org/10.1111/1467-8721.00160

Engle, R. W. (2018). Working memory as executive attention: A revisit. *Perspectives on Psychological Science*, 12, 190–193. https://doi.org/10.1177/1745691617720478

Farmer, T. A., Fine, A. B., Misyak, J. B., & Christiansen, M. H. (2017). Reading span task performance, linguistic experience, and the processing of unexpected syntactic events. *Quarterly Journal of Experimental Psychology*, 70(3), 413–433. https://doi.org/10.1080/17470218.2015.1131310

Felser, C. (2016). Binding and coreference in non-native language processing. In A. Holler & K. Suckow (Eds.), *Empirical perspectives on anaphora resolution*. (pp. 229–253). De Gruyter. https://doi.org/10.1515/9783110464108-012

Felser, C. (2019). Structure-sensitive constraints in non-native sentence processing. *Journal of the European Second Language Association*, 3(1), 12–22. https://doi.org/10.22599/jesla.52

Felser, C., & Cunnings, I. (2012). Processing reflexives in a second language: The timing of structural and discourse-level constraints. *Applied Psycholinguistics*, 33(3), 571–603. https://doi.org/10.1017/S0142716411000488

Felser, C., & Roberts, L. (2007). Processing wh-dependencies in a second language: A cross-modal priming study. *Second Language Research*, 23(1), 9–36. https://doi.org/10.1177/0267658307071600

Felser, C., Roberts, L., Marinis, T., & Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics*, 24(3), 453–489. https://doi.org/10.1017/S0142716403000237

Felser, C., Sato, M., & Bertenshaw, N. (2009). The on-line application of binding Principle A in English as a second language. *Bilingualism: Language and Cognition*, 12(4), 485–502. https://doi.org/10.1017/S1366728909990228

Foote, R. (2011). Integrated knowledge of agreement in early and late English – Spanish bilinguals. *Applied Psycholinguistics*, 32(1), 187–220. https://doi.org/10.1017/S0142716410000342

Fujita, H. (2019). Reanalysis processes in native and non-native language comprehension (Unpublished doctoral dissertation). University of Reading.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition*, 14(1), 25–38. https://doi.org/10.1017/S0272263100010457

Hofmeister, P., Jaeger, T. F., Arnon, I., Sag, I. A., & Snider, N. (2013). The source ambiguity problem: Distinguishing the effects of grammar and processing on acceptability judgments. *Language and Cognitive Processes*, 28(1–2), 48–87. https://doi.org/10.1080/01690965.2011.572401

Hofmeister, P., & Sag, I. A. (2010). Cognitive constraints and island effects. *Language*, 86(2), 366–415. https://doi.org/10.1353/lan.0.0223

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, 29(1), 33–56. https://doi.org/10.1177/0267658312461803

Hopp, H. (2014). Working memory effects in the L2 processing of ambiguous relative clauses. *Language Acquisition*, 21(3), 250–278. https://doi.org/10.1080/10489223.2014.892943

Hopp, H. (2018). The bilingual mental lexicon in L2 sentence processing. *Second Language*, 17, 5–27.

Jacob, G., & Felser, C. (2016). Reanalysis and semantic persistence in native and non-native garden-path recovery. *Quarterly Journal of Experimental Psychology*, 69(5), 907–925. https://doi.org/10.1080/17470218.2014.984231

Jäger, L. A., Mertzen, D., Van Dyke, J. A., & Vasishth, S. (2020). Interference patterns in subject-verb agreement and reflexives revisited: A large-sample study. *Journal of Memory and Language*, 111, 104063. https://doi.org/10.1016/j.jml.2019.104063

Juffs, A. (2004). Representation, processing and working memory in a second language. *Transactions of the Philological Society*, 102(2), 199–225. https://doi.org/10.1111/j.0079-1636.2004.00135.x

Juffs, A. (2005). The influence of first language on the processing of wh-movement in English as a second language. *Second Language Research*, 21(2), 121–151. https://doi.org/10.1191/0267658305sr255oa

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching*, 44(2), 137–166. https://doi.org/10.1017/S0261444810000509

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99(1), 122–149. https://doi.org/10.1037/0033-295X.99.1.122

Just, M. A., Carpenter, P. A., & Keller, T. A. (1996). The capacity theory of comprehension: New frontiers of evidence and arguments. *Psychological Review*, 103(4), 773–780. https://doi.org/10.1037/0033-295X.103.4.773

Kan, I. P., Teubner-Rhodes, S., Drummey, A. B., Nutile, L., Krupa, L., & Novick, J. M. (2013). To adapt or not to adapt: The question of domain-general cognitive control. *Cognition*, 129(3), 637–651. https://doi.org/10.1016/j.cognition.2013.09.001

Keating, G. D. (2010). The effects of linear distance and working memory on the processing of gender agreement in Spanish. In B. VanPatten & J. Jegerski (Eds.), *Language Acquisition and Language Disorders*, 53, 113–134. https://doi.org/10.1075/lald.53.05kea

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends in Cognitive Sciences*, 22(2), 154–169. https://doi.org/10.1016/j.tics.2017.11.006

Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207(4427), 203–205. https://doi.org/10.1126/science.7350657

Lago, S., Shalom, D. E., Sigman, M., Lau, E. F., & Phillips, C. (2015). Agreement attraction in Spanish comprehension. *Journal of Memory and Language*, 82, 133–149. https://doi.org/10.1016/j.jml.2015.02.002

Leeser, M., & Sunderman, G. (2016). Methodological implications of working memory tasks for L2 processing research. In G. Granena, D. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in second language processing and acquisition* (pp. 89–104). John Benjamins. https://doi.org/10.1075/bpa.3.05lee

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29(3), 375–419. https://doi.org/10.1207/s15516709cog0000_25

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, 10(10), 447–454. https://doi.org/10.1016/j.tics.2006.08.007

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, 21(4), 861–883. https://doi.org/10.3758/s13423-013-0565-2

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, 109(1), 35–54. https://doi.org/10.1037/0033-295X.109.1.35

Mathy, F., Chekaf, M., & Cowan, N. (2018). Simple and complex working memory tasks allow similar benefits of information compression. *Journal of Cognition*, 1(1), 31. https://doi.org/10.5334/joc.31

McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, 55(3), 381–401. https://doi.org/10.1016/j.jml.2006.06.006

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research*, 29(2), 111–123. https://doi.org/10.1023/A:1005184709695

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, 48(1), 67–91. https://doi.org/10.1016/S0749-596X(02)00515-6

Miller, K. (2014). Accessing and maintaining referents in L2 processing of wh-dependencies. *Linguistic Approaches to Bilingualism*, 4, 167–191. https://doi.org/10.1075/lab.4.2.02mil

Navon, D. (1984). Resources – A theoretical soup stone? *Psychological Review*, 91, 216–234. https://doi.org/10.1037/0033-295X.91.2.216

Novick, J. M., Hussey, E., Teubner-Rhodes, S., Harbison, J. I., & Bunting, M. F. (2014). Clearing the garden-path: Improving sentence processing through cognitive control training. *Language, Cognition and Neuroscience*, 29(2), 186–217. https://doi.org/10.1080/01690965.2012.758297

Pan, H.-Y., & Felser, C. (2011). Referential context effects in L2 ambiguity resolution: Evidence from self-paced reading. *Lingua*, 121(2), 221–236. https://doi.org/10.1016/j.lingua.2010.08.003

Pan, H.-Y., Schimke, S., & Felser, C. (2015). Referential context effects in non-native relative clause ambiguity resolution. *International Journal of Bilingualism*, 19(3), 298–313. https://doi.org/10.1177/1367006913515769

Parker, D., Shvartsman, M., & Van Dyke, J. A. (2017). The cue-based based retrieval theory of sentence comprehension: New findings and new challenges. In L. Escobar, V. Torrens, & T. Parodi (Eds.), *Language processing and disorders* (pp. 121–144). Cambridge Scholars.

Pozzan, L., & Trueswell, J. C. (2016). Second language processing and revision of garden-path sentences: A visual word study. *Bilingualism: Language and Cognition*, 19(3), 636–643. https://doi.org/10.1017/S1366728915000838

Roberts, L., Gullberg, M., & Indefrey, P. (2008). Online pronoun resolution in L2 discourse: L1 influence and general learner effects. *Studies in Second Language Acquisition*, 30(3), 333–357. https://doi.org/10.1017/S0272263108080480

Sagarra, N., & Herschensohn, J. (2010). The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua*, 120(8), 2022–2039. https://doi.org/10.1016/j.lingua.2010.02.004

Sorace, A. (2011). Pinning down the concept of "interface" in bilingualism. *Linguistic Approaches to Bilingualism*, 1(1), 1–33. https://doi.org/10.1075/lab.1.1.01sor

Sorace, A., & Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Language Research*, 22(3), 339–368. https://doi.org/10.1191/0267658306sr271oa

Sprouse, J., Wagers, M., & Phillips, C. (2012). A test of the relation between working-memory capacity and syntactic island effects. *Language*, 88(1), 82–123. https://doi.org/10.1353/lan.2012.0004

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, 48(3), 542–562. https://doi.org/10.1016/S0749-596X(02)00536-3

Tanner, D., Herschensohn, J., & Osterhout, L. (2012). Electrophysiological markers of interference and structural facilitation in native and nonnative agreement processing. In A. Biller, A. Chung, & A. Kimball (Eds.), *Proceedings of the 36th annual Boston University Conference on Language Development* (pp. 294–606). Cascadilla.

Turner, M., & Engle, R. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28, 127–154. https://doi.org/10.1016/0749-596X(89)90040-5

Van Dyke, J. A., & Johns, C. L. (2012). Memory interference as a determinant of language comprehension: Interference in comprehension. *Language and Linguistics Compass*, 6(4), 193–211. https://doi.org/10.1002/lnc3.330

Van Dyke, J. A., Johns, C. L., & Kukona, A. (2014). Low working memory capacity is only spuriously related to poor reading comprehension. *Cognition*, 131(3), 373–403. https://doi.org/10.1016/j.cognition.2014.01.007

Vasishth, S., Nicenboim, B., Engelmann, F., & Burchert, F. (2019). Computational models of retrieval processes in sentence processing. *Trends in Cognitive Sciences*, 23(11), 968–982. https://doi.org/10.1016/j.tics.2019.09.003

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, 61(2), 206–237. https://doi.org/10.1016/j.jml.2009.04.002

Waters, G., & Caplan, D. (2003). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments, & Computers*, 35, 550–564. https://doi.org/10.3758/BF03195534

Wen, Z., Mota, M. B., & McNeill, A. (Eds.). (2015). *Working memory in second language acquisition and processing*. Multilingual Matters. https://doi.org/10.21832/9781783093595

Witzel, J., Witzel, N., & Nicol, J. (2012). Deeper than shallow: Evidence for structure-based parsing biases in L2 sentence processing. *Applied Psycholinguistics*, 33, 419–456. https://doi.org/10.1017/S0142716411000427

Woodard, K., Pozzan, L., & Trueswell, J. C. (2016). Taking your own path: Individual differences in executive function and language processing skills in child learners. *Journal of Experimental Child Psychology*, 141, 187–209. https://doi.org/10.1016/j.jecp.2015.08.005

Heritage language speakers inform the critical period hypothesis for first and second language acquisition

Silvina Montrul
University of Illinois at Urbana-Champaign

This chapter considers how heritage language speakers, bilinguals exposed to their L1 since birth either exclusively or simultaneously with the majority language, shed new light on the role of age in language acquisition. Investigating the ultimate attainment of bilinguals with early and late exposure to their L1 and L2 helps disentangle how the roles of previous linguistic knowledge and of quantity and quality of input interact with age of onset of bilingualism. Comparing the linguistic and processing abilities of heritage speakers and L2 learners allows us to understand how the timing of input interacts with situational factors and elucidates the developmental schedule of different modules of the grammar. Early exposure is critical for aspects of phonology, syntax, and morphology.

Keywords: heritage language acquisition, simultaneous bilingualism, age effects, second language acquisition, morphology, syntax, phonology, processing

1. Introduction

It is widely acknowledged that adults and children learn languages differently, and that the timing of language exposure and learning has a deterministic impact on the ultimate ability achieved in different grammatical domains. Although there is compelling evidence that there is a critical period for the acquisition of a first language (L1) (Mayberry & Kluender, 2018), the existence of a critical period for adult second language (L2) acquisition continues to be disputed: Age effects in L2 acquisition are compatible with both maturational explanations that prioritize the role of the genetic endowment (Newport, 2018; DeKeyser, 2018), and with proposals that place the burden of proof on environmental factors and experience (Flege, 2018; White, 2018; among many others).

https://doi.org/10.1075/lald.67.11mon

In this chapter, I consider how heritage language speakers, bilinguals exposed to their L1 since birth either exclusively or simultaneously with the majority language spoken in their sociolinguistic environment, shed new light on the role of age in language acquisition. In Montrul (2008) I claimed that heritage speakers provide a different but critical angle on age effects in bilingual acquisition, and I continue to maintain so. Like monolingual children, heritage speakers were exposed to their heritage language since birth through their primary caregivers, either in a monolingual or a bilingual context. Yet, unlike monolingual children, the majority of young adult heritage speakers studied to date show many of the same patterns of non-native attainment in different aspects of their heritage grammars as adult L2 learners (Ionin & Montrul, 2012; Montrul, 2008). A critical question originally posed by Au et al. (2002) is whether early exposure to a language brings advantages in several aspects of the grammar compared to later language experience in bilinguals. Almost two decades of research comparing heritage speakers and L2 learners of different proficiency levels and in different linguistic domains have directly or indirectly addressed this question and provided suggestive answers.

In a review article on age effects in language acquisition, Mayberry and Kluender (2018) argue that age-onset effects for L1 and L2 language acquisition are different. L1 acquisition arises from post-natal brain development interacting with environmental linguistic experience. By contrast, L2 learning after childhood is built on prior childhood L1 acquisition, both linguistically and neurally, making it a less clear test of the critical period for language. The empirical support for these claims comes from case studies and cross-sectional studies of the L1 and L2 acquisition of American Sign Language, some of which collected behavioral and neurolinguistic data. For example, Mayberry (1993) tested knowledge of grammatical and semantic relations in native speaker signers (exposed to ASL ages birth to 3), childhood signers (exposure between ages 5–8), late L1 signers (exposure to ASL ages 9–13) and English-speaking L2 learners of ASL (exposure to ASL ages 9–15). The native signers performed at ceiling, while the other groups were very inaccurate. However, among the L1 groups, the late L1 signers were significantly more inaccurate than the childhood signers. The L2 learners were no different from the childhood signers on grammatical properties but were more accurate than the childhood signers on semantics. Mayberry (1993) showed clear age effects for L1 acquisition, and while the L2 learners did not reach ceiling performance like the native signers, they were close to the child signers, suggesting that having a native language both helps and hampers L2 acquisition. In their review of the latest research on this issue, Mayberry and Kluender (2018) conclude that language acquisition and development of the brain language system appear to reciprocally affect one another, but only when the onset of language experience is synchronous with the onset of post-natal brain development. Therefore, L1 acquisition and development of the brain language system can be considered an example of critical period learning. Because in L2 acquisition brain development and L1 acquisition are established when the acquisition of another language starts, the age effects noted are different, and not as critical as in L1 acquisition. There are also age effects, but not a critical period for heritage language acquisition, I argue, because the heritage language was acquired during brain development, and heritage speakers were exposed to language during the critical period for learning.

Even though there are age effects, but not a critical period in bilingualism, I argue that investigating the ultimate attainment of bilinguals with early and late exposure to their L1 and L2 helps disentangle, in some linguistic areas, how the roles of previous linguistic knowledge and of quantity and quality of input interact with age of acquisition (Montrul, 2008, 2016). Comparing the linguistic and processing abilities of heritage speakers and L2 learners allows us to understand how the timing of input (age) interacts with external factors (the type of input) and elucidates the developmental schedule of different modules of the grammar (phonology, morphology, syntax). We will see that early exposure is critical for aspects of phonology, syntax and morphology.

2. Age effects in child bilingual acquisition: Maturation or language exposure?

A main goal of generative linguistics is to characterize the linguistic competence of mature native speakers of a language and to explain how very young children acquire the properties of their native language (L1). A related goal has been to extend the same approach to investigate the nature of L2 acquisition and 'interlanguage' (Selinker, 1972) competence, from initial state to ultimate attainment (Gregg, 1996; White, 1989, 2003). Understanding the sharp difference in the linguistic outcome of L1 acquisition by children and L2 acquisition by adults has guided a great deal of the research agenda in generative approaches to L2 acquisition and has justified the experimental comparison between native and non-native speakers. At issue has been whether interlanguages are natural language systems, constrained by the same innate linguistic principles and language specific parameters as a native language. Assuming Universal Grammar guides native language acquisition, do L2 learners use the same domain-specific linguistic mechanisms as L1 learners? Answers to this question fall into two broad and opposing theoretical positions: Yes, L2 acquisition is constrained by the same linguistic principles as L1 acquisition (Epstein, Flynn, & Martohardjono, 1996; Schwartz & Sprouse, 1996; White, 2003); and No, L2 acquisition is fundamentally different from L1 acquisition and does not rely on linguistic-specific mechanisms (Bley-Vroman, 1990; Clahsen & Felser, 2006;

Hawkins & Chan, 1997; Meisel, 2011; Tsimpli, 2013). Although both views acknowledge that variability and fossilization are possible outcomes of L2 acquisition, despite many years of exposure and use of the target language (Schachter, 1990; Sorace, 1993), they explain them in different ways. The former position attributes variability, incomplete acquisition and fossilization to processing or other extra-grammatical factors, while the latter, known as the deficit view, to maturational effects. Debates on which of these two positions is most compatible with the available empirical evidence has been going on for over three decades, and a summary of the arguments and evidence is beyond the scope of this chapter. I focus instead on the issue of age of onset of bilingual acquisition, age of intensive exposure to majority language input, and interruption to input in childhood, which are critical to understand heritage language acquisition and how it contributes to debates on the critical period for language.

A vast body of evidence shows that in an immigration setting the earlier the age of arrival and onset of L2 acquisition, the more native-like the proficiency eventually achieved in adulthood, and these findings apply for phonology (Yeni-Komshian, Flege, & Liu, 2000) and morphosyntax (Abrahamsson, 2012; Abrahamsson & Hyltenstam, 2009; Hartshorne, Tenenbaum, & Pinker, 2018; Johnson & Newport, 1989, 1991). Other cognitive and neurocognitive approaches to L2 acquisition, which do not necessarily posit an innate linguistic module like Universal Grammar (DeKeyser, 2003; Paradis, 2004; Ullman, 2001) account for differences between L1 and L2 acquisition as arising from different memory systems and types of learning: L1 acquisition relies on implicit learning and knowledge stored in procedural memory, while L2 acquisition is primarily accomplished via explicit learning and reliance on declarative knowledge, with minimal implication of implicit learning . The critical period hypothesis seems to apply to the ability to learn language implicitly (DeKeyser, 2003; Paradis, 2004; Ullman, 2001). The fact that L2 learners use explicit, domain-learning mechanisms explains why they can sometimes emulate the successful outcomes of L1 acquisition, especially those who score high on language aptitude (Abrahamsson & Hyltenstam, 2009; DeKeyser, 2003; Granena & Long, 2013).

Early studies of age effects in L2 acquisition have generally posited that puberty (11–13 years old) is the critical age when the ability to learn another language declines, and this suggests that early bilinguals (with onset of bilingualism before puberty) are more likely to reach native-like grammatical competence than late bilinguals (with onset of bilingualism after puberty). Within early bilingualism, simultaneous bilinguals are exposed to two L1s since birth or within the first three years of life, and sequential bilinguals are monolingual during the first three years of life and begin exposure to a L2 after approximately age 3;00. According to Meisel (2010, 2011, 2013), simultaneous bilingual children are like monolingual children.

The acquisition of their two languages is guided by Universal Grammar and follows the same developmental patterns of the languages acquired by monolingual children. However, age effects for the acquisition of morphosyntax start as early as ages 4–6 in bilingual children, and Meisel (2013) posits multiple sensitive periods between the ages of 4 and 15 years for different aspects of morphosyntax in sequential or successive bilingualism. In support of his claim, Meisel (2018) reports on a study of gender in French by German-speaking children and shows that the children starting acquisition of French at age 3;00 already show different accuracy rates and patterns from monolingual French-speaking children and simultaneous French-German speaking children and resemble adult L2 learners. Meisel also claims that different syntactic properties (gender, root infinitives, finiteness, etc.) show different sensitive periods, even in the same children. Although Meisel has studied the very early stages of bilingualism during the pre-school period, the implication of his findings is that simultaneous bilinguals can attain native ability in their two languages because they were exposed to the two languages early in life and both languages are learned through Universal Grammar. In contrast, child sequential bilinguals and adult L2 learners have very different grammars from L1 and simultaneous bilingual learners because the timing of L2 acquisition is after age 3.

In light of Meisel's claims, research on heritage language acquisition, which includes simultaneous bilingual children whose L1 is a minority language acquired before age 3;00 provides convincing evidence that age of onset of acquisition is not a sufficient condition for monolingual-like development of the two languages of simultaneous bilinguals in adulthood and even earlier. Even bilingual children younger than 3;00 do not always develop equal, balanced, monolingual-like age-appropriate proficiency in the two languages at that young age because significant exposure to the two languages, other than just age, is as important as biology for successful language development. Flores (2010), Montrul (2016), and Polinsky (2018) show that the language development and outcome of simultaneous and sequential bilingual children depend on the sociopolitical context of acquisition. By sociopolitical context, I mean the ideologies of the territories where the minority language is spoken, which determine whether there is education and other public presence of the minority language in the society, and the attitudes of the minority and majority language speakers toward the minority language. For example, the United States has overt and covert English-only ideologies, attitudes, and politics, which affect the prestige and the availability of heritage languages. And while Meisel (2013) has delineated the boundary between simultaneous and sequential bilinguals at about age 3, 5–6 years old is the most critical age in heritage language acquisition because it marks school entry in most parts of the world. Many heritage speakers are monolingual or dominant in the heritage language before age 5, but language dominance shifts dramatically after that age, especially in the United

States (Carreira & Kagan, 2011), where the majority of the heritage speakers are schooled exclusively in English. In some other cases, input in the heritage language can be interrupted for several years in childhood, as in cases of heritage language reversal (Flores, 2015).

The typical outcome of heritage language acquisition, just like the outcome of adult L2 acquisition, is variable and most often unevenly non-native in several aspects of the heritage language. Montrul (2018) showed that, as in adult L2 acquisition, many aspects of syntax and semantics are not problematic for heritage language learners, but morphology seems to be most prone to fossilization and the acquisition *bottleneck* (Slabakova, 2008; Montrul, 2018). For example, gender agreement in nouns, aspectual distinctions, and prepositions are highly problematic in both populations (Alarcón, 2011; Laleko & Polinsky, 2016; Montrul, 2009). Based on these findings, in Montrul (2008), I re-examined the claim that age effects apply mostly to the adult L2 acquisition situation. I argued and demonstrated that age effects play a decisive role in cases of L1 loss in early bilingualism: The younger the bilingual child when input in the heritage language is reduced or interrupted altogether (as in internationally adopted children), the faster and more pervasive the extent of language loss. By contrast, late bilinguals, or native speakers with age of onset of bilingualism in adulthood in immigration contexts are unlikely to lose aspects of their L1 to the same extent as has been found in childhood attrition, regardless of amount of language use (Schmid, 2018). For example, Ahn et al. (2017) investigated age effects in the speech perception of Korean stops in heritage speakers with ages of reduced contact to Korean ranging from 3 to 15 years of age, and found that compared to control participants, the Korean heritage speakers were less accurate on L1-specific contrasts, and their accuracy was significantly correlated with age of reduced contact with Korean as L1. These findings suggest that the earlier bilinguals are extensively exposed to the L2, the less likely they are to perceive L1 sounds natively.

Although an early start is important, heritage language speakers challenge the claim that an early start is all that is needed for native-like ultimate attainment on at least two counts. First, many heritage speakers do not show native or age-appropriate ability in all aspects and domains of their heritage language in early and later childhood and beyond. Second, many heritage speakers do eventually reach native ability in most aspects and domains of their L2, even if acquired after age 4 (Chang, 2016; Montrul & Ionin, 2012).

In support of uneven acquisition despite an early start, Antonova Ünlü and Li (2016, 2018) report on a case study of a Turkish-Russian bilingual child growing up in a Turkish dominant environment. The child was studied longitudinally from age 11 months to 4 years and her acquisition of case marking, lexical aspect and gender agreement in Russian, the weaker language, was examined. Antonova Ünlü and Li found no difference between the bilingual child's acquisition of case marking and aspectual marking compared to monolingual Russian children of the same age. However, gender agreement was not age-appropriate by age 4;00, consistent with a pattern of incomplete acquisition. Therefore, acquiring the two languages before age 3;00 does not guarantee monolingual-like development in all structural properties when there is insufficient input in one of the languages. Other studies show incomplete acquisition of aspects of the heritage language *after* the first three years. Silva-Corvalán (2014) studied two of her grandchildren from first productions until age 6;00. The children were growing up in the United States, exposed to English through the mother and Spanish through the father and the grandmother. Daily amount of input ranged from 30% to 35% Spanish and 65% to 70% English for the older child (Nico), and 20% to 28% Spanish and 72 to 80% English for the younger sibling (Bren). Silva-Corvalán measured MLUs in both languages and studied the emergence and mastery of simple and complex verbal tenses in Spanish and English, the copulas *ser* and *estar* and the overt/null pronoun subject distribution in Spanish. She found monolingual-like development of Spanish and English up to age 3;00. Afterwards, about 30% of input in Spanish or less was not sufficient for the children to develop the verbal system of Spanish, although the children had no problems with the syntactic distribution of *ser* and *estar*. However, 75% of input in English was sufficient to promote the age-appropriate growth of English in an English-speaking society: by age 6;00 the children used productively the complex tenses in English. Similar to what Antonova Ünlü and Li's (2016, 2018) case study found, by age 6;00 Silva-Corvalán's grandchildren showed age-appropriate mastery of *some* properties of the heritage language, but not all of them. The incomplete acquisition of some grammatical properties attested at such early age is most likely due to insufficient exposure and use of the weaker language and not to maloperation of the language faculty or a maturational effect because the children had an early start.

Regardless of whether the heritage language already lags in development before or after age 3;00, its weakening is likely to continue and become more extensive when bilingual children start school and there is no academic support of the heritage language. With the onset of literacy and schooling, language development continues: there is expansion of vocabulary and learning of more abstract words. Children learn about denotation and connotation of words as used in literal and figurative language; children need to use language to define and describe complex concepts and they gain awareness of the multidimensional meanings of words and expressions. In terms of morphosyntax, children are called on to use more complex structures such as passives and complex sentences (noun, adjective and adverbial clauses) in spoken and written language. They need to use modals and conditionals (to argue and make hypotheses) and they become aware of registers: when to use formal or informal language, the difference between speaking and writing, and how to spell words. The input children receive at school is critical to promoting language development and bilingual balance and can alter the course of development of the two languages. Thordardóttir (2019) studied the vocabulary skills of first and third grade simultaneous and sequential bilingual children (English-French) in Montreal, where the two languages have equal status. She found that the difference between simultaneous and sequential bilinguals was more pronounced in the first-grade children than in the third-grade children, and the children who received 39% or less exposure to French on a daily basis consistently scored significantly lower than monolinguals in French. Thordardóttir concluded that amount and recency of input trump timing of input (age) in both simultaneous and sequential bilingual development. Although simultaneous bilingual children may acquire the two languages guided by innate mechanisms of UG like monolingual native speakers, it is not guarantee that later in life the two languages will reach native-like development. This most recent evidence confirms that sufficient language exposure in early childhood matters greatly because it has long-term consequences (Mayberry & Kluender, 2018). Sufficient input during childhood, more than maturation itself, plays a significant role in the degree of acquisition of the two languages of bilinguals. Insufficient language exposure and language use in childhood lead to the incomplete acquisition of aspects of the heritage language, even before age 3:00.

3. Are heritage speakers like second language learners?

In young adulthood many simultaneous and sequential bilingual heritage speakers display variable levels of proficiency, mostly non-native, in their heritage language. Even if the heritage language was their L1, as in sequential bilinguals, societal pressure and lack of access to schools with instruction in later childhood and adolescence lead to the heritage language not developing at the same level as the majority language and falling short of age-matched monolingual speakers of the language. Like late bilinguals with age of acquisition of the L2 after puberty, insufficient and variable input restricted to a particular environment (the home in heritage speakers, the formal classroom in many L2 learners) appears to be related to non-native, variable linguistic proficiency in the weaker language in the two bilingual groups, regardless of whether the weaker language was acquired as a L1 in heritage speakers or as a L2 in adult L2 learners. At the same time, heritage speakers and adult L2 learners vary in age of acquisition and experience with the weaker language in terms of timing (early for heritage speakers, late for L2 learners), environment or setting (predominantly naturalistic at home for heritage speakers vs. predominantly in an instructed, formal environment for L2 learners), and the modality of the input (auditory for heritage speakers and both written and auditory for adult L2 learners). In Montrul (2008, p. 126) I advanced the *Weaker Language as L1 Hypothesis*: "even when it may lag behind in development due to insufficient exposure to input and use, the weaker language in simultaneous bilingual acquisition is acquired as a first language, through the same cognitive and linguistic means to acquire the stronger language available in early childhood." The prediction of this hypothesis is that in adulthood heritage speakers retain signatures of L1 acquisition because they acquired the language within the critical period for language. Meisel (2014) supports this prediction. He claims that adult L2 learners do not acquire native competence in the L2 due to late acquisition while heritage language learners do not acquire native competence in their L1 due to impoverished input. While late learning of an L2 impacts the language learning mechanisms (there is a critical period for L2), impoverished input before the critical period does not impact the language learning mechanisms (no critical period for simultaneous bilinguals).

The prediction is that areas of language that develop early and require little input will be susceptible to critical period effects, such as phonological perception and to some extent phonological production. Although phonetic learning can be affected by experience past childhood, phonetic learning exhibits the two principles consistent with a critical period (Kuhl et al., 2005): A lack of exposure early in development to natural language results in the lack of normal language (Fromkin, Krashen, Curtiss, Rigler, & Rigler, 1974; Mayberry & Lock, 2003), and early experience with a particular language has indelible effects on speech perception (Flege, Yeni-Komshian, & Liu, 1999). Phonetic perception might exhibit a critical period in development. The relationship between perception and production is not straightforward, however. Child L1 studies show that perception develops before production in children, and that time and use are needed to develop stable target-like speech production (Strange & Broen, 1981). Therefore, phonological perception and production will develop at native or near-native levels in heritage language speakers.

Core syntax (word order) also develops early in L1 acquisition but is not that problematic in L2 acquisition, showing less evidence of age effects (Johnson & Newport, 1989). Grammatical areas that are more dependent on quantity and frequency of input, such as vocabulary and inflectional morphology will look "apparently" similar in L2 learners and heritage speakers. If implicit knowledge of language is subject to critical period effects, tasks that tap implicit knowledge of language will reveal native-like effects of morphology and syntax in heritage speakers but nonnative effects in L2 learners, especially in learners with lower proficiency in the language.

Au et al. (2002) was a pioneering study addressing the potential benefits of early exposure to the native language in heritage speakers with receptive knowledge of

Spanish (overhearers) and incipient L2 learners of Spanish. Participants were evaluated on overall accent and pronunciation of voiced and voiceless stops in Spanish. Morphosyntax, broadly conceived, was tested with a grammaticality judgment task and a phrase completion task. The results showed that the heritage speakers had significant advantages over L2 learners on accent ratings and pronunciation of stops, suggesting that early language exposure confers advantages for phonology even in very low proficiency bilinguals. However, the findings for morphosyntax were different: equal performance by heritage speakers and L2 learners on the two tasks suggested no discernible advantages of early language exposure in this domain. Since Au et al.'s trendsetting study, research has been delving deeper into the question of whether and how heritage speakers and L2 learners may differ from each other, focusing on specific grammatical areas, in different languages, and in bilinguals of higher proficiency.

In the next section we consider more findings of differences and similarities between L2 learners and heritage speakers on different modules of the grammar. We will see that the *Weaker Language as L1 Hypothesis* is borne out in phonology and that it is also born out in aspects of morphology and syntactic processing.

4. The impact of age of acquisition in different linguistic modules

4.1 Phonology

Heritage speakers have superior phonological abilities than L2 learners in phonological perception and production (Chang et al., 2011; Gor, 2014; Kim, 2020; Lukyanchenko & Gor, 2011). Chang et al. (2011) investigated the production of the English and Mandarin vowels /o^{u}, u/ and Mandarin /y/, long and short VOT stop consonants in the two languages, and the Mandarin fricatives /ʂ, ɕ, ʃ/ by heritage speakers and L2 learners of Mandarin. Although the heritage speakers did not realize all categories in the same way as the native speakers, they were more native-like than the L2 learners, supporting the early exposure advantage for phonology. Chang (2016) examined Korean heritage speakers' phonological perception of syllable final stops in Korean and in English and found that the heritage speakers were indistinguishable from the Korean native speakers: they even surpassed the perceptual abilities of English native speakers in English. Kim (2020) tested perception and production of lexical stress in Spanish, which is contrastive with some verb forms (*canto* 'I sing', *cantó*, 'she/he sang'), in nuclear position, prenuclear position and in unaccented contexts. In perception, the heritage speakers did not differ statistically from the native speakers, while the L2 learners performed at chance with two syllable words. In production, however, the heritage speakers patterned with the L2 learners instead of with the native speakers. The discrepancy found between heritage speakers' perception and production suggests that the link between speech perception and production is not always direct and that language use facilitates pronunciation (Oh et al., 2003).

Summarizing, heritage speakers show native phonological perceptual abilities in the heritage language and these abilities are likely due to very early exposure to the heritage language in infancy. There are age effects in the perception and production of phonology: the younger the exposure to the majority language (the L2) the less native the ability in the heritage language (Ahn et al., 2017; Montrul, 2008). Finally, perceptual advantages of early bilingualism may carry over to native abilities in the L2 as well (Chang, 2016). As far as production is concerned, some heritage speakers may display a heritage "accent", which is non-native but still more native than L2 accent (Polinsky, 2018).

4.2 Syntax and morphology

An early study showing that early language exposure is also relevant for aspects of syntax is Håkansson (1995), who studied heritage speakers and L2 learners of Swedish on their written and oral production of V2 and nominal morphology (gender, number and definiteness). Although the study was small and included only 11 participants, the results were very robust. All the L2 learners made between 8% and 58% errors with V2 (average 24%); except for one heritage speaker who made 3% errors, the other four made no errors with V2. With morphology, by contrast, the error rate for the L2 learners was 37% average, while for the heritage speakers it was 53%. This study demonstrates the importance of investigating morphosyntax in more detail, because different aspects of morphology and syntax have different developmental schedules and need different amounts of exposure and use for their acquisition and mastery.

Clitic pronouns (weak personal pronouns) straddle morphology and syntax. In Spanish, the preverbal or postverbal placement of object clitics in the sentence depends on the finiteness of the verb. In a production task, Montrul (2010) confirmed that L2 learners of Spanish do not produce as many clitics as Spanish native speakers and heritage speakers of Spanish. However, when they do produce clitic pronouns, bilingual speakers follow finiteness constraints. Spanish also has clitic climbing and clitic doubling. Clitic climbing is optional with many verbs, but native speakers prefer to place clitics before the finite verb in a sequence of an auxiliary verb and an infinitive, and so do heritage speakers, who do not differ from native speakers in this respect. By contrast, Montrul found that the L2 learners overwhelmingly preferred to leave clitics in the lower position, next to the non-finite verb, as in English. Montrul reported that heritage speakers were more native-like than the L2 learners in their acceptance of clitic doubling in clitic left-dislocations.

Further evidence of early exposure advantage for heritage speakers in syntax comes from two studies of comprehension of relative clauses in Spanish with syntactic ambiguity (*Alguien disparó contra el criado de la actriz que estaba en el balcón con su marido*. 'Somebody shot the servant of the actress who was in the balcony with her husband.'). Spanish is a high-attachment language (the servant was shot), while English is a low-attachment language (the actress was shot). In a written comprehension task testing comprehension of who was doing what, Keating, Van Patten, and Jegerski (2011) found that the L2 learners of Spanish preferred low attachment, as they do in English, while native Spanish speakers and heritage speakers preferred high attachment. Jegerski, Keating, and VanPatten (2016) used an online measure with monolingual native speakers in Mexico and heritage speakers of Mexican descent and found no differences between the native speakers and the heritage speakers in accuracy and reaction times: both groups preferred high attachment. In conclusion, research to date indicates that there are advantages of early childhood exposure detectable in specific aspects of syntax, which would be consistent with the *Weaker Language as L1 Hypothesis*, but more research is needed to confirm and expand the strength of these findings.

Morphology is perhaps the most vulnerable area in both L2 acquisition and heritage language acquisition. Although inflectional morphology appears early in young children (before age 3), it does not reach about 90% accuracy in production and comprehension until much later (depending on the morphology) because its eventual mastery and maintenance depend heavily on exposure. It has long been observed that L2 learners have significant problems with inflectional morphology, and this is an area where fossilization errors are very common (Lardiere, 2006; Franceschina, 2005). Slabakova (2008) has argued that compared to syntax and semantics, morphology is the sticky point and the bottleneck in L2 acquisition and Montrul (2018) reached the same conclusion for heritage language acquisition. Unlike research on phonology and syntax, research on morphology has not found as many advantages for heritage speakers due to early exposure to the language. Despite early emergence, this is the module of grammar most vulnerable to insufficient language exposure later in childhood, as the study by Silva-Corvalán (2014) clearly shows.

Case marking is a good example of difficulty in both groups. Hindi is a split ergative language with a rich morphological case system. Subjects of transitive perfective predicates are marked with ergative case (*-ne*). Human specific direct objects, indirect objects, and dative subjects are marked with the particle *-ko*. Montrul et al. (2019) compared knowledge of case marking in Hindi-English balanced bilinguals from India, L2 learners of Hindi with age of acquisition (AoA) of Hindi in adulthood and Hindi heritage speakers with AoA of Hindi since birth in oral production and acceptability judgments. The balanced bilinguals outperformed the L2 learners and the heritage speakers, who showed similarly lower command of the Hindi case marking system. The results of the two tasks showed no difference between the L2 learners and the heritage speakers, despite the fact that the two had very different language learning experiences.

However, Chung (2018) found that heritage speakers of Korean patterned more with native adult and child Korean speakers, compared to L2 learners of Korean, on knowledge of case ellipsis. Although Korean marks nominative and accusative case, in casual conversational contexts accusative (*-lul*) and nominative (*-ka*) markers are often dropped with canonical subjects and objects. In an elicited production task and a written task, Chung tested the linguistic factors that trigger case ellipsis (focus, animacy, definiteness) and found that the heritage speakers were more influenced by focus when prioritizing case ellipsis whereas the L2 learners paid more attention to animacy and definiteness. When it comes to a conversational phenomenon, heritage speakers outperform L2 learners in morphology.

Another area that has been the subject of intensive research is gender agreement in noun phrases. Montrul, Foote, and Perpiñán (2008) was the first study finding an effect of task and experience in gender agreement. Heritage speakers of Spanish and L2 learners of Spanish completed an oral production task and two written tasks. The native speaker group was at ceiling on the three tasks but the heritage speakers and the L2 learners were overall similarly inaccurate. However, the heritage speakers were more native-like and significantly outperformed the L2 learners in the oral production task whereas the L2 learners outperformed the heritage speakers in the two written tasks. Alarcón (2011) partially replicated Montrul et al.'s (2008) study with advanced proficiency L2 learners of Spanish and heritage speakers. She also found that heritage speakers were accurate with gender and were better than the L2 learners in an oral task.

To bring more clarity to the issue of modality and explicitness of tasks, Montrul et al. (2013) and Montrul et al. (2014) focused on the processing of spoken language exclusively. Montrul et al. (2014) implemented tasks to tap the participants' more automatic and implicit knowledge of grammatical gender. Spanish native speakers, L2 learners and heritage speakers of intermediate to advanced proficiency in Spanish completed three spoken word recognition experiments that varied on the degree of explicitness of the task: a gender monitoring task (GMT), a grammaticality judgment task (GJT) and a repetition task (RT). All the groups demonstrated sensitivity to gender agreement violations in Spanish noun phrases in general, but the heritage speakers were more native-like than the L2 learners in the more implicit task (the RT): the heritage speakers patterned with the native speakers, while the L2 learners showed the reverse response. We then have more evidence that when we control for modality, the explicitness of the task matters for these two types of learners. Therefore, the type of language learning experience early and later in life – naturalistic vs. instructed – impacts how L2 learners and heritage speakers store, process, and retrieve linguistic knowledge in different experimental tasks.

Montrul et al. (2013) examined the interaction of gender marking in nouns with diminutive formation to address whether the type of input received makes a difference in the acquisition in the two groups, especially with non-canonical ending nouns. Diminutives are a hallmark of Child Directed Speech in early language development and a highly productive morphological mechanism that facilitates the acquisition of declensional noun endings in many languages (Savickienė & Dressler, 2007). The hypothesis tested in this study was that Spanish heritage speakers should be more accurate at producing diminutives and at gender agreement with non-canonical nouns than L2 learners. Because heritage speakers were exposed to Spanish since birth, they were potentially also exposed to many instances of diminutives through Child Directed Speech, whereas L2 learners of Spanish were not exposed to such forms in early childhood. In an elicited production task, the native speakers performed at ceiling (100% accuracy). The L2 learners and the heritage speakers made gender errors, but heritage speakers were more accurate than the L2 learners with gender agreement in general, and with non-canonical ending nouns in particular. There was a much higher incidence of native-like ability in the heritage speakers than in the L2 learners. This study confirmed that early language experience and the type of input received confer some advantages to heritage speakers over L2 learners with early-acquired aspects of language.

To summarize, the collective results from all these studies suggest that it is hard to detect the influence of age effects in the acquisition of morphology because the acquisition of morphology also depends on specific language experience – as it relates to type of input and input modality. Input modality (auditory vs visual) and experience with written language affect the processing of language and linguistic performance of heritage speakers and L2 learners as measured by different tasks. Heritage speakers outperform L2 learners typically in tests that test vocabulary and morphology typical of early language development (such as diminutives) and do not tap metalinguistic knowledge. If linguistic knowledge elicited in this way is closer to the essence of grammatical competence than the knowledge elicited through reading and writing, then one may say that the heritage speakers have linguistic advantages not only on phonology, but on aspects of morphosyntax and syntax discourse as well.

4.3 Processing morphology

Montrul et al. (2014) was a language processing study and showed that heritage speakers were closer to native speakers in their sensitivity to gender violations as reflected in both accuracy and reaction times. Foote (2011) tested integrated knowledge of agreement, both gender agreement in nouns and adjectives and plural morphology in verbs (subject-verb agreement), in high proficiency heritage speakers and L2 speakers of Spanish. Participants completed a written agreement task, a working memory span task and a moving window visual sentence comprehension task. Foote did not find age of acquisition effects regardless of type of agreement (nominal or verbal): that is, the heritage speakers did not perform more accurately or faster than the L2 learners. Both L2 learners and heritage speakers were quite accurate and sensitive to agreement violations, demonstrating integrated knowledge of agreement. Foote's conclusion was that advanced proficiency and explicit training through classroom instruction may have helped the L2 learners, particularly because the main task used was written.

Another grammatical area where both L2 learners and heritage speakers of Spanish show non-native performance is with the production and interpretation of differential object marking (Arechabaleta, 2019). In Spanish, objects that are human, definite and specific are obligatorily marked with the preposition "a". Inanimate objects and indefinites are typically unmarked. Sagarra, Sánchez, and Bel (2019) investigated the processing of DOM in object relative clauses in Spanish. They found that the heritage speakers tested in their experiments exhibited processing patterns closer, but not identical, to those shown by monolingual native speakers than to L2 learners. While in accuracy the heritage speakers patterned with the L2 learners, in processing speed they patterned with the native speakers. Finally, Jegerski and Sekerina (2020) investigated production and online processing of differential object marking in as answers to subject and object wh-questions in Spanish using the Visual World paradigm, which is ideal to test participants with low levels of literacy, like heritage speakers. The task relies on auditory comprehension and does not require metalinguistic reflection. Although the study did not include a group L2 learners of Spanish of similar proficiency as the heritage speakers, Jegerski and Sekerina found that the heritage speakers were less accurate than the group of Spanish-speaking first generation immigrants tested as comparison group in the off-line measures. Interestingly, the processing patterns and proportion of looks to the target pictures used in the experimental design was no different for the two groups. This study also confirmed that at least in some aspects of morphological processing, when the tasks minimize explicit knowledge or writing, heritage speakers pattern with native speakers.

In summary, language experience as it relates to type of input and modality seems to affect the performance of heritage speakers and L2 learners on different tasks and experimental designs testing knowledge of morphology. Heritage speakers outperform L2 learners in tests that minimize metalinguistic knowledge and especially on oral production tasks an in morphological phenomena that is more frequent in conversational speech. Even though they are often inaccurate, they seem to process inflectional morphology like native speakers (a signature of L1 acquisition). The question that these findings raise is whether these effects are due to early age of acquisition of the heritage language or to how language exposure was experienced, that is, the modality of input (auditory) and spoken language use. One way to answer this question more directly, perhaps, would be to conduct studies of heritage speakers with different types of input experience (auditory only and auditory and written) and L2 learners with predominantly auditory experience as well, such as naturalistic L2 acquisition.

Many of the findings of these studies confirm what classroom-based research and teachers report about what they see in the classroom. While heritage speakers approach language tasks like linguistically naïve native speakers, L2 learners approach language like language teachers and textbooks. In a study of the use and interpretation of subjunctive in relative clauses in Spanish, Torres (2013) assigned a group of heritage language learners and a group of L2 learners a control group who received no focused instruction or to two groups that received explicit instruction (+ complex instruction and – complex instruction). All groups completed oral and written production tasks prior to the instruction (as baseline or pre-test), immediately after the instruction (immediate post-test) and one to two weeks later (delayed post-test). Torres found that the heritage language learners oriented primarily to the content of the tasks; that is, they were concerned with interpreting meaning of the prompts rather than learning the grammar. By contrast, the L2 learners recognized right away that the focus of the study was contrasting the forms of the present subjunctive and the indicative and paid close attention to form. In all aspects of grammar investigated, it is in morphology where experience with language exposure and recency of exposure seem to trump, depending on the task, the effects of age of onset of acquisition.

Thus, when comparing phonology, syntax and morphology, we find that there seem to be clear advantages for heritage speakers over L2 learners suggesting age effects for phonology and syntax. Even when many heritage speakers are quantitatively different from fully fluent native speakers, overall, the *Weaker Language as L1 Hypothesis* seems to be supported, since heritage speakers show more signatures of native-likeness in their heritage language than the L2 learners exhibit in the second language. For morphology, experience with input modality and recency of exposure appear to override timing of acquisition, but at present this depends on the task used in the studies to date. Studies of language processing, which seem to tap more implicit knowledge than explicit knowledge of language, also point to advantages for heritage speakers over L2 learners.

5. Discussion

Heritage language acquisition contributes to the debate on how age of acquisition of the language and environmental exposure contribute to bilingual language development. We have seen that many heritage speakers do not reach full competence in several grammatical domains due to insufficient exposure and use of the language during the critical period for language, which ranges between birth and puberty, depending on the grammatical phenomenon. Their heritage language retains signals of L1 acquisition in phonological perception, in aspects of syntax and in morphosyntactic processing. While this research is incipient, there is evidence that heritage speakers demonstrate implicit grammatical knowledge and native language processing, when compared to L2 learners with a later onset of acquisition of the L2. We have also seen that there is a spoken language advantage for heritage speakers compared to L2 speakers: they perform more native-like in oral production and comprehension tasks and when tested explicitly or implicitly on morphosyntactic structures that are more frequent in spoken language (e.g., case ellipsis in Korean, clitic climbing and clitic left dislocations in Spanish).

How do heritage speakers contribute to the debate on the critical period for language? They confirm that there are age effects in bilingualism, but perhaps not a strict critical period. Unlike late acquirers of American Sign Language, heritage speakers, like L2 speakers, received linguistic exposure during the period of brain maturation. This suggests that they would be more accurate in their L1 than late acquirers of ASL, who began to receive input after age 9. Heritage speakers are like L2 learners because they both developed language during the period of brain maturation. However, heritage speakers differ on which language was acquired when, and the timing of acquisition of the weaker language plays a role in native-like representations and processing. Thus, the non-native effects of heritage language speakers are mostly due to impoverished input during the critical period, when a language stabilizes. In their case, brain maturation and L1 development happened at the same time. By contrast, the non-native effects in L2 acquisition could be due to different learning mechanisms resulting from their learning experience, like enhanced reliance on explicit learning (Meisel, 2013). This is clear in phonology, and we showed that it also holds to come extent for morphology and syntax, but it can be harder to show for morphosyntax because we also know that a handful of very proficient L2 learners can achieve native ability (Foote, 2011). One possible venue of research to pursue would be a study of adult naturalistic L2 learners (with no formal education in the L2) compared to heritage speakers who did not receive formal instruction in the HL. If we can match them on proficiency and input, we may be more successful at isolating more precisely the effects of AoA.

Hartshorne, Tannenbaum, and Pinker (2018) consider that the decline in L2 ability with age is an epiphenomenon of culture and society, such as transitioning to the workforce. In addition to interference from the L1, late-emerging neural maturation processes compromise the circuitry responsible for successful language acquisition, whether specific to language or not. Mayberry and Kluender (2018) show that the brain areas activated during early and late L1 acquisition are different. Brain areas activated in L2 learners and late L1 learners are also different. Would different areas be activated in L2 learners and heritage speakers? Future research examining behavioral and neurolinguistic evidence could inform this debate.

References

Abrahamsson, N. (2012). Age of onset and nativelike L2 ultimate attainment of morphosyntactic and phonetic intuition. *Studies in Second Language Acquisition*, 34, 187–214. https://doi.org/10.1017/S0272263112000022

Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning*, 59, 249–306. https://doi.org/10.1111/j.1467-9922.2009.00507.x

Ahn, S., Chang, C., DeKeyser, R., & Lee-Ellis, S. (2017). Age effects in first language attrition: Speech perception by Korean-English bilinguals. *Language Learning*, 67, 694–733. https://doi.org/10.1111/lang.12252

Alarcón, I. (2011). Spanish gender agreement under complete and incomplete acquisition: Early and late bilinguals' linguistic behavior within the noun phrase. *Bilingualism: Language and Cognition*, 14, 332–350. https://doi.org/10.1017/S1366728910000222

Antonova Ünlü, E., & Li, W. (2016). Examining the effect of reduced input on language development: The case of gender acquisition in Russian as a non-dominant and dispreferred language by a bilingual Turkish-Russian child. *International Journal of Bilingualism*, 22, 215–233. https://doi.org/10.1177/1367006916666390

Antonova Ünlü, E., & Li, W. (2018). The acquisition of the weaker language. Evidence from the acquisition of Turkish cases by a Turkish-Russian child. *Linguistic Approaches to Bilingualism*, 8, 637–666. https://doi.org/10.1075/lab.16029.ant

Arechabaleta, B. (2019). Tracing language variation in Spanish: A multidisciplinary approach (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.

Au, T., Knightly, L, Jun, S., & Oh, J. (2002). Overhearing a language during childhood. *Psychological Science*, 13, 238–243. https://doi.org/10.1111/1467-9280.00444

Bley-Vroman, R. (1990). The logical problem of foreign language learning. *Linguistic Analysis*, 20, 3–49.

Carreira, M., & Kagan, O. (2011). The results of the national heritage language survey: implications for teaching, curriculum design, and professional development. *Foreign Language Annals*, 44, 40–64. https://doi.org/10.1111/j.1944-9720.2010.01118.x

Chang, C. (2016). Bilingual perceptual benefits of experience with a heritage language. *Bilingualism: Language and Cognition*, 19, 791–809. https://doi.org/10.1017/S1366728914000261
Chang, C. B., Yao, Y., Haynes, E. F., & Rhodes, R. (2011). Production of phonetic and phonological contrast by heritage speakers of Mandarin. *Journal of the Acoustical Society of America*, 129, 3964–3980. . https://doi.org/10.1121/1.3569736
Chung, E. S. (2018). Second and heritage language acquisition of Korean case drop. *Bilingualism: Language and Cognition*, 21, 63–79. https://doi.org/10.1017/S1366728916001218
Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, 27, 3–42. https://doi.org/10.1017/S0142716406060024
DeKeyser, R. (2003). Implicit and explicit learning. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 313–348). Blackwell. https://doi.org/10.1002/9780470756492.ch11
DeKeyser, R. (2018). The critical period hypothesis: A diamond in the rough. *Bilingualism: Language and Cognition*, 21, 915–916. https://doi.org/10.1017/S1366728918000147
Epstein, S., Flynn, S., & Martohardjono, G. (1996). Second language acquisition. Theoretical and experimental issues in contemporary research. *Behavioral and Brain Sciences*, 19, 677–758. https://doi.org/10.1017/S0140525X00043521
Flege, J. (2018). It's input that matters most, not age. *Bilingualism: Language and Cognition*, 21, 919–920. https://doi.org/10.1017/S136672891800010X
Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, 41(1), 78–104. https://doi.org/10.1006/jmla.1999.2638
Flores, C. (2010). The effect of age on language attrition: Evidence from bilingual returnees. *Bilingualism: Language and Cognition* 13, 533–546. https://doi.org/10.1017/S136672890999054X
Flores, C. (2015). Losing a language in childhood: A longitudinal case study on language attrition. *Journal of Child Language*, 42(3), 562–590. https://doi.org/10.1017/S0305000914000233
Foote, R. (2011). Integrated knowledge of agreement in early and late English-Spanish bilinguals. *Applied Psycholinguistics*, 32, 187–220. https://doi.org/10.1017/S0142716410000342
Franceschina, F. (2005). *Fossilized second language grammars. The acquisition of grammatical gender*. John Benjamins. https://doi.org/10.1075/lald.38
Fromkin, V., Krashen, S., Curtiss, S., Rigler, D., & Rigler, M. (1974). The development of language in Genie: a case of language acquisition beyond the "critical period". *Brain and language*, 1(1), 81–107. https://doi.org/10.1016/0093-934X(74)90027-3
Gor, K. (2014). Raspberry, not a car: Context predictability and a phonological advantage in early and late learners' processing of speech in noise. *Frontiers in Psychology*, 5, 1449. https://doi.org/10.3389/fpsyg.2014.01449
Granena, G., & Long, M. (2013). Age of onset, length of residence, language aptitude, and ultimate attainment in three linguistic domains. *Second Language Research*, 29, 311–343. https://doi.org/10.1177/0267658312461497
Gregg, K. (1996). The logical and developmental problems of second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *The handbook of second language acquisition* (pp. 49–81). Academic Press. https://doi.org/10.1016/B978-012589042-7/50004-9
Håkansson, G. (1995). Syntax and morphology in language attrition. A study of five bilingual expatriate Swedes. *International Journal of Applied Linguistics*, 5, 153–171. https://doi.org/10.1111/j.1473-4192.1995.tb00078.x
Hartshorne, J., Tenenbaum, J., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*, 177, 263–277. https://doi.org/10.1016/j.cognition.2018.04.007

Hawkins, R., & Chan, C. (1997). The partial availability of universal grammar in second language acquisition. The 'failed functional features hypothesis'. *Second Language Research*, 13, 187–226. https://doi.org/10.1191/026765897671476153
Jegerski, J., Keating, G. D., & VanPatten, B. (2016). On-line relative clause attachment strategy in heritage speakers of Spanish. *International Journal of Bilingualism*, 20, 254–268. https://doi.org/10.1177/1367006914552288
Jegerski, J., & Sekerina, I. A. (2020). The processing of input with differential object marking by heritage Spanish speakers. *Bilingualism: Language and Cognition*, 23(2), 274–282. https://doi.org/10.1017/S1366728919000087
Johnson, J., & Newport, E. (1989). Critical period effects in second language learning. The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21, 60–99. https://doi.org/10.1016/0010-0285(89)90003-0
Johnson, J., & Newport, E. (1991). Critical period effects on universal principles of language. The status of subjacency in the acquisition of a second language. *Cognition*, 39, 215–258. https://doi.org/10.1016/0010-0277(91)90054-8
Keating, G. D., VanPatten, B., & Jegerski, J. (2011). Who was walking on the beach? Anaphora resolution in Spanish heritage speakers and adult second language learners. *Studies in Second Language Acquisition*, 33, 193–221. https://doi.org/10.1017/S0272263110000732
Kim, J. Y. (2020). Discrepancy between heritage speakers' use of suprasegmental cues in the perception and production of Spanish lexical stress. *Bilingualism: Language and Cognition*, 23(2), 233–250. https://doi.org/10.1017/S1366728918001220
Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early speech perception and later language development: Implications for the "critical period". *Language Learning and Development*, 1, 237–264. https://doi.org/10.1207/s15473341lld0103&4_2
Laleko, O., & Polinsky, M. (2016). Between syntax and discourse: Topic and case marking in heritage speakers and L2 learners of Japanese and Korean. *Linguistic Approaches to Bilingualism*, 6, 396–439. https://doi.org/10.1075/lab.14018.lal
Lardiere, D. (2006). *Ultimate attainment in second language acquisition. A case study*. Lawrence Erlbaum Associates.
Lukyanchenko, A., & Gor, K. (2011). Perceptual correlates of phonological representations in heritage speakers and L2 learners. In *Proceedings of the 35th annual Boston University Conference on Language Development* (pp. 414–426). Cascadilla Press.
Mayberry, R. (1993). First-language acquisition after childhood differs from second-language acquisition: The case of American Sign Language. *Journal of Speech, Language and Hearing Research*, 36, 1258–1270. https://doi.org/10.1044/jshr.3606.1258
Mayberry, R., & Kluender, R. (2018). Rethinking the critical period for language: New insights into an old question from American Sign Language. *Bilingualism: Language and Cognition*, 21, 886–905. https://doi.org/10.1017/S1366728917000724
Mayberry, R. I., & Lock, E. (2003). Age constraints on first versus second language acquisition: Evidence for linguistic plasticity and epigenesis. *Brain and Language*, 87(3), 369–384. https://doi.org/10.1016/S0093-934X(03)00137-8
Meisel, J. (2010). Age of onset in successive acquisition of bilingualism. In M. Kail & M. Hickmann (Eds.), *Language acquisition across linguistic and cognitive systems* (pp. 225–248). John Benjamins. https://doi.org/10.1075/lald.52.16mei
Meisel, J. (2011). *First and second language acquisition: Parallels and differences*. Cambridge University Press. https://doi.org/10.1017/CBO9780511862694

Meisel, J. (2013). Sensitive phases in successive bilingual acquisition. The critical period hypothesis revisited. In C. Boeckx & K. Gromann (Eds.), *The Cambridge handbook of biolinguistics* (pp. 69–85). Cambridge University Press. https://doi.org/10.1017/CBO9780511980435.007

Meisel, J. (2014). Heritage language learners. Incomplete acquisition of grammar in early childhood. In E. Arias, Gutiérrez, M., Landa, I. & Ocampo, F., *Perspectives in the study of Spanish language variation. Papers in honor of Carmen Silva-Corvalán* (Verba, Anexo, 72, pp. 427–454). Universidad de Santiago de Compostela.

Meisel, J. (2018). Early child second language acquisition: French gender in German children. *Bilingualism: Language and Cognition*, 21, 656–673. https://doi.org/10.1017/S1366728916000237

Montrul, S. (2008). *Incomplete acquisition in bilingualism: Re-examining the age factor*. John Benjamins. https://doi.org/10.1075/sibil.39

Montrul, S. (2009). Knowledge of tense-aspect and mood in Spanish heritage speakers. *International Journal of Bilingualism*, 13, 239–269. https://doi.org/10.1177/1367006909339816

Montrul, S. (2010). How similar are L2 learners and heritage speakers? Spanish clitics and word order. *Applied Psycholinguistics*, 31, 167–207. https://doi.org/10.1017/S014271640999021X

Montrul, S. (2016). *The acquisition of heritage languages*. Cambridge University Press.

Montrul, S. (2018). The Bottleneck Hypothesis and heritage language acquisition. In J. Cho, T. Judy, M. Iverson, T. Leal, & E. Shimanskaya (Eds.), *Meaning and structure in second language acquisition* (pp. 149–177). John Benjamins. https://doi.org/10.1075/sibil.55.06mon

Montrul, S., Bhatt, R., Bhatia, A., & Puri, V. (2019). Case in Hindi as a weaker language. *Frontiers in Psychology*, 10, 461. https://doi.org/10.3389/fpsyg.2019.00461

Montrul, S., Davidson, J., de la Fuente, I., & Foote, R. (2014). Early language experience facilitates gender agreement processing in Spanish heritage speakers. *Bilingualism: Language and Cognition*, 17, 118–138. https://doi.org/10.1017/S1366728913000114

Montrul, S., de la Fuente, I., Davidson, J., & Foote, R. (2013). The role of experience in the acquisition and production of diminutives and gender in Spanish: Evidence from L2 learners and heritage speakers. *Second Language Research*, 29, 87–118. https://doi.org/10.1177/0267658312458268

Montrul, S., Foote, R., & Perpiñán, S. (2008). Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language Learning*, 58, 503–553. https://doi.org/10.1111/j.1467-9922.2008.00449.x

Montrul, S., & Ionin, T. (2010). *Transfer effects in the interpretation of definite articles by Spanish heritage speakers*. Bilingualism: Language and Cognition, 13(4), 449-473.

Montrul, S., & Ionin, T. (2012). Dominant language transfer in Spanish heritage speakers and second language learners in the interpretation of definite articles. *The Modern Language Journal*, 96(1), 70-94.

Newport, E. (2018). Is there a critical period for L1 but not L2? Comments on Mayberry & Kluender. *Bilingualism: Language and Cognition*, 21, 928–929. https://doi.org/10.1017/S1366728918000305

Oh, J. S., Jun, S. A., Knightly, L. M., & Au, T. K. F. (2003). Holding on to childhood language memory. *Cognition*, 86(3), B53–B64. https://doi.org/10.1016/S0010-0277(02)00175-0

Paradis, M. (2004). *A neurolinguistic theory of bilingualism*. John Benjamins. https://doi.org/10.1075/sibil.18

Polinsky, M. (2018). *Heritage languages and their speakers*. Cambridge University Press. https://doi.org/10.1017/9781107252349

Sagarra, N., Sánchez, L., & Bel, A. (2019). Processing DOM in relative clauses. Salience and optionality in early and late bilinguals. *Linguistic Approaches to Bilingualism*, 9, 120–160. https://doi.org/10.1075/lab.16020.sag

Savickienė, I., & Dressler, W. (2007). *The acquisition of diminutives: A crosslinguistic perspective*. John Benjamins. https://doi.org/10.1075/lald.43

Schachter, J. (1990). On the issue of completeness in second language acquisition. *Second Language Research*, 6, 93–124.

Schmid, M. (1918). The impact of frequency of use and length of residence on L1 attrition. In M. Schmid & B. Köpke (Eds.), *The Oxford handbook of language attrition* (pp. 288–293). Oxford University Press.

Schwartz, B., & Sprouse, R. (1996). L2 cognitive states and the full transfer/full access hypothesis. *Second Language Research*, 12, 40–72. https://doi.org/10.1177/026765839601200103

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics*, 10, 209–231. https://doi.org/10.1515/iral.1972.10.1-4.209

Silva-Corvalán, C. (2014). *Bilingual language acquisition: Spanish and English in the first six years*. Cambridge University Press. https://doi.org/10.1017/CBO9781139162531

Slabakova, R. (2008). *Meaning in the second language*. Berlin: De Gruyter. https://doi.org/10.1515/9783110211511

Sorace, A. (1993). Incomplete vs. divergent representations of unaccusativity in native and non-native Italian grammars of Italian. *Second Language Research*, 9, 22–47. https://doi.org/10.1177/026765839300900102

Strange, W., & Broen, P. (1981). The relationship between perception and production of /w/, /r/, and /l/ by three-year-old children. *Journal of Experimental Child Psychology*, 31(1), 81–102. https://doi.org/10.1016/0022-0965(81)90005-9

Thordardottir, E. (2019). Amount trumps timing in bilingual vocabulary acquisition: Effects of input in simultaneous and sequential school-age bilinguals. *International Journal of Bilingualism*, 23, 236–255. https://doi.org/10.1177/1367006917722418

Torres, J. (2013). Heritage and second language learners of Spanish: The roles of task complexity and inhibitory control (Unpublished doctoral dissertation). Georgetown University.

Tsimpli, I. M. (2013). (Evidence for) the language instinct. In C. Boeckx & K. Grohmann (Eds.), *The Cambridge handbook of biolinguistics* (pp. 49–68). Cambridge University Press. https://doi.org/10.1017/CBO9780511980435.006

Ullman, M. (2001). The neural basis of lexicon and grammar in first and second language, the declarative/procedural model. *Bilingualism: Language and Cognition*, 4, 105–122. https://doi.org/10.1017/S1366728901000220

White, L. (1989). *Universal grammar and second language acquisition*. John Benjamins. https://doi.org/10.1075/lald.1

White, L. (2003). *Second language acquisition and universal grammar*. Cambridge University Press. https://doi.org/10.1017/CBO9780511815065

White, L. (2018). Nonconvergence on the native speaker grammar: Defining L2 success. *Bilingualism: Language and Cognition*, 21, 934–935. https://doi.org/10.1017/S1366728918000214

Yeni-Komshian, G. H., Flege, J. E., & Liu, S. (2000). *Pronunciation proficiency in the first and second languages of Korean–English bilinguals*. Bilingualism: Language and cognition, 3(2), 131–149.

A different type of RC attachment resolution

Comparing bilingual and trilingual processing

Marina Sokolova and Roumyana Slabakova
University of Southampton

This study investigates attachment resolution of ambiguous relative clauses (RC) by second (L2) and third (L3) language speakers of English and Russian. Participants' sensitivity to the language of testing, social conventions, and a linguistic effect of the matrix verb (perception, non-perception) is investigated. Monolingual controls confirm the established cross-linguistic variation: high attachment in Russian and low attachment in English. Non-native speakers demonstrate a tendency to perform in the target-like manner in their non-native languages. Both native and non-native speakers favor high attachment after a perception matrix verb. Neither monolinguals nor L2/L3 learners rely on social conventions to interpret ambiguous RCs. In other words, non-native sentence comprehension appears to be sensitive to syntactic cues prompted by a perception verb in RC resolution.

Keywords: non-native processing, bilingual, trilingual parsing, structural, top-down, projection, bottom-up

1. Introduction

The study reported in this chapter approaches non-native processing at earlier stages of language acquisition in adulthood. It focuses on how second language (L2) and third language (L3) learners of intermediate proficiency process ambiguous relative clauses (RC). It also aims to describe the specific characteristics of their language processing. The study examines the issue of whether non-native languages, be they L2, L3 or L*n*, use similar strategies in RC resolution. If so, the known facts about L2 processing could be generalized to the entire field of non-native processing.

When developing a new L*n* linguistic system, a learner develops a certain set of processing strategies that ensure successful comprehension of a new language. Even though recent research investigating differences between native and non-native processing has offered different explanations (Dussias, 2003; Dussias & Sagarra,

https://doi.org/10.1075/lald.67.12sok

2007; Hahne & Frederici, 2001; Hopp, 2014a, 2014b), two approaches have put forward proposals about exactly how full comprehension in non-native languages is achieved: Full Transfer / Full Access / Full Parse (FT/FA/FP) (Dekydtspotter, Schwartz, & Sprouse, 2006) and the Shallow Structure Hypothesis (SSH) (Clahsen & Felser, 2018).

According to the first approach, FT/FA/FP, non-native speakers use structure-based parsing in their newly acquired languages. This proposal ties sentence processing to L2 acquisition. To acquire a new language, the parser performs input analysis and accumulates linguistic information, which can be added to the existing interlanguage grammar (Dekydtspotter, Schwartz, & Sprouse, 2006; Schwartz & Sprouse, 1996). In other words, the parser detects new linguistic features in the L2, which would be impossible without a full structural analysis of the incoming linguistic information.

According to the second approach, the SSH, non-native processing is *shallow*, i.e., non-native speakers rely on extra-syntactic information to interpret sentences in their L2 (Clahsen & Felser, 2006, 2018). Under this approach, a learner might disregard syntactic cues in L2 processing and use other sources of linguistic information to comprehend a sentence, for example contextual, lexical and pragmatic information.

The experiment described in this chapter addresses predictions from both approaches. To test whether the parser is sensitive to the selectional properties of the matrix predicate and is capable of parsing a sentence accordingly, our stimuli include perception verbs in matrix clauses, which is a linguistic cue favoring high attachment of the RC (full analysis is provided in the next section). To test whether non-native processing is governed by non-structural information, we operationalize lexical and pragmatic information in terms of social biases, since they prompt a certain type of RC resolution. Specifically, social biases in this study represent social conventions accepted in society; that is, an established tendency to assign certain activities as typically performed by certain social groups, for example, by men, women, children, adults, the elderly (see the next section).

The study is interested in capturing the moment of sentence processing when parsing decisions are made. For this reason, we use a self-paced reading technique, such that the participants cannot go back and reread the test sentences.[1] The interpretation decisions on RC attachment made by native and non-native speakers result from certain processing strategies implemented by the participants. The study

---

1. In this chapter, we only report the data of the participants' comprehension answer choice, which is representative of the processing mechanisms used by the participants in sentence comprehension. An additional analysis of the reading times, which is not included in this chapter, did not add any statistically significant factors that influence sentence processing.

focuses on the analysis of these strategies. It also broadens the scope of investigation of non-native processing mechanisms through comparisons between processing in L2 and L3.

2. Theoretical motivation

The study investigates attachment resolution of ambiguous RCs in English and Russian and uses cross-linguistic variation to ascertain whether non-native speakers of these languages process the RC as in their native language[2] or as in the target language. The study also investigates whether L2 and L3 speakers are sensitive to a perception verb in the main clause and use it as a linguistic prompt when parsing the rest of the sentence. Such a finding would be considered evidence for structural processing in non-native languages (Dekydtspotter et al., 2006). An alternative possibility is that participants use, instead, social conventions as the main cue for sentence processing in non-native languages. The latter result would be in line with the assumptions of the SSH (Clahsen & Felser, 2006, 2018).

To begin with, let us examine the structural ambiguity of the RC in (1), where both answers to the comprehension question are grammatical. In linguistic terms, option (a) is the result of high attachment (HA) of the RC to the first NP ('the mother'), a strategy that is generally preferred in Russian (Sekerina, 2002). Option (b) results from attaching the RC to a lower NP ('the boy') (LA) which is mostly preferred in English (see Fodor, 2002 for a summary).[3]

(1) Bill saw the mother of the boy [$_{RC}$ that was talking about cosmetics in the yard]. Who was talking about cosmetics?
   a. the mother
   b. the boy

Thus, in a sentence such as (1), RC resolution can have two answers, but only one is typically preferred in each language. RC-resolution choices can shed light on whether RCs are processed in a native-like fashion or not. Our experimental

2. For L3 speakers of English, the L1 effect in RC resolution is non-distinguishable from the effect of the L2. The choice of languages is deliberate and is explained in the section on Participants.

3. Cross linguistic variation in RC resolution has been widely studied in monolingual populations. It has been established that native speakers (NS) of Russian, French, Dutch, German, Greek, Spanish and Italian prefer HA, answer (a) in (1) (Cuetos & Mitchell, 1988; Hemforth et al., 1998; Zagar, Pynte, & Rtiveau, 1997), whereas NSs of English, Norwegian, Romanian, and Swedish prefer LA, answer (b) in (1) (Fernandez, 1999; Fodor, 2002).

design allows us to draw conclusions about the processing strategies that non-native speakers use to comprehend ambiguous RCs.

The two approaches to non-native processing mentioned above disagree about the role that syntactic information plays when L2 speakers process their L2. The first approach advocates for a full structural parse in the L2 (Dekydtspotter et al., 2006; see also Sprouse, 2011). The second approach claims that learners possess reduced capabilities for structure building in real time, suggesting instead that L2 processing is governed by non-structural information (Clahsen & Felser, 2006, 2018).

The account offered by Dekydtspotter and colleagues provides a detailed analysis of the role of processing in L2 acquisition. The authors claim that, when processing new linguistic input, the parser manages to establish even minimal L1–L2 distinctions. As a result, meaning is assigned to the newly detected features and they can be successfully acquired in the L2 (Lardiere, 2009; Slabakova & Montrul, 2008, among many others; see Jiang, 2004 for a counterargument). In a nutshell, Dekydtspotter et al. (2006) argue that parsing drives L2 acquisition, which would be impossible if sentence parsing in a non-native language is incomplete or shallow.

To support the notion that learners can fully parse the L2, Dekydtspotter et al. (2008) tested second-year learners of French at an American University to investigate whether learners could develop sensitivity to L2-specific properties from relatively early stages of L2 acquisition. Dekydtspotter et al. (2008) tested L2 speakers' knowledge of the default prosodic differences between their L2 (French) and their L1 (English). Thus, the study tested the predictions of the Implicit Prosody Hypothesis (Fodor, 2002), arguing that different placement of prosodic breaks in French and English entailed different structural parses of the ambiguous RC. Results showed that learners demonstrated a preference for HA in their L2-French, while also preferring LA in their native language, English. The authors concluded that the default prosodic organization of the new language was successfully acquired and that L2 speakers preferred HA in French because they were sensitive to its prosodic structure as early as the second year of learning. Therefore, the authors argued that L2 processing was governed by the same principles as processing in L1, which in their approach was structure building.

The findings by Dekydtspotter et al. (2008) were supported by Witzel, Witzel, and Nicol (2012) and by Hopp (2014b). Witzel et al. (2012) tested L1 speakers of Chinese who were highly proficient in their L2 English. The participants read part-way ambiguous RCs, which were disambiguated towards either HA or LA at the end of the sentence. The study measured the participants' reading time and established that disambiguation towards LA was more difficult for L2 speakers of English. The authors concluded that the participants demonstrated a clear preference for a certain type of RC attachment (HA), suggesting that they performed structure-based parsing in their L2, even though they did not fully demonstrate the LA preference normally attested for English.

Hopp (2014b; see also Hopp, 2014a) investigated the role of individual differences on non-native processing. The study matched native and non-native speakers by working memory capacity. The participants were asked to read ambiguous RCs while their eye-movements were monitored. The results showed that native and non-native speakers of English demonstrated similar processing behavior when their working memory capacity matched.

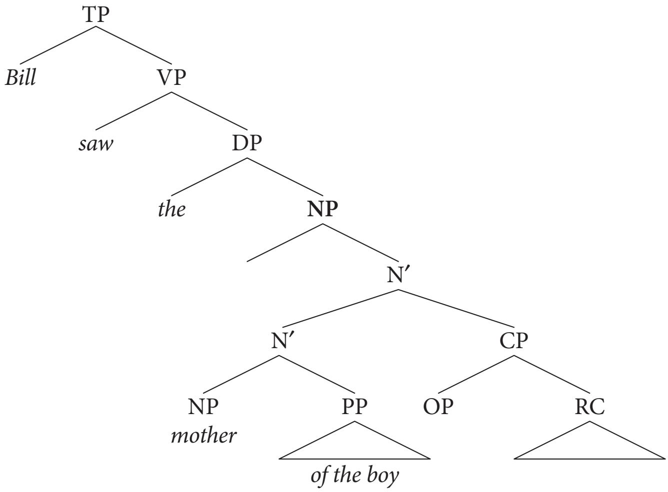
The second account discussed above proposes that non-native sentence processing is governed by non-structural information (Clahsen & Felser, 2018) and sentence parsing is performed in a "bottom-up manner" (Felser, 2018; personal communication). As articulated in the SSH, non-native speakers have trouble building syntactic structures in online processing. Instead, non-native processing is primarily governed by non-structural information (lexical, pragmatic, etc.) that allows the comprehender to interpret a sentence.

The SSH has received experimental support in studies by Felser. et al. (2003), Papadopoulou and Clahsen (2003), Felser and Cunnings (2012), among others. These studies compared native and non-native processing and attested several behavioral differences. For example, advanced non-native speakers in Papadopoulou and Clahsen (2003) did not show a clear preference in RC resolution and performed at chance, whereas native speakers manifested their respective language-specific patterns of RC resolution.

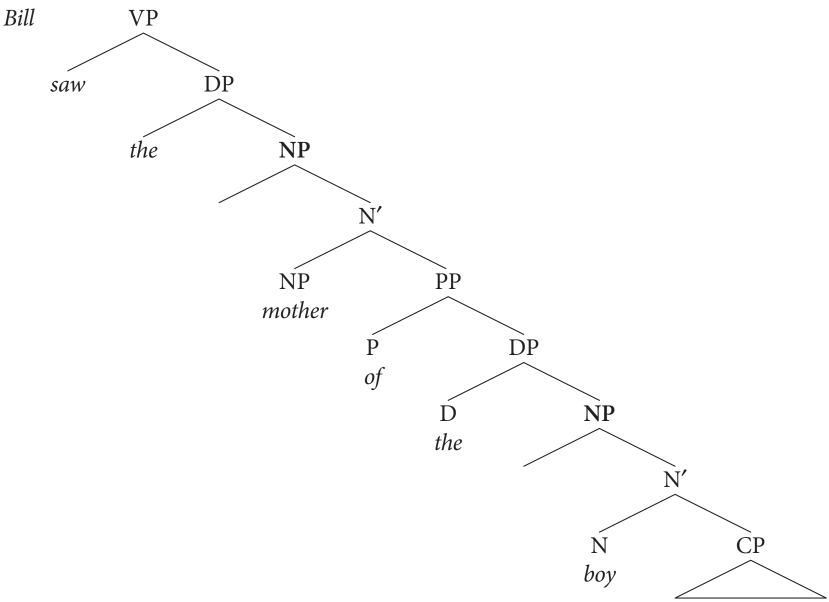
More support for the SSH is provided by an eye-tracking study by Felser and Cunnings (2012), where non-native speakers read complex sentences with reflexives. In processing, the informants consulted the ungrammatical antecedents which matched the reflexives in gender. The eye movements of native speakers stayed within the grammatical options. The authors concluded that nonnative speakers relied on lexical semantic information, whereas, native speakers demonstrated a structure-based parse (see also Jiang, 2004 for another argument for the SSH).

According to the SSH, non-native processing may rely on various non-structural cues. We include social conventions as providing such cues, allowing learners to interpret sentences such as those in (1). Social conventions are biases established in society that perceive certain activities to be typically masculine or typically feminine, etc. For example, the action of *talking about cosmetics* in (1) is more likely to be performed by a woman rather than a boy. Such a social bias would prompt HA of the RC in (1), as the agent of *talking about cosmetics* is presumed to be a woman, *the mother*. The idea is that social convention information can shape the sentence interpretation either towards HA, as in (2), or towards LA, as in (3).

(2) Bill saw **the mother** of the boy that was talking about **cosmetics**.



(3) Bill saw the mother of **the boy** that was talking about **cartoons**.

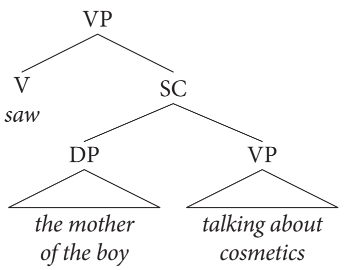


In order to test the assumptions of the SSH for non-native processing, the experiment we report in this chapter includes social conventions as a variable. This is because the SSH predicts that participants tested in their non-native languages will be sensitive to non-syntactic information, as a result prefering HA in sentences such as (2) and LA in sentences such as (3). On the other hand, because the SSH assumes that native language processing is largely structural, we would expect monolingual English speakers to show a preference for LA in English while monolingual Russian speakers should prefer HA in Russian regardless of the social conventions reflected in the examples.

To test the predictions for structural processing in non-native languages, as proposed by Dekydtspotter et al. (2006), our study includes a perception verb in the matrix clause. This verb type is expected to favor HA even in a LA language such as English (Grillo et al., 2015). The verb-type effect will be a processing cue signaling adjustment of mental structure favoring HA. Following Dekydtspotter et al. (2008), our study investigates whether L2 and L3 learners of English and Russian at intermediate stages of non-native language proficiency show sensitivity to linguistic cues in L2/L3 processing.

What is the prediction of structural processing based on? The processing effects of a perception verb were studied by Grillo and Costa (2014), who argue that when a perception verb is placed in the matrix clause of a sentence with an ambiguous RC, it triggers a structural anticipation for an eventive complement, like *Bill saw (what event?)*. The eventive complement in (4) is structurally different from the RC in (5).

(4) Bill saw [$_{\text{SC}}$ the mother of the boy talking about cosmetics in the yard].
[$_{\text{S}}$ NP$_{\text{Bill}}$ [VP$_{\text{saw}}$ [SC]]]

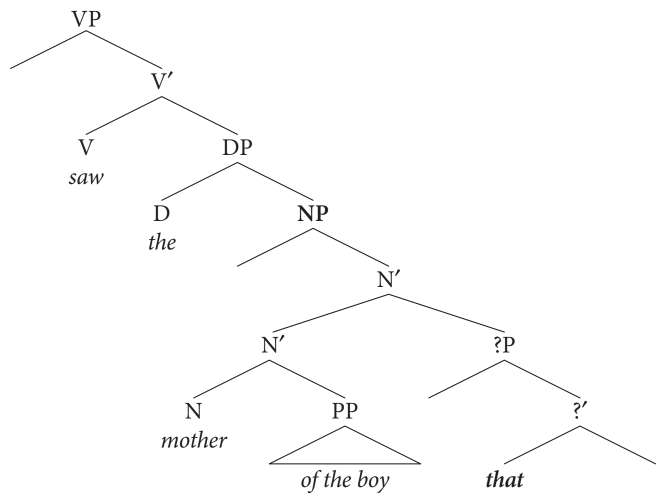


The subordinate clause in (4) is a Small Clause (SC); it provides an event-oriented interpretation of the sentence, *Bill saw (what event?)* – the event of "talking about cosmetics performed by the mother of the boy." The eventive complement, the SC, modifies the matrix predicate. It makes the NP '*the mother*' the only grammatically licensed doer of the action of *talking*.

Grillo and Costa (2014) propose that a perception verb such as *see* in the matrix clause in sentences such as in (1) favors HA cross-linguistically. This prediction has received experimental evidence from Grillo et al. (2015) and from Sokolova and Slabakova (2019). Grillo et al. (2015) tested monolingual speakers of English, who changed their preference for LA to HA when the sentences had a perception verb in the matrix clause. Sokolova and Slabakova (2019) checked the effect of a perception verb with native and non-native speakers of English and Russian. A perception verb prompted a change to HA in English and maintained HA in Russian. Non-native speakers followed the structural prompt of a perception verb in the same way as monolinguals.

The effect found in Grillo et al. (2015) can be explained through the following order of structural operations. The parser generates a structural anticipation for an eventive complement in (4), which is supported by the incoming string of words until the complementizer *that* is processed. When the parser encounters *that*, it "realizes" that the structure generated for the SC (4) is impossible and performs a minimal adjustment of the structure, as shown in (5).

(5) Bill [$_{VP}$ saw [$_{DP}$ the mother of the boy [$_{?P}$ that…]]]



After the head NP is processed, the parser anticipates that the SC analysis will continue and generates a phrase for the upcoming verbal element, (?P) in (5). On processing *that*, the parser dismisses the originally anticipated projection for the VP in the SC (see (4)) and adjusts the parse to accommodate the incoming RC. The RC is attached to the already generated head NP and modifies its head. Consequently, HA of the RC is preferred.

Our research continues the line of investigation started by Grillo et al. (2015) and by Sokolova and Slabakova (2019). We extend the scope of our research to non-native processing that includes L2 and L3. If all target groups prove to be sensitive to the effects of a perception verb, the assumption by Dekydtspotter et al. (2006, 2008) – that sentence processing is structural – would be supported and could be extended to a broader field of processing in L1, L2, L3 or L*n*.

In summary, our study investigates whether non-native speakers at the intermediate (B1)[4] level of proficiency are sensitive to the attachment preference of the TL and, thus, whether they prefer HA in Russian and LA in English. The study also examines whether native and non-native speakers use different processing strategies. Under the SSH, non-native speakers are expected to rely more on social conventions in RC resolution, whereas monolinguals should show language specific RC resolution (Clahsen & Felser, 2018). On the other hand, if both native and non-native processing uses a structural approach to parsing, then sensitivity to the effects of a perception verb can be expected in both groups of speakers.

We put forward the following hypothesis, based on the literature reviewed above: Native and non-native processing is governed by similar psycholinguistic mechanisms; both native and non-native processing is sensitive to structural prompts and does not rely on non-syntactic information in sentence processing for comprehension.

The hypothesis motivated the following Research Questions (RQ).

- RQ1: Do non-native speakers of English and Russian show L1-like patterns of RC resolution in their respective L2s or L3s?
- RQ2: Is sentence processing structural in both native language and non-native languages?
- RQ3: Is non-native processing governed by social biases?

3. Method

The RQs motivated the choice of the target groups of participants and informed the experimental method and its design. Before conducting the experiment, an IRB approval, protocol #1602915700, was obtained. Participants were provided with all the necessary information related to their role in the experiment. Participation in the study was voluntary and the participants could quit the experiment at any moment without any consequences.

Our experiment focuses on eliciting the preferred patterns of RC resolution in native and non-native languages. It investigates a possible switch of attachment preference between high and low attachment languages, i.e., between Russian, be it the L1 or L2, and English, either the L2 or L3. In other words, the study aims to determine whether learners can switch in RC resolution preferences to conform to the target language preferences.

4. B1 is the 3rd level of proficiency in the Common European Framework of Reference for Languages, which covers the range of 6 levels, from *beginner* to *proficient user*: A1-A2-**B1**-B2-C1-C2.

The aim of the study with the bilingual groups is straight-forward – to determine whether speakers can switch to the TL-like pattern of RC resolution in their L2 at the intermediate level of proficiency.[5] This question is addressed through comparisons of the participants' performance in Russian and English, as L2s for Russian-English and English-Russian bilinguals and L1s for English and Russian monolinguals. The trilingual groups are L3 speakers of English at an intermediate level of proficiency. By comparing the trilingual groups to L2 speakers, the study seeks to determine whether a TL-like preference for LA in English can be acquired when more than one language in the participants' background is a HA-language.

For the trilingual groups, both L2s, French and German, belong to the HA group for RC resolution (see Fodor, 2002 for summary). Their native language, Russian, is also a HA language. If we establish a possible preference for HA in L3 English, influence from either the L1 or the L2 cannot be distinguished. This choice of experimental groups is deliberate; the study does not investigate how previously learnt languages influence L3 acquisition (see Rothman, 2010 and Slabakova, 2017 for relevant discussion). Rather, our experiment compares processing in L2 and L3 to establish whether non-native processing in different acquisition contexts is governed by similar mechanisms.

3.1 Participants

The participants in the current study were adult speakers of English and Russian. The study included 10 monolingual speakers of English, 9 monolingual speakers of Russian, 14 native speakers of Russian with English as their L2 (Russian – English), and 14 native speakers of English with Russian as their L2 (English – Russian). We also included two groups of L3 speakers of English. Both L3 groups were native speakers of Russian. However, the participants' L2s were different. Fifteen people spoke French as their L2 (the Russian – French – English group, RFE). Eleven participants spoke German as their L2 (the Russian – German – English group, RGE).

The level of proficiency was measured in the target languages of the study; the L2 for the English-Russian group and the Russian-English bilingual group, and the L3 for the Russian-French-English group and the Russian-German-English groups. A C-test asked participants to fill in 60 gaps across 3 independent texts (20 gaps per text). The texts were balanced for number of sentences and sentence length. Those participants who completed 30% to 60% of the C-test correctly were classified as

5. The level of participants' proficiency in their respective L2s or L3s is called 'intermediate' to provide correlations with the Common European Framework of Reference for Languages, where it would be labelled as B1. For the study, the level of the English language classes the participants were taking and the results of a C-test were taken into account for the acceptance criteria.

intermediate level speakers of English or Russian and were invited to participate in the experiment. Furthermore, the cut-off point of 30% was calculated as the lowest possible score which could not be obtained through incidental guessing of a preposition or a repeated word. All participants were recruited from college-level classes of English or Russian. The classification of the participants as intermediate matched the level of the English or Russian language course they were currently taking. For the learners of English, their courses were designed using textbooks pertaining to the B1 level of the Common European Framework of Reference for Languages.

For L3 speakers of English, proficiency in their respective L2s was assessed independently. They took language proficiency exams in their respective L2s at their Universities in Russia twice a year. The exams contained reading, writing, speaking, and listening parts. By the time of participation in the experiment, their most recent University exams in either L2-German or L2-French corresponded to the advanced (C1) level of proficiency in the Common European Framework of Language Proficiency Assessment. Besides, most participants had visited the countries of their respective L2s several times for both work and study. All the trilingual participants were taking at least one of their University courses in their L2 at the time of the experiment. The background information of the participants, including their level of proficiency in the target language is given in Table 1.

Table 1. Participant background information

Group	**E (NSs of English)**	**R (NSs of Russian)**	**ER (L1-English, L2-Russian)**	**RE (L1-Russian, L2-English)**	**RFE (L1 Russian, L2 French, L3 English)**	**RGE (L1 Russian, L2 German, L3 English)**
N participants	10	9	14	14	15	11
C-test % correct	–	–	37% 30–60	45% 30–60	54% 30–60	56% 40–57
Length of exposure to the target language	–	–	2 years: 4 hrs/week	4 years: 2 hrs/week	6.7 years (4 hrs/week)	5.4 years (4 hrs/week)
Mean age	40	29	21	30	24	25

In summary, all non-native participants demonstrated an intermediate level of proficiency in the target languages of the experiment (English and Russian), according to the C-test. Despite the fact that the average C-test score varies across the groups, linear regression analysis with R revealed no significant difference between the four groups of non-native speakers, $p < .3$. The analysis checked whether there were statistically significant differences between groups, when such factors as

C-test score, length of L2/L3 exposure and self-assessment were taken into account. The proficiency scores demonstrated that the participants were not beginners, but they were not yet advanced in their respective L2 or L3. The L3 groups were quite proficient in their L2s, which was confirmed by their general academic record and their results in standardized proficiency tests at their home Universities. All the participants were adults, either college students or professionals with degrees not lower than BA.

3.2 Materials

The study used a two-by-two design in both English and Russian. The first condition employed a perception vs. a non-perception matrix verb. Recall that a perception verb is expected to favor HA across all languages and in all experimental groups. A non-perception verb would, in turn, favor a language-specific pattern of RC resolution.

The second condition manipulated social biases that would prompt a certain type of RC resolution. The social conventions used in the experiment were selected based on a survey taken by 20–25 adult native speakers in the U.S. and Russia. The survey included a list of activities like *talking about cosmetics*, *buying flowers*, *playing in the yard* and a list of possible doers, like *a man*, *a woman*, *a child*, etc. There were 20 activities in the full list. Participants were asked to match an activity with the most likely doer of this activity. The participants' choices at 85% or higher were selected to design the experimental stimuli for social bias. A sample set of experimental sentences for English is given in Table 2.

Table 2. Sample stimuli quadruple in English

Perception / HA Bias	Bill *saw* **the mother** of the man that was talking about **cosmetics**.
Non-Perception / HA Bias	Bill <u>arrested</u> **the mother** of the man that was talking about **cosmetics**.
Perception / LA Bias	Bill *saw* the son of **the woman** that was talking about **cosmetics**.
Non-Perception / LA Bias	Bill <u>arrested</u> the son of **the woman** that was talking about **cosmetics**.

The English stimuli included NP object head nouns of two different genders and used social gender biases to assign certain activities to be performed by either men or women. This approach could not be used in Russian because grammatical gender is overtly marked and head nouns of different genders would entail gender marking on the complementizer, which would disambiguate the target sentence. Therefore, Russian stimuli used a different convention, splitting head nouns between different social groups by age. Table 3 shows a sample set of target sentences for Russian.

Table 3. Sample stimuli quadruple in Russian (English equivalents are shown)

Perception / HA Bias	Bill *saw* **the son** of the man that was **playing** in the yard.
Non-Perception / HA Bias	Bill <u>arrested</u> **the son** of the man that was **playing** in the yard.
Perception / LA Bias	Bill *saw* the father of **the boy** that was **playing** in the yard.
Non-Perception / LA Bias	Bill <u>arrested</u> the father of **the boy** that was **playing** in the yard.

The experiment contained 40 target sentences and 40 distractors. The distractors were complex sentences that did not contain ambiguous RCs, like (6):

(6) My friend likes the coffee that I brought her from Brazil last year.

The order of the sentences was randomized by the program Linger so that each participant saw a unique sequence of experimental items.

3.3 Procedure

The experiment included three stages: a pre-experimental part, the experiment, and a wrap-up phase. During the pre-experimental part, participants completed the linguistic background questionnaire and completed the proficiency measure in the non-native languages in which they were tested. Monolingual speakers of English and Russian were exempt from the language proficiency test. The pre-testing took the monolinguals 5–7 minutes and the non-native speakers 20–25 minutes to complete.

The main experiment started with a trial session where participants were introduced to the format of the experiment (a self-paced reading task). Each self-paced sentence was followed by a comprehension question. The comprehension questions had two answer choices which could be selected by pressing either the key F or the key J. To move forward, the participants had to press the SPACE bar. The main experiment took the participants 30–40 minutes to complete. Upon completion of the experiment, the participants had an opportunity to ask questions about the study and their results.

3.4 Data analysis

Data were analyzed in R using a mixed linear model, software package lmer4 (Bates et al., 2015). The dependent variable was Noun Choice, standing for the answer choice in the comprehension question that showed high- or low-attachment preferences in RC resolution.

The independent variables were the Verb Type (perception vs. non-perception), Social Biases (favoring HA vs. LA), and Group (ER, RFE, RGE, RE in analysis 1, and

E, R, ER, RFE, RGE, RE in analysis 2). Verb Type, or the type of the matrix predicate, tested whether a perception verb favored HA of the RC across the two languages of the experiment. Social Bias measured whether the answer to a comprehension question depended on the activity expressed by the embedded verb and a social bias to assign this activity to a certain head noun. The third factor was Group, which allowed for comparisons between native and non-native speakers as well as for comparisons between the groups of L2 speakers vs. the groups of L3 speakers. The dependent variable, Noun Choice, was the data calculated as percentile preference for a certain type of RC resolution, HA or LA. The results are presented with HA as a reference category.

The statistical analysis used a mixed linear effect model to measure the effect of three factors: Group, Verb Type, and Social Bias on the answer choice (Nchoice) in a comprehension question: model.english = lmer(PctNoun1 ~ VerbType_factor*SocialBias_factor*Group_factor + (1 | Participant) + (1 | Item), data = full_data_set, REML = FALSE). The model had Participant and Item as random effects. The Results section below reports the significance or non-significance of the main factors, supported by the lsmean data. Two separate analyses were carried out: once with the non-native speaker groups only, and then with all the participant groups. Therefore, in the first analysis the Group factor had 4 levels, while in the second analysis it had 6 levels.

3.5 Results

Results are presented in two stages. In the first stage, non-native speakers are compared to each other. In the second stage, the two monolingual control groups are added to the analysis. The need for a two-stage data presentation was motivated by the complexity of the factor Group, since the study included 6 participant groups in total. Recall that the target languages include a LA-language, English, alongside a HA-language, Russian.

L2 and L3 Speakers. The results of all non-native speakers analyzed together are presented in Table A in the Appendix. The analysis yielded only one significant main effect: Verb Type (*Estimate* = 0.07, *SE* = 0.02, *df* = 162, *t* = 3.50, *p* < .001). This suggests that a perception verb type favored overall preference for HA in the four experimental groups.

Table 4 demonstrates how verb type (perception verb vs. non-perception verb) affected participants' behavior. The Perception condition returned 58% preference for HA vs. the Non-Perception condition 51% for HA.

Table 4. Verb type effect on high attachment preference

		After a perception verb		After a non-perception verb	
Preference for high attachment		**58%**		**51%**	
VerbType_factor	lsmean	SE	df	lower.CL	upper.CL
Non-Perception	0.511	0.024	83.555	0.462	0.561
Perception	0.579	0.024	83.555	0.529	0.628

There was no effect of social bias on RC attachment resolution in either the L2 or L3. As demonstrated by Table 5, social bias favoring HA represented 55% of all HA choices, whereas the bias towards LA reduced this score by 1% only, $p < .8$ (see Appendix A for full statistical analysis).

Table 5. Effect of social bias on high attachment preference

		Favoring HA		Favoring LA	
Preference for high attachment		**55%**		**54%**	
Social_factor	lsmean	SE	df	lower.CL	upper.CL
forHA	0.547	0.024	83.555	0.498	0.597
forLA	0.543	0.024	83.555	0.493	0.592

There was no significant difference between the L2ers and L3ers. The group effect *p*-value for the populations tested in their non-native languages was $p < .3$ (see Appendix A for full analysis). Table 6 demonstrates the preferred percentage score for HA in each group of non-native speakers.

Table 6. Group effect on high attachment preference

		ER	RFE	RGR	RE
Preference for high attachment		**51%**	**60%**	**51%**	**57%**
Group_factor	lsmean	SE	df	lower.CL	upper.CL
ER	0.505	0.044	58.333	0.417	0.594
RFE	0.597	0.042	58.333	0.511	0.682
RGE	0.507	0.049	58.333	0.407	0.607
RE	0.571	0.044	58.333	0.483	0.660

Non-Native Speakers and Monolinguals. The results of all the experimental groups together allow for a comparison between monolingual processing and processing in non-native languages (Appendix, Table B). When the data of native and non-native speakers were analyzed in one pool, the analysis unveiled two significant main effects: Group (*Estimate* = 0.06, *St. Error* = 0.02, *df* = 73, *t* = 4.12, $p < .001$) and Verb Type (*Estimate* = 0.27, *St. Error* = 0.06, *df* = 219, *t* = 3.81, $p < .001$). In other words, a perception verb had a similar effect on the entire population of participants (VerbType effect, $p < .001$). At the same time, the preference for RC resolution differed by Group (Group effect, $p < .001$). These seemingly contradictory results are explained below.

To begin with, a perception verb in the matrix clause favored HA in the entire data pool, as shown in Table 7.

Table 7. Verb type effect on high attachment preference, all groups combined

		After a perception verb		**After a non-perception verb**	
Preference for high attachment		**56%**		**50%**	
VerbType_factor	lsmean	SE	df	lower.CL	upper.CL
Non-Perception	0.496	0.022	96.788	0.451	0.541
Perception	0.555	0.022	96.788	0.510	0.601

As becomes evident from Table 7, the participants preferred HA after a perception verb 6% more often than after a non-perception one. The homogeneous effect of the matrix verb was supported by the lack of a significant interaction VerbType*Group. Nevertheless, the group data on the effect of the matrix predicate in Figure 1 present insightful information. These data are used for illustrative purposes.

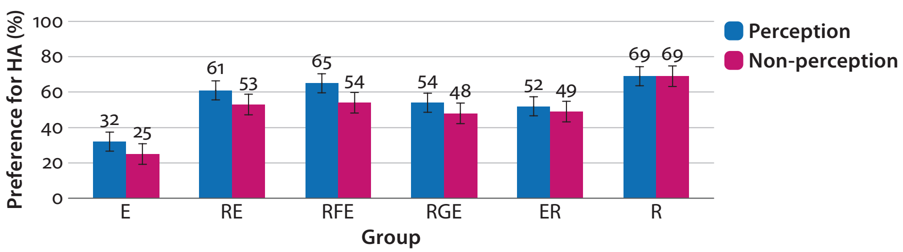


Figure 1. Descriptive Statistics: Noun choice by verb type in each group

Figure 1 demonstrates a noticeable difference between the effect of a perception verb in Russian and English. A perception verb did not override the default preference for LA in English monolinguals (E), although it did lead to an increase in HA responses. Moreover, it had no effect on HA in Russian monolinguals (R). It is also noticeable that a perception verb tended to influence RC resolution in English more than in Russian, whether in native or non-native processing.

The between-group difference in the participants' sensitivity to the effect of a perception verb becomes clearer in the analysis of the Group factor. The Group factor included 6 experimental groups, which were compared to each other. The statistical significance of the Group factor indicates that RC resolution varied depending on whether the participants were monolingual (Russian / English), or L2 learners of either English or Russian, or L3 learners of English. Table 8 demonstrates how the contrasts within the Group factor were set in R.

Table 8. Levels of group factor in R

contrasts(full_data_set$Group_factor)

	[1]	[2]	[3]	[4]	[5]
E	−0.863	−0.239	−0.130	−0.082	−0.034
ER	0.136	−0.739	−0.130	−0.082	−0.034
RFE	0.136	0.260	−0.630	−0.082	−0.034
RGE	0.136	0.260	0.369	−0.582	−0.034
RE	0.136	0.260	0.369	0.417	−0.465
R	0.136	0.260	0.369	0.417	0.534

As can be gathered from Table 8, there were 5 levels of comparison within the Group factor. They are in columns and marked as [1], [2], [3], [4] and [5] in the table. The order of comparisons can be tracked by the changing balance between the negative and the positive values. In R, the decimals demonstrate how the variables are distributed around zero on the X axis. For example, in column [1] group E is placed to the left of zero (negative), and groups ER, RFE, RGE, RE and R to the right of zero (positive). This demonstrates that the comparison occurs between groups E and ER. In column [2], there are two groups whose data are distributed to the left of zero (negative) – E and ER. The data of the other four groups are distributed to the right of zero (positive). Thus, the second level of comparison (column [2]) places the main contrast between groups ER and RFE. The other three columns follow the same pattern of distribution.

The model returned only *one statistically significant* contrast within the Group factor. It was the first level of analysis reflected in column [1], (see Table B in the Appendix). In particular, English monolinguals were significantly different from the Anglophone learners of Russian (ER). This difference can be observed in the individual data on RC resolution provided in Figure A in the Appendix.

The lack of a significant difference at the other levels of comparison means that L2 learners of Russian were not significantly different from L2/L3 learners of English (column [2]). There was no significant difference that would separate the two trilingual groups (column [3]). A better visual representation of the RC resolution in groups RFE and RGE can be obtained from Figure C in the Appendix.

L3 speakers of English (RFE and RGE) were not different from L2 speakers of English (RE) (column [4]). It is important to notice that Russian monolinguals (R) were not statistically different from Russian learners of English (RE, column [5]). Please, consult the individual data for groups R and RE in Figure B in the Appendix.

The percentile score of HA preferences and R statistics for the comparisons within the Group factor are provided in Table 9 below.

Table 9. Group effect on high attachment preference

	E	**ER**	**RFE**	**RGE**	**RE**	**R**
Preference for high attachment	**29%**	**51%**	**60%**	**51%**	**57%**	**69%**
Group_factor	lsmean	SE	df	lower.CL		upper.CL
E	0.287	0.055	73.999	0.176		0.398
ER	0.505	0.047	73.999	0.411		0.599
RFE	0.596	0.047	73.999	0.505		0.687
RGE	0.506	0.053	73.999	0.400		0.613
RE	0.571	0.047	73.999	0.477		0.665
R	0.688	0.058	73.999	0.571		0.806

Monolingual speakers of English showed a preference for LA, which was as expected. Monolingual speakers of Russian, on the other hand, showed a preference for HA, which patterns with the RC resolution most often preferred in Russian. L2 and L3 speakers tested in English displayed a general preference for HA, i.e., the preference was above 50%. However, the rate of their HA choices was still lower than in the monolingual Russian group (see data in Figures B and C in the Appendix).

Group RGE, L3 English, demonstrated the same preference for HA (51%) as the ER group, L2 Russian. We take this as evidence for a potential to acquire TL-like parsing in the non-native language. Bilingual and multilingual speakers of English demonstrate noticeably lower preference for HA in English than in their L1 Russian. At the same time L2 speakers of Russian prefer HA much more often in their L2 Russian than in their L1 English.

The results in Table 9 demonstrate that English monolinguals stayed within the preference for LA in their native language (see Figure 1). The data in Figure 1 and Table 9 clearly illustrate that a perception verb did not override the default preference for LA in RC resolution in native English, i.e. there was no switch to HA preference in the sentences with a perpcetion verb. It had almost no effect in Russian (R and ER), where HA is generally preferred. All groups tested in English, be it their native or non-native language, showed a greater effect of a perception verb on RC resolution than those tested in Russian. In other words, the effect of the verb type was mediated by the language in which participants were tested and not by the native language of the participants.

The statistical analysis did not return any significant effect of Social Bias on RC resolution; that is, the entire population of participants did not rely on this type of non-structural information in RC resolution. While a possible distinction between native and non-native speakers could be revealed through a significant interaction of the factors Group and Social Bias, no such interaction was found (see Appendix for detail). Table 10 provides the relevant data.

Table 10. Effect of social bias on high attachment preference

		Favoring HA		**Favoring LA**	
Preference for high attachment		**53%**		**52%**	
Social_factor	lsmean	SE	df	lower.CL	upper.CL
forHA	0.528	0.022	96.788	0.483	0.573
forLA	0.523	0.022	96.788	0.478	0.568

In summary, our results showed that verb type had a robust influence on RC attachment resolution across groups (monolinguals, L2 and L3 speakers). The participants demonstrated a tendency to prefer HA more often after a perception verb, even through in English the overall preference for LA was maintained. In the group analysis, only English monolinguals, who showed a strong preference for LA, were significantly different from the other experimental groups. Finally, social biases did not guide sentence processing in either native or non-native languages.

4. Discussion

Our study investigated RC resolution in non-native English and Russian, determining whether L2 and L3 speakers demonstrate the use of TL-like parsing strategies at the intermediate level of proficiency in their non-native languages. In addition, the study sought experimental evidence on whether non-native languages, L2 and L3, utilize the same mechanisms of sentence parsing.

Our first research question explored the influence of the native or previously learnt languages on RC resolution in English, the L2 or L3 of the participants, and in Russian, an L2. The results of the experiment suggested a negative answer for L1 English influence on attachment preferences. There was no evidence of the L1-like English preference (LA) in L2 Russian; rather, the Russian L2ers appear to be moving towards a HA preference. We would also like to claim that there was relatively little L1 Russian effect on L2/L3 processing in English. Even though the preference for HA in L2/L3 English was around 50% or more, this preference was lower than in the participants' L1 Russian (69% HA), though the difference did not reach statistical significance.. Bearing in mind that the participants were not highly proficient in the L2/L3, we interpret our results as evidence of the development of TL-like processing in non-native English. Ideally, future studies with more advanced non-native speakers of English and Russian should be conducted to determine whether LA is eventully achieved.

Let us assume that the intermediate level of Ln proficiency provides the learner with enough linguistic information to realize that RC resolution, while ambiguous in both languages, shows a tendency to be high in Russian and low in English. In this case, the learner's mental grammar has been set to accommodate two grammatically possible options for RC parsing. If so, the preference for RC resolution of around 50% might be a result of the acquired flexibility in such attachment.

Our second research question investigated whether the presence of a perception verb impacted RC resolution. It examined whether both native and non-native speakers could adjust their structural parse to favor HA when prompted by a perception verb. Our results pointed to an affirmative answer, although this statement requires further clarification.

First, there was a significant main effect of a perception verb on RC resolution and no significant Verb Type – Group interaction. This effect suggests that a perception verb influenced RC resolution favoring HA for the entire population of participants. However, a closer look at the descriptive statistics in Figure 1 reveals differences in Verb Type effect between experimental groups, varying from no effect in NSs of Russian to the 7% difference between HA and LA in NSs of English. A LA-language like English, then, was sensitive to a parsing prompt by a perception verb, whereas Russian preferences remained HA with or without a perception verb.

The effect of a perception verb demonstrated by non-native speakers points to the TL-like sensitivity to this effect in non-native languages. The three groups tested in their non-native English demonstrated the effect of a perception verb, favoring HA up to 11% more often after a perception verb. The group tested in their non-native Russian was less sensitive to the effect of a perception verb.

We view the effect produced by perception verbs as evidence for the participants' sensitivity to a syntactic prompt in native and non-native language. In doing so, we extend the theoretical analysis by Grillo and Costa (2014) to language processing. The study provides experimental evidence for the effect of a main-clause perception verb, suggesting that non-native processing is based on mental structure building in the same way as native processing is.

Our results generally support the assumptions by Grillo and Costa (2014), showing perception verbs affect RC resolution. However, our findings did not fully support the analysis of Grillo et al. (2015), who concluded that RC ambiguity resolution totally depends on the linguistic environment created by a perception verb. Our experimental results did not support the notion that the effect of a perception verb was powerful enough to trigger a full switch to HA in RC resolution in a LA language like English. English monolinguals preserved their overall preference for

LA and increased it in the direction of HA in sentences with a perception verb, but did not reach a HA preference of over 50%.

In addition, the existing literature and our own results demonstrated that Russian is a HA language independently of whether there is a perception or a non-perception verb in the matrix clause (see Figure 1). While a perception verb may provide an additional parsing cue supporting the original preference for HA in Russian, it does not define RC attachment resolution in this language. In some sense, one can think of the perception verb cue applying vacuously in Russian, where its effect is not felt because the attachment preferences are already high.

Our last research question investigated whether native and non-native speakers employ different processing strategies, in particular whether non-native speakers rely on non-structural information in RC resolution more than native speakers do. Evidence in support of shallow processing would involve a different pattern of RC processing between native and non-native speakers. In our experiment, social biases constituted non-structural information that could guide parsing. However, this proved not to be the case in our experiment. Neither native nor non-native processing showed a significant effect of social conventions. This finding should be corroborated in future studies with a higher number of participants and other language combinations.

In sum, our study supports the theoretical assumptions presented in Dekydtspotter et al. (2006) that both native and non-native processing are fundamentally similar, both making use of structural (rather than shallow) parsing. Thus, we suggest that learners, when they acquire a new language, also acquire the processing strategies essential for the target language. Evidence for developing TL-like processing can be attested as early as at the intermediate level of L2/L3 proficiency.

5. Conclusions

This chapter reports on a study of non-native processing, where L2 and L3 speakers, as well as monolinguals, showed sensitivity to the effect of a perception verb in RC resolution. We take this result to constitute evidence of structural processing in native and non-native languages. Similarity of effects in L2 and L3 processing suggests that all non-native processing is fundamentally similar. In terms of RC resolution, L2 and L3 speakers at the intermediate level of proficiency show the beginnings of parsing non-native sentences in a targetlike manner. Further studies with advanced speakers of L2 and L3, as well as studies with a higher number of participants, are needed to investigate the developmental trajectory in processing ambiguous sentences in non-native languages.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Clahsen, H., & Felser, C. (2006). Continuity and shallow structures in language processing. *Applied Psycholinguistics*, 27(2), 107–126. https://doi.org/10.1017/S0142716406060206

Clahsen, H., & Felser, C. (2018). Notes on the Shallow Structure Hypothesis. *Studies in Second Language Acquisition*, 40(3), 693–706. https://doi.org/10.1017/S0272263117000250

Cuetos, F., & Mitchell, D. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition*, 30, 73–105. https://doi.org/10.1016/0010-0277(88)90004-2

Dekydspotter, L., Donaldson, B., Edmonds, A. C., Fultz, A. L., & Petrush, R. A. (2008). Syntactic and prosodic computation in the resolution of relative clause attachment ambiguity by English-French learners. *Studies in Second Language Acquisition*, 30, 453–480. https://doi.org/10.1017/S0272263108080728

Dekydspotter, L., Schwartz, B. D., & Sprouse R. A. (2006). The comparative fallacy in L2 processing research. In M. Grantham O'Brien, C. Shea, & J. Archibald (Eds.), *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006)* (pp. 33–40). Cascadilla Proceedings Project.

Dussias, P. E. (2003). Syntactic ambiguity resolution in L2 learners. *Studies in Second Language Acquisition*, 25(4), 529–557. https://doi.org/10.1017/S0272263103000238

Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish-English bilinguals. *Bilingualism: Language and Cognition*, 10(1), 101–116. https://doi.org/10.1017/S1366728906002847

Felser, C., & Cunnings, I. (2012). Processing reflexives in English as a second language: The role of structural and discourse-level constraints. *Applied Psycholinguistics*, 33, 571–603. https://doi.org/10.1017/S0142716411000488

Felser, C., Roberts, L., Gross, R., & Marinis, T. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics*, 24, 453–489. https://doi.org/10.1017/S0142716403000237

Fernandez, E. (1999). Processing strategies in second language acquisition: Some preliminary results. In E. Klein & G. Martohardjono (Eds.), *The development of second language grammars: A generative approach* (pp. 217–239). John Benjamins. https://doi.org/10.1075/lald.18.12fer

Fodor, J. (2002). Psycholinguistics cannot escape prosody. *Speech Prosody 2002*.

Grillo, N., & Costa, J. (2014). A novel argument for the universality of parsing principles. *Cognition*, 133(1), 156–187. https://doi.org/10.1016/j.cognition.2014.05.019

Grillo, N., Costa, J., Fernandes, B., & Santi, A. (2015). Highs and lows in English attachment. *Cognition*, 144, 116–122. https://doi.org/10.1016/j.cognition.2015.07.018

Hahne, A., & Frederici, A.D. (2001). Processing a second language: Late learners' comprehension mechanisms as revealed by eventrelated brain potentials. *Bilingualism: Language and Cognition*, 4(2), 123–141.

Hemforth, B., Konieczny, L., Scheepers, C., & Strube, G. (1998). Syntactic ambiguity resolution in German. *Syntax and Semantics*, 31, 293–309. https://doi.org/10.1108/S0092-4563(1998)0000031018

Hopp, H. (2014a). Individual differences in the L2 processing of object-subject ambiguities. *Applied Psycholinguistics*, 36(2), 129–173. https://doi.org/10.1017/S0142716413000180

Hopp, H. (2014b). Working memory effects on the L2 processing of ambiguous relative clauses. *Language Acquisition*, 21(3), 250–278. https://doi.org/10.1080/10489223.2014.892943

Lardiere, D. (2009). Some thoughts on the contrastive analyses of features in second language acquisition. *Second Language Research*, 25(2), 173–227. https://doi.org/10.1177/0267658308100283

Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing: A study of relative clause attachment in Greek. *Studies in Second Language Acquisition*, 25, 501–528. https://doi.org/10.1017/S0272263103000214

Rothman, J. (2010). On the typological economy of syntactic transfer: Word order and relative clause high/low attachment preference in L3 Brazilian Portuguese. *IRAL* 48, 245-173

Schwartz, B., & Sprouse, R. (1996). L2 cognitive states and Full Transfer/Full Access Model. *Second Language Acquisition*, 12(1), 40–72. https://doi.org/10.1177/026765839601200103

Sekerina, I. (2002). The Late Closure Principle in processing of ambiguous Russian sentences. *Proceedings of the 2nd Conference on Formal Description of Slavic Languages* (pp. 1–17). Retrieved on 1 March 2022 from: https://www.academia.edu/2173650/_2002_The_Late_Closure_Principle_in_processing_of_ambiguous_Russian_sentences

Slabakova, R. (2017). The scalpel model of third language acquisition. *International Journal of Bilingualism*, 1-15

Slabakova, R., & Montrul, S. A. (2008). Aspectual shifts: Grammatical and pragmatic knowledge in L2 acquisition. In J. M. Liceras, H. Zobl, & H. Goodluck (Eds), *The role of formal features in second language acquisition* (pp. 455–483). Lawrence Erlbaum Associates.

Sokolova, M., & Slabakova, R. (2019). L3-sentence processing: Language-specific or phenomenon-sensitive. *Languages*, 4(5), 1–17. https://doi.org/10.3390/languages4030054

Sprouse, R, A. (2011). The Interface Hypothesis and Full Transfer/Full Access/Full Parse: A brief comparison. *Linguistic Approaches to Bilingualism*, 1, 97–100. https://doi.org/10.1075/lab.1.1.16spr

Witzel, J., Witzel, N., & Nicol, J. (2012). Deeper than shallow: Evidence for structure-based parsing biases in second-language sentence processing. *Applied Psycholinguistics*, 33, 419–456. https://doi.org/10.1017/S0142716411000427

Zagar, D., Pynte, J., & Rativeau, S. (1997). Evidence for early closure attachment on first-pass reading times in French. *Quarterly Journal of Experimental Psychology*, 50A, 421–438. https://doi.org/10.1080/713755715

Appendix. Statistical analysis, full output

Table A. R statistics for the four groups of non-native speakers

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	0.548148	0.021646	53.999998	25.324	< 2e–16 ***
VerbType	0.068519	0.019595	162.000001	3.497	0.000608***
SocialBias	–0.003704	0.019595	162.000001	–0.189	0.850317
Group_factor1	0.062538	0.049856	53.999998	1.254	0.215106
Group_factor2	–0.057543	0.052092	53.999998	–1.105	0.274207
Group_factor3	0.064610	0.064088	53.999998	1.008	0.317878
VerbType:SocialBias	–0.062963	0.039189	162.000001	–1.607	0.110082
VerbType:Group_factor1	0.053820	0.045132	162.000001	1.193	0.234802

Fixed effects:					
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**
VerbType:Group_factor2	−0.041407	0.047156	162.000001	−0.878	0.381197
VerbType:Group_factor3	0.012338	0.058016	162.000001	0.213	0.831859
SocialBias:Group_factor1	−0.013896	0.045132	162.000001	−0.308	0.758553
SocialBias:Group_factor2	0.019351	0.047156	162.000001	0.410	0.682088
SocialBias:Group_factor3	0.044156	0.058016	162.000001	0.761	0.447702
VerbType:SocialBias:Group1	0.046104	0.090263	162.000001	0.511	0.610207
VerbType:SocialBias:Group2	−0.093506	0.094312	162.000001	−0.991	0.322939
VerbType:SocialBias Group3	0.158442	0.116031	162.000001	1.366	0.173986

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table B. R statistics for all groups: Native and non-native speakers

Fixed effects:					
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**
(Intercept)	0.529795	0.020686	72.999995	25.611	< 2e−16 ***
VerbType	0.062329	0.016344	219.000002	3.814	0.000178 ***
SocialBias	−0.004795	0.016344	219.000002	−0.293	0.769526
Group_factor1	0.256467	0.062278	72.999995	4.118	9.95e−05 ***
Group_factor2	0.077220	0.054942	72.999995	1.405	0.164116
Group_factor3	−0.028178	0.056116	72.999995	−0.502	0.617077
Group_factor4	0.123341	0.065311	72.999995	1.889	0.062929.
Group_factor5	0.117460	0.075514	72.999995	1.555	0.124157
VerbType: SocialBias	−0.058904	0.032687	219.000002	−1.802	0.072912.
VerbType:Group_factor 1	−0.019717	0.049203	219.000002	−0.401	0.689016
VerbType:Group_factor 2	0.046281	0.043407	219.000002	1.066	0.287510
VerbType:Group_factor 3	−0.056486	0.044335	219.000002	−1.274	0.203984
VerbType:Group_factor 4	−0.017821	0.051599	219.000002	−0.345	0.730145
VerbType:Group_factor 5	−0.060317	0.059661	219.000002	−1.011	0.313127
SocialBias:Group_factor1	0.042367	0.049203	219.000002	0.861	0.390143
SocialBias:Group_factor2	−0.012408	0.043407	219.000002	−0.286	0.775263
SocialBias:Group_factor3	0.022327	0.044335	219.000002	0.504	0.615051
SocialBias:Group_factor4	0.050108	0.051599	219.000002	0.971	0.332570
SocialBias:Group_factor5	0.011905	0.059661	219.000002	0.200	0.842024
VerbType:SocialBias:Group1	−0.150023	0.098407	219.000002	−1.525	0.128821
VerbType:SocialBias:Group2	0.025667	0.086815	219.000002	0.296	0.767772
VerbType:SocialBias:Group3	−0.134380	0.088670	219.000002	−1.516	0.131086
VerbType:SocialBias:Group4	0.076696	0.103199	219.000002	0.743	0.458168
VerbType:SocialBias:Group5	−0.163492	0.119321	219.000002	−1.370	0.172032

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

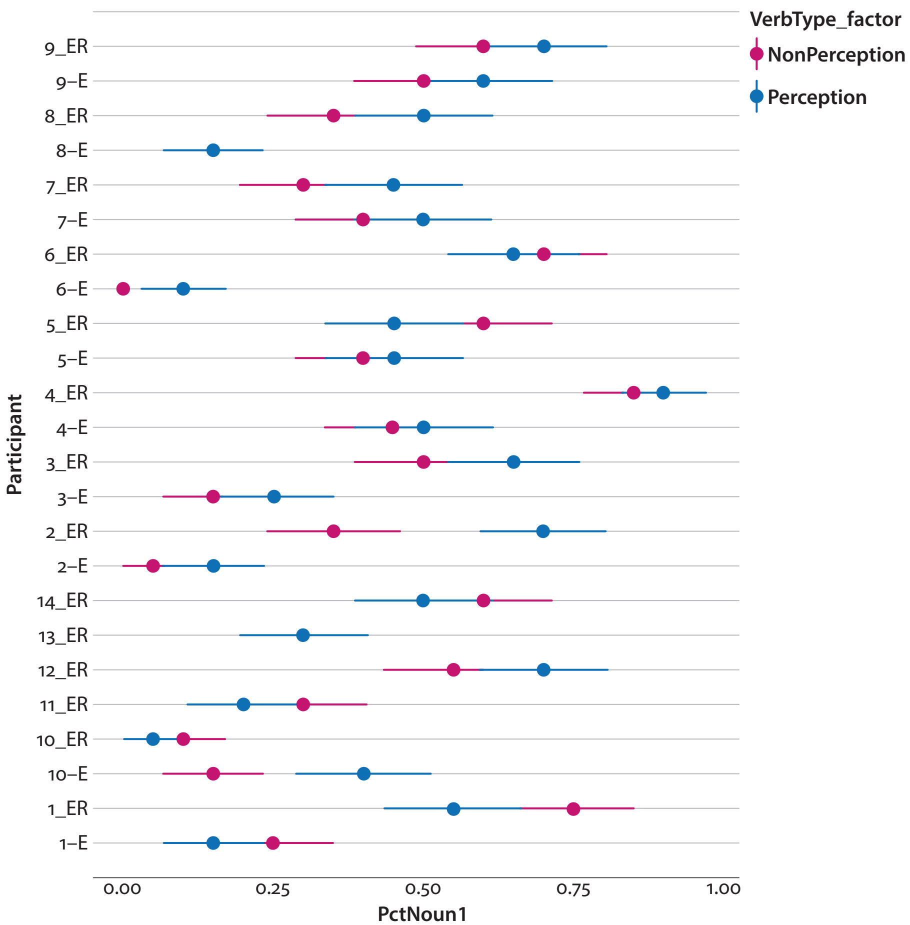


Figure A. Percentage Noun choice with HA as a Reference Category: English monolinguals and English-Russian L2ers

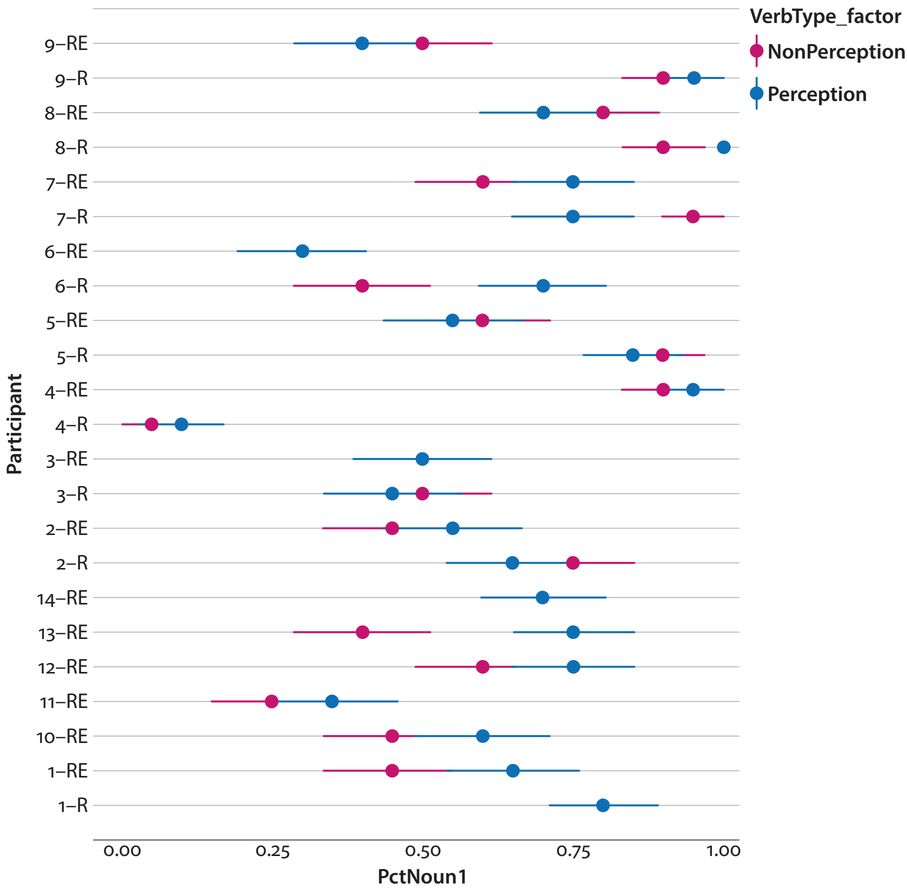


Figure B. Percentage Noun choice with HA as a Reference category: Russian monolinguals and Russian-English L2ers

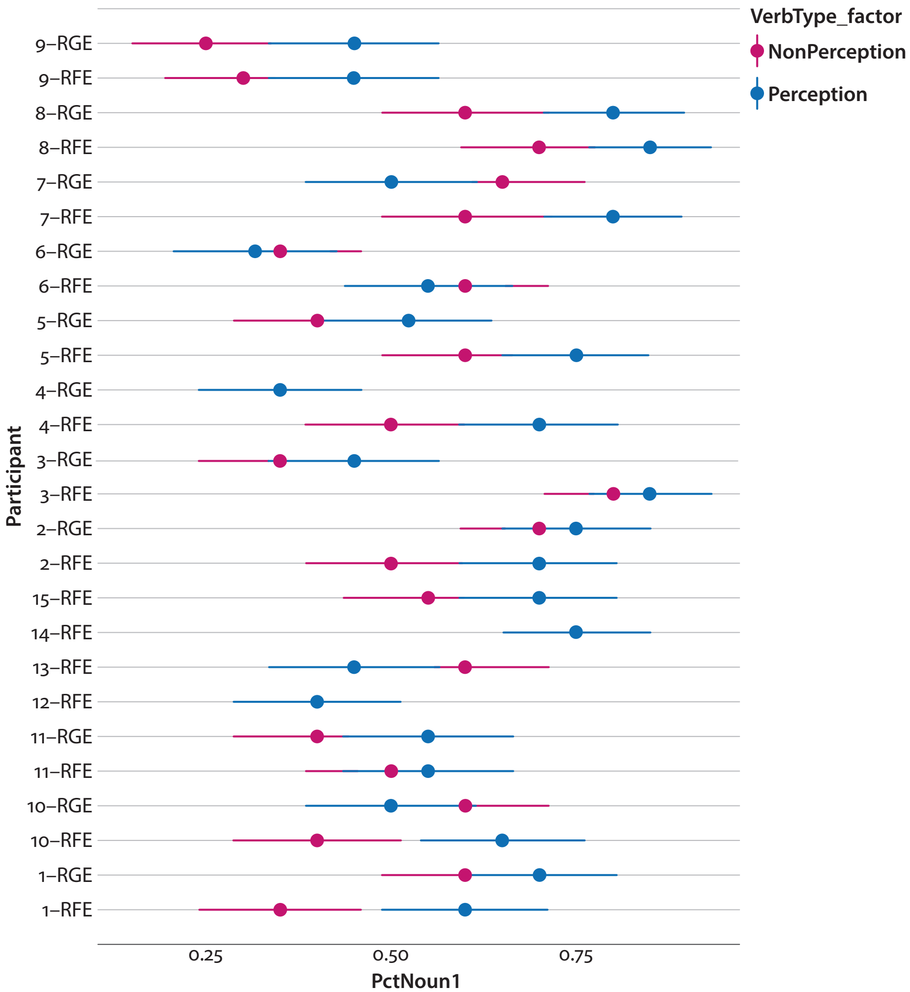


Figure C. Percentage Noun choice with HA as a Reference Category: L3 speakers of English

Subject index

This volume brings together empirical studies and keynote addresses presented at the 15th Generative Approaches to Second Language Acquisition conference hosted by the University of Nevada, Reno in 2019. The studies selected for the volume reflect how the latest developments in generative syntactic theory and psycholinguistic methodologies have impacted second language acquisition research in the last decade, from the linguistic properties under investigation and L1-L2/Ln language pairings down to the specific research questions in each study. The minimalist view of language architecture is at the center of studies investigating L2 acquisition of raising, scope, definiteness, phonological representations, and interlanguage transfer. The volume also showcases the latest research on interface phenomena, language processing, and working memory. Studies analyze data collected with a variety of L2 populations from adult foreign language learners to adolescent L3 learners and heritage speakers.

ISBN 978 90 272 1144 6

9 789027 211446

JOHN BENJAMINS PUBLISHING COMPANY