DE GRUYTER
MOUTON

# ISSUES IN JAPANESE PSYCHOLINGUISTICS FROM COMPARATIVE PERSPECTIVES

## VOLUME 1: CROSS-LINGUISTIC STUDIES

*Edited by Masatoshi Koizumi*

NINJAL
National Institute for Japanese Language and Linguistics

THE MOUTON-NINJAL LIBRARY OF LINGUISTICS

DE G

**Issues in Japanese Psycholinguistics from Comparative Perspectives**

# The Mouton-NINJAL Library of Linguistics

Edited by
Yukinori Takubo
Haruo Kubozono
Yo Matsumoto

## Volume 5

# Issues in Japanese Psycholinguistics from Comparative Perspectives

Volume 1: Cross-Linguistic Studies

Edited by
Masatoshi Koizumi

DE GRUYTER
MOUTON

ISBN 978-3-11-077880-9
e-ISBN (PDF) 978-3-11-077894-6
e-ISBN (EPUB) 978-3-11-077906-6
ISSN 2626-9201
e-ISSN 2626-921X
DOI https://doi.org/10.1515/9783110778946



**Library of Congress Control Number: 2023934194**

**Bibliographic information published by the Deutsche Nationalbibliothek**
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the internet at http://dnb.dnb.de.



Cover image: Ben1183/iStock/Getty Images Plus
Typesetting: Integra Software Services Pvt. Ltd.
Printing and binding: CPI books GmbH, Leck

www.degruyter.com

In memory of Hiroko Hagiwara

# Series preface

*The Mouton-NINJAL Library of Linguistics* (MNLL) series is a new collaboration between De Gruyter Mouton and NINJAL (National Institute for Japanese Language and Linguistics), following the successful twelve-volume series *Mouton Handbooks of Japanese Language and Linguistics*. This new series publishes research monographs as well as edited volumes from symposia organized by scholars affiliated with NINJAL. Every symposium is organized around a pressing issue in linguistics. Each volume presents cutting-edge perspectives on topics of central interest in the field. This is the first series of scholarly monographs to publish in English on Japanese and Ryukyuan linguistics and related fields.

NINJAL was first established in 1948 as a comprehensive research organization for Japanese. After a period as an independent administrative agency, it was re-established in 2010 as the sixth organization of the Inter-University Research Institute Corporation "National Institutes for the Humanities". As an international hub for research on Japanese language, linguistics, and Japanese language education, NINJAL aims to illuminate all aspects of the Japanese and Ryukyuan languages by conducting large-scale collaborative research projects with scholars in Japan and abroad. Moreover, NINJAL also aims to make the outcome of the collaborative research widely accessible to scholars around the world. The MNLL series has been launched to achieve this second goal.

The authors and editors of the volumes in the series are not limited to the scholars who work at NINJAL but include invited professors and other scholars involved in the collaborative research projects. Their common goal is to disseminate their research results widely to scholars around the world.

The current volume originated from an international conference jointly held by Tohoku University and NINJAL and collects papers on psycholinguistics related to the Japanese language from comparative perspectives. Aiming to bridge the gap in the field between theoretical and psycholinguistic studies, it covers L1 and L2 acquisition as well as language comprehension and production. It will benefit both students and experts alike by providing information needed to carry out their research as well as information concerning what is happening in the state of the art in their subfields.

Yukinori Takubo
Haruo Kubozono


https://doi.org/10.1515/9783110778946-202

# Preface

*Issues in Japanese Psycholinguistics from Comparative Perspectives* came out of the International Symposium on Issues in Japanese Psycholinguistics from Comparative Perspectives (IJPCP) held online in September 2021. IJPCP consisted of twenty-nine papers in ten sessions over two days. It was jointly organized by the JSPS Grant-in-Aid for Scientific Research (S) Project "Field-Based Cognitive Neuroscientific Study of Word Order in Language and Order of Thinking from the OS Language Perspective" and the NINJAL (National Institute for Japanese Language and Linguistics) Collaborative Research Project "Cross-linguistic Studies of Japanese Prosody and Grammar" and cosponsored by the Advanced Institute of Yotta Informatics (AI Yotta), Tohoku University, Japan.

*Issues in Japanese Psycholinguistics from Comparative Perspectives* is in two volumes: *Cross-Linguistic Studies* (Volume 1) and *Interaction Between Linguistic and Nonlinguistic Factors* (Volume 2). The two volumes combined together include 27 papers that were all presented at the conference except for two papers by Takuya Kubo and Jungho Kim, respectively, who were unable to attend the symposium. All the papers went through peer review, and I would like to thank those who kindly acted as inside or outside reviewers.

In organizing the international symposium and editing the volumes, I received invaluable assistance from numerous people. First and foremost, I am grateful to Yukinori Takubo (Director-General of NINJAL) and Haruo Kubozono (former Deputy Director-General of NINJAL) for their continuous support that made this project possible. Sachiko Kiyama, Kexin Xiong, Maho Morimoto, Misato Ido, Min Wang, Ge Song, Liya Cheng, and Rei Emura helped organize the conference. Thanks are also due to Michaela Göbels and De Gruyter Mouton for their support. The conference and the editing of the volumes were funded by NINJAL and JSPS KAKENHI Grant Number 19H05589.

Masatoshi Koizumi


https://doi.org/10.1515/9783110778946-203

# Contents

# Contributors

**Kensuke Emura**
Center for the Advancement of Higher Education, Iwate Prefectural University
152-52 Sugo, Takizawa, Iwate, 020-0693, JAPAN
emura_k@iwate-pu.ac.jp

**Rei Emura**
Graduate School of Arts and Letters, Tohoku University
27-1 Kawauchi, Aoba-ku, Sendai 980-8576, JAPAN
rei.emura.r4@dc.tohoku.ac.jp

**Akinori Ito**
Graduate School of Engineering, Tohoku University
6-6-5 Aramaki aza Aoba, Aoba-ku, Sendai, 980-8579, JAPAN
aito@spcom.ecei.tohoku.ac.jp

**Jungho Kim**
Faculty of Letters, Kyoto Women's University
35 Kitahiyoshi-cho, Imakumano, Higashiyama-ku, Kyoto, 605-8501, JAPAN
kimj@kyoto-wu.ac.jp

**Naoki Kimura**
Department of Business Management, Tokyo Fuji University
3-8-1 Takadanobaba, Shinjuku-ku, Tokyo, 169-0075, JAPAN
nkimura@fuji.ac.jp

**Sachiko Kiyama**
Department of Linguistics, Tohoku University
27-1 Kawauchi, Aoba-ku, Sendai 980-8576, JAPAN
skiyama@tohoku.ac.jp

**Masatoshi Koizumi**
Graduate School of Arts and Letters, Tohoku University
27-1 Kawauchi, Aoba-ku, Sendai, 980-8576, JAPAN
koizumi@tohoku.ac.jp

**Takuya Kubo**
Administrative Headquarters, Shinshu University
3-1-1 Asahi, Matsumoto, Nagano, 390-8621, JAPAN
kubotaku@shinshu-u.ac.jp

**Mari Kugemoto**
College of Humanities & Fine Arts, University of Massachusetts Amherst
N408 Integrative Learning Center, 650 North Pleasant Street, Amherst, MA, 01003, U.S.A.
mkugemoto@umass.edu

**Ichiro Kuriki**
Graduate School of Science and Engineering, Saitama University
255 Shimo-Okubo, Sakura-ku, Saitama, 338-8570, JAPAN
ikuriki@mail.saitama-u.ac.jp

**Cornelia D. Lupsa**
Center for the Advancement of Higher Education, Iwate Prefectural University
152-52 Sugo, Takizawa, 020-0693, JAPAN
c-lupsa@iwate-pu.ac.jp

**Satoru Mizuochi**
Graduate School of Engineering, Tohoku University
6-6-5 Aramaki aza Aoba, Aoba-ku, Sendai, 980-8579, JAPAN
satoru.mizuochi.p3@spcom.ecei.tohoku.ac.jp

**Shota Momma**
College of Humanities & Fine Arts, University of Massachusetts Amherst
N408 Integrative Learning Center, 650 North Pleasant Street, Amherst, MA, 01003, U.S.A.
snegishi@umass.edu


https://doi.org/10.1515/9783110778946-205

**Toshiaki Muramoto**
International Research Institute of Disaster Science, Tohoku University
468-1 Aoba, Aramaki, Aoba-ku, Sendai 980-8572, JAPAN
muramoto@cog.is.tohoku.ac.jp

**Kuniya Nasukawa**
Department of English, Tohoku Gakuin University
1-3-1 Tsuchitoi, Aoba-ku, Sendai, 980-8511, JAPAN
nasukawa@mail.tohoku-gakuin.ac.jp

**Keiyu Niikuni**
Department of Clinical Psychology, Niigata Seiryo University
1-5939 Suido-cho, Chuo-ku, Niigata 951-8121, JAPAN
keiyu@n-seiryo.ac.jp

**Takashi Nose**
Graduate School of Engineering, Tohoku University
6-6-5 Aramaki aza Aoba, Aoba-ku, Sendai, 980-8579, JAPAN
nose@tohoku.ac.jp

**Hajime Ono**
College of Liberal Arts, Tsuda University
2-1-1 Tsuda-machi, Kodaira, Tokyo, 187-8577, JAPAN
hajime@tsuda.ac.jp

**Yohei Oseki**
Graduate School of Arts and Sciences, University of Tokyo
3-8-1 Komaba, Meguro-ku, Tokyo, 153-8901, JAPAN
oseki@g.ecc.u-tokyo.ac.jp

**Koichi Otaki**
School of Global Studies, Chukyo University
5-31-2 Yamate-dori, Showa-ku, Nagoya, Aichi, 466-0815, JAPAN
otaki@lets.chukyo-u.ac.jp

**Yuko Otsuka**
Faculty of Foreign Studies, Sophia University
7-1 Kioi-cho, Chiyoda-ku, Tokyo, 102-8554, JAPAN
yotsuka@sophia.ac.jp

**Hiromu Sakai**
Faculty of Science and Engineering, Waseda University
3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, JAPAN
hsakai@waseda.jp

**Manami Sato**
College of Global and Regional Culture, Okinawa International University
2-6-1 Ginowan, Ginowan City, Okinawa, 901-2701, JAPAN
m.sato@okiu.ac.jp

**Satoshi Shioiri**
Research Institute of Electrical Engineering, Tohoku University
2-1-1 Katahira, Aoba-ku, Sendai, 980-8577, JAPAN
Satoshi.shioiri.b5@tohoku.ac.jp

**Koji Sugisaki**
School of Humanities, Kwansei Gakuin University
1-155 Uegahara Ichiban-cho, Nishinomiya, Hyogo, 662-8501, JAPAN
sugisaki@kwansei.ac.jp

**Rumi Tokunaga**
Graduate school of Global and Transdisciplinary Studies, Chiba University
1-33 Yayoicho, Inage-Ku, Chiba, 263-8522, JAPAN
tokunaga@chiba-u.jp

**Kexin Xiong**
Frontier Research Institute for Interdisciplinary Sciences, Tohoku University
6-3 Aramaki aza Aoba, Aoba-ku, Sendai 980-8578, JAPAN
xiong@tohoku.ac.jp

**Matthew Wagers**
Department of Linguistics, University of California, Santa Cruz
1156 High St., Santa Cruz, CA, 95064-1077, U.S.A.
mwagers@ucsc.edu

**Hiroko Yamashita**
Department of Modern Languages and Cultures, Rochester Institute of Technology
92 Lomb Memorial Drive, Rochester, NY, 14623, U.S.A.
hxygsl@rit.edu

**Kazuko Yatsushiro**
Leibniz-Center General Linguistics (ZAS)
Schützenstr. 18, 10117 Berlin, GERMANY
yatsushiro@leibniz-zas.de

**Noriaki Yusa**
Department of English, Miyagi Gakuin Women's University
9-1-1 Sakuragaoka, Aoba-ku, Sendai, 981-8557, JAPAN
yusa@mgu.ac.jp

Masatoshi Koizumi

# Chapter 1 Japanese Psycholinguistics from Comparative Perspectives: Cross-linguistic studies

## 1 Introduction

*Issues in Japanese Psycholinguistics from Comparative Perspectives* consists of two volumes compiling 27 state-of-the-art articles on Japanese psycholinguistics and related topics. It emphasizes the importance of using comparative perspectives when conducting psycholinguistic research.

Psycholinguistic studies of the Japanese language have contributed greatly to the field from a cross-linguistic perspective. However, the target languages for comparison have been limited. Most research focuses on English and a few other typologically similar languages, which are nominative-accusative and subject-before-object (SO) languages, as is Japanese. As a result, many current theories fail to acknowledge the nature of ergative-absolutive and/or object-before-subject (OS) languages and treat the nature of nominative-accusative subject-before-object languages as universal to human language. A detailed consideration of the language processing stages of more diverse languages (in addition to familiar languages), in comparison with Japanese, is essential to clarify the universality and individuality of human language and to correctly situate Japanese among human languages.

The cross-linguistic approach is not the only method of comparison in psycholinguistics. Other prominent comparative aspects include comprehension vs. production, prosodic vs. syntactic processing, syntactic vs. semantic processing, semantic vs. pragmatic processing, native speakers vs. second language learners, typical development vs. development of language by people with autism spectrum disorder, typical vs. aphasic language development, language vs. action, and language vs. memory. Comparative studies have proved fruitful in revealing the nature of various components of human cognition as well as how they interact with each other. Many of these approaches are underrepresented in Japanese psycholinguistics.

The studies reported in the two volumes attempt to fill these gaps. Using various experimental and/or computational techniques, they address issues of the universality/diversity of the human language and the nature of the relationship between human cognitive modules. Special reference is made to the mechanisms in which languages are processed and represented in the mind and brain.

**Acknowledgments:** Part of this work is supported by JSPS KAKENHI Grant Number 19H05589.


https://doi.org/10.1515/9783110778946-001

## 2 Outline of Volume 1

In addition to this chapter, Volume 1 contains 14 papers related to cross-linguistic research, each summarized in the following paragraphs.

Chapter 2, "Dimensions in the investigations of human language processing," is an excellent survey of research on language processing related to Japanese. In this chapter, Hiroko Yamashita discusses the importance of studies on human language processing (HLP) and production in less-commonly investigated (LCI) languages – which typically have linguistic properties different from English. They have significantly contributed to the understanding of how humans use the knowledge of the language. The chapter highlights how the studies on Japanese have impacted the advancement of HLP. It then presents an argument that additional diverse approaches such as in-depth investigations of language-specific phenomena, the creation of comprehensive theories for each LCI language, and research of speakers with atypical cognitive mechanisms will further advance the understanding of the mechanisms of HLP in general.

The subsequent six chapters explore adult LCI language processing (both comprehension and production) from a comparative perspective. Chapter 3 "Encoding interference in verb-initial languages" by Matthew Wagers investigates the nature of encoding interference in verb-initial languages. Encoding interference provides a satisfying explanation of why initially-disfavored object-extracted relative clause (ORC) structures are sensitive to the similarity of their arguments. It argues that encoding interference can occur whenever there is reanalysis and not merely when two arguments occur in linear succession. These two sources are confounded in the noun–noun–verb (N–N–V) word orders widely examined in verb-medial/verb-final languages. However, in languages with verb-initial main clauses, relative clauses often have noun–verb–noun (N–V–N) orders. Focusing on two such languages, Chamorro (Austronesian, Mariana Islands) and Zapotec (Oto-Manguean, Oaxaca, Mexico), the author demonstrates that similarity affects the ORC parse even when the arguments involved are never processed in close succession.

In Chapter 4, "Cross-cultural comparison of lexical partitioning of color space," Satoshi Shioiri, Rumi Tokunaga, and Ichiro Kuriki elaborate on a cross-cultural comparison of lexical partitioning of color space. As may be expected from variations of cultures, color terms are not determined independently for different languages. It is known that there is a common set of processes for color lexicons across different cultures; more precisely, these processes are likely related to physiological factors, influencing color naming systems in most languages. To investigate the effect of cultural differences in more detail, the authors introduce new techniques concerning data analyses, i.e., *k*-means clustering and motif analyses. These methods indicate the similarities and differences in color lexicons among Japanese, Taiwanese Chinese, and standard American English.

Chapter 5, "Word orders, gestures, and a view of the world from OS languages," by Hajime Ono, Takuya Kubo, Manami Sato, Hiromu Sakai, and Masatoshi Koizumi considers the relationship between word orders, gestures, and a view of the world from object-before-subject (OS) languages. Based on a gesture production study, Goldin-Meadow et al. (2008) argued that the subject-object-verb (SOV) (or actor-patient-action) order is a natural order of an event description. Because all the languages examined in previous studies have subject-before-object (SO) word order as their basic word order, Ono and his colleagues examined Kaqchikel (Mayan, Guatemala), whose basic word order is VOS. In Experiment 1, in contrast to previous studies, Kaqchikel speakers produced as many subject-verb-object (SVO) gestures as SOV gestures. In Experiment 2, the participants had a practice phase before the task and produced considerably more SVO gestures, but crucially not in the object-human condition, in which S is an inanimate object and O is a human. This supports the perspectives of Hall, Mayberry, and Ferreira (2013) and Hall, Ferreira, and Mayberry (2014) regarding the increase in SVO gestures in events with two human entities. As such, the authors of this chapter suggest that patient-boundedness plays a more decisive role than the complexity of action in determining the choice between SVO and SOV gesture orders.

In Chapter 6, "Factors affecting the choice of word order in Kaqchikel: Evidence from discourse saliency," Takuya Kubo examines the factors affecting the choice of word order in Kaqchikel. From the perspective of discourse analysis, Gundel (1988) proposed two independent principles that determine the choice of word order: the *given-before-new* principle and the *first-things-first* principle. However, psycholinguistic studies of sentence production to date have established only a tendency to follow the former and not the latter. Using a picture description task in which discourse contexts were manipulated, Kubo explored how these principles affect the choice of word order in Kaqchikel. In particular, Kaqchikel speakers tended to produce verb-object-subject (VOS) active sentences more often when the agent was contextually salient, implying that the *first-things-first* principle played a greater role than the *given-before-new* principle. Moreover, the author discusses the interaction between discourse principles and psycholinguistic principles based on this result.

In Chapter 7, "Sentence comprehension in Central Alaskan Yup'ik: The effects of case marking, agreement, and word order," author Rei Emura tackles sentence comprehension in Central Alaskan Yup'ik (Eskimo–Aleut, southeast Alaska), an ergative language with free word order. Emura examines the effects of word order and their interaction with case marking and verb agreement on the judgment of grammatical relations in this language. The acceptability judgment experiment presents the preference for subject-object order regardless of ambiguity. The sequence in which objects are immediately followed by verbs was less acceptable only in the case-ambiguous sentences. Furthermore, agreement ambiguity has no effects, in contrast to case marking. These results indicate that Yup'ik speakers use the word order cue and the case marking cue, but not the agreement cue to determine grammatical relations. Finally, the obtained findings regarding Yup'ik are compared with those of studies involving Japanese.

In Chapter 8 "Producing long-distance dependencies in English and Japanese," using structural priming, Mari Kugemoto and Shota Momma investigate how English and Japanese speakers plan long-distance *wh*-dependencies in sentence production. Specifically, Experiment 1 in English revealed that priming the optional complementizer *that* had a slow-down effect on the onset latency of subject-extracted *wh*-questions, where *that* cannot be used grammatically, but not on the onset latency of object-extracted *wh*-questions. In Experiment 2 in Japanese, the embedded *wh*-scope and matrix *wh*-scope had a speed-up and slow-down effect on the onset latency of the matching scope targets, respectively. According to the authors, these results imply early planning of the structural properties of *wh*-dependencies before uttering the sentence-initial *wh*-phrase in both English and Japanese.

The subsequent five chapters approach issues in first and/or second language acquisition from a comparative perspective. In Chapter 9, "Case and word order in children's comprehension of *wh*-questions: A cross-linguistic study," Koichi Otaki, Manami Sato, Hajime Ono, Koji Sugisaki, Noriaki Yusa, Yuko Otsuka, and Masatoshi Koizumi consider case and word order in children's comprehension of *wh*-questions from a cross-linguistic perspective. Building on their own experimental data of typologically distinct languages, the authors identify the source of the subject preference widely observed in children's comprehension of *wh*-questions. Previous studies on the acquisition of *wh*-questions have focused exclusively on typologically similar languages with a nominative/accusative case system and SO word order. The exclusive research makes it difficult to consider the role of case and word order in children's subject-over-object preferences. The authors tested major hypotheses regarding children's subject preferences against the experimental data obtained from children acquiring Japanese, Tongan (Austronesian, Tonga), and Kaqchikel. They demonstrate the structural distance between a moved *wh*-phrase and its gap strongly affecting children's comprehension of *wh*-questions, thereby arguing for the structural distance hypothesis.

In Chapter 10, "Cross-linguistic investigation of the acquisition of disjunction," Kazuko Yatsushiro compares children's interpretation of disjunction in Japanese (*ka* and *ka. . .ka*) and German (*oder* and *entweder. . .oder*) in an upward entailing environment. Yatsushiro observes some differences between the following two groups of children: four and five-year-olds (Japanese and German) and six-year-olds and older (German). She speculates that the difference between the two languages may stem from morphological differences between the disjunction of these two languages.

Chapter 11 "Effects of annual quantity of second language input on pronunciation in EFL environments" by Noriaki Yusa, Cornelia D. Lupsa, Naoki Kimura, Kensuke Emura, Jungho Kim, Kuniya Nasukawa, Masatoshi Koizumi, and Hiroko Hagiwara investigates whether the *earlier-is-better* rule of thumb generally observed in the acquisition of pronunciation in the second language (L2) environments applies to English as a foreign language (EFL) environments. To identify the effects of early exposure to English on the production of English stop consonants, they followed their participants' changes in voice onset time (VOT) for four years. The results indicate that not only the total quantity of L2 input but also the annual quantity affects the production of L2 VOT. Accordingly, this implies the importance of L2 input that is both continuous and consistent in quantity in EFL environments.

While cross-language word-form overlap facilitates cognate recognition, little is known about the effect of sub-lexical information. In Chapter 12 "Asymmetric effects of sub-lexical orthographic/phonological similarities on L1-Chinese and L2-Japanese visual word recognition," Kexin Xiong, Keiyu Niikuni, Toshiaki Muramoto, and Sachiko Kiyama discuss using eye-tracking to investigate the sub-lexical (character) form overlapping effects on Chinese (first language, L1)-Japanese (L2) bilinguals' cognate reading ability. The results of their experiments indicate that the sub-lexical form overlap affects cognate reading differently in these languages. When reading in L2, the greater orthographic similarity of the initial character induced longer first fixation duration, whereas more orthographic and phonological overlaps required shorter reaction times (RTs). However, the first fixation duration of L1 reading decreased with phonological similarity and RTs declined with orthographic overlaps of the initial character. The authors argue that logographic cognate reading is driven by sub-lexical form overlaps and, moreover, that the phonological information activates even when it is not necessary for visual word recognition.

In Chapter 13 "Cortical neural activities related to processing Japanese scrambled sentences by Japanese L2 learners: An fMRI study," Jungho Kim reports on cortical neural activities related to processing scrambled Japanese sentences by Japanese L2 learners using functional magnetic resonance imaging. Japanese is a well-known free word order language; thus, it is assumed that when an object is scrambled to a position that precedes a subject, it leaves a "trace" in its original position and creates "a filler-gap dependency." The author clarified the mechanism underlying sentence processing during the parsing of Japanese scrambled sentences by native speakers of Korean and Chinese. The direct comparison of data between scrambled and canonical tasks in Korean and Chinese participants exhibited cortical activation in the left interior frontal gyrus (LIFG) in Broca's area. This result indicates that despite typological differences between the languages (e.g., SOV vs. SVO), Broca's area is a syntactically modulated region both in L1 and L2. In addition, this result implies that the LIFG is activated when the syntactic structure of a presented sentence is more complex.

The last two chapters combine comparative psycholinguistics and natural language processing. In Chapter 14 "Spoken term detection from utterances of minority languages," Akinori Ito, Satoru Mizuochi, and Takashi Nose present a new method of spoken term detection from utterances of minority languages. Many attempts have been made to create databases of the speech of the endangered languages. To realize a function to search the speech database without a speech recognizer, the authors propose the query-by-example spoken term detection (QbE-STD), which searches a speech database using speech as a query. They examined this to combine Japanese and English posteriorgrams (vectors of phoneme posterior probabilities) to search the speech database of another language (Kaqchikel). Drawing on the experimental results, the authors improved the search performance for the Kaqchikel language using the proposed method compared with the posteriorgram from a single language and the conventional acoustic feature.

Finally, in Chapter 15 "Human language processing in comparative computational psycholinguistics," Yohei Oseki discusses HLP in comparative computational psycholinguistics. The author advocates a comparative approach to computational psycholinguistics, which he terms as comparative computational psycholinguistics. The comparative computational psycholinguistics constructs and evaluates computational models of HLP from comparative perspectives. Specifically, he presents the results of modeling hierarchical syntactic structure with recurrent neural network grammars (Dyer et al. 2016). This demonstrates that the hierarchical syntactic structure universally makes computational models more human-like, although optimal parsing strategies may vary with respect to head directionality (Yoshida et al. 2021). He then presents the results of modeling cue-based memory retrieval with Transformer architectures (Vaswani et al. 2017; Merkx and Frank 2021), suggesting that these are too powerful for languages with few long-distance dependencies, which can be rendered more human-like through context limitations (Kuribayashi et al. 2021, 2022).

## References

Dyer, Chris, Adhiguna Kuncoro, Miguel Ballesteros, & Nsoah Smith. 2016. Recurrent Neural Network Grammars. In Kevin Knight, Ani Nenkova, Owen Rambow (eds.), *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 199–209.

Goldin-Meadow, Susan, Wing Chee So, Aslı Özyürek, & Carolyn Mylander. 2008. The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of* *the National Academy of Sciences of the United States of America* 105(27). 9163–9168. https://doi.org/10.1073/pnas.0710060105

Gundel, Jeanette K. 1988. Universals of topic-comment structure. *Studies in Syntactic Typology* 17(1). 209–239. https://doi.org/10.1075/tsl.17.16gun

Hall, Matthew L., Victor S. Ferreira, & Rachel I. Mayberry. 2014. Investigating constituent order change with elicited pantomime: A functional account of SVO emergence. *Cognitive Science* 38(5). 943–972. https://doi.org/10.1111/cogs.12105

Hall, Matthew L., Rachel I. Mayberry, & Victor S. Ferreira. 2013. Cognitive constraints on constituent order: Evidence from elicited pantomime. *Cognition* 129(1). 1–17. https://doi.org/10.1016/j.cognition.2013.05.004

Kuribayashi, Tatsuki, Yohei Oseki, Ana Brassard, & Kentaro Inui. 2022. Context limitations make neural language models more human-like. https://arxiv.org/abs/2205.11463

Kuribayashi, Tatsuki, Yohei Oseki, Takumi Ito, Ryo Yoshida, Masayuki Asahara, & Kentaro Inui. 2021. Lower perplexity is not always human-like. *Proceedings of the 59th Annual Meeting of the Associatiosn for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, 5203–5217.

Merkx, Danny & Stefan Frank. 2021. Human sentence processing: Recurrence or attention? *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, 12–22.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan Gomez, Łukasz Kaiser, & Illia Polosukhin. 2017. Attention Is all you need. *Proceedings of NIPS*, 5998–6008.

Yoshida, Ryo, Hiroshi Noji, & Yohei Oseki. 2021. Modeling human sentence processing with left-corner recurrent neural network grammars. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 2964–2973.

Hiroko Yamashita

# Chapter 2
# Dimensions in the investigations of human language processing

## 1 Introduction

Research in human language processing (HLP), the cognitive processes that take place as humans use their knowledge of language, has witnessed a remarkable development over the last half century (see Sanz, Laka, and Tanenhaus 2015 for summary). Earlier studies were based in English, a head-initial language with relatively rigid word order and obligatory overt pronouns. As the field further developed, studies in languages with linguistic characteristics distinct from English emerged. Although English is still the most studied language, these less commonly investigated (LCI) languages have significantly contributed to the understanding of HLP.[1]

This chapter highlights the role LCI languages have played in deepening the understanding of HLP and further discusses some potential impacts if more dimensions in the focus in the study of HLP are added. It will focus on the case of Japanese, one of the LCI languages due to its distinct linguistic characteristics in contrast to English.

The goals of the chapter are trifold. First, the advancement of the field of HLP in general made by studies of Japanese, a head-final language with flexible word order and null pronouns, is highlighted. It shows how different types of language may contribute to understanding the mechanism of HLP in general.[2] Then, two additional dimensions to the study of HLP are proposed: in-depth investigations of language-specific phenomena to establish a processing model of each individual language, and studies of HLP in language users with atypical cognitive mechanisms

1 In this chapter, language processing is broadly defined and includes both human sentence processing and production.

2 In this chapter, the term "null pronoun" refers to cases where arguments are not phonologically realized in Japanese. Note, however, some have demonstrated that the phenomenon in Japanese differs in its function from null pronouns in Spanish and may be better captured as "argument ellipsis" (e.g., Oku 1998; Otaki 2014; Saito 2004; Sakamoto 2016; among others).

**Acknowledgments:** I would like to thank Joseph Bochner, Franklin Chang, and anonymous reviewers for their valuable comments on the chapter. All shortcomings are my own.


https://doi.org/10.1515/9783110778946-002

in each language. It is suggested that these dimensions will not only benefit studies in each language but also further advance the understanding of the mechanisms of HLP in general.

## 2 Exploring the overarching principles of HLP in all languages

Most studies of HLP in LCI languages reported at international venues such as journals and conferences have traditionally been investigations on overarching principles of HLP common to all languages. The data and the studies' theoretical interpretations of LCI languages tended to center around the challenges found in, or the corroborations of, hypotheses and theories of HLP and production mainly developed through the study of English. Often those reported challenges significantly advanced the theories.

One such example is a seminal paper by Mazuka and Lust (1987), which questioned the hypothesis of sentence processing strategies driven by the grammatical head, that is, a verb, preposition, noun, among others (e.g., Kimball 1973; Pritchett 1991). Head-driven processing models hypothesize that syntactic phrases are projected based on the information of the head; for example the verb phrase is projected when the verb information becomes available to the parser, and the propositional phrase with the information from the proposition. They predict that the parser processes input strings efficiently in a top-down manner for a head-initial language such as English, whose phrase structure starts with a head, as in (1).

(1) $[_{S}$ Jane $[_{VP}$ ate $[_{NP}$ a salad] $[_{PP}$ at $[_{NP}$ the café]]$[_{PP}$ at noon]]].

Mazuka and Lust (1987), however, pointed out such models would predict increased processing cost and difficulty when applied to a language like Japanese. Due to its head-final nature, word order flexibility, and null pronouns, the beginning of a sentence or phrase in Japanese is not an effective predictor of (a) forthcoming phrase(s). Even the nominative-marked noun phrase in (2a), a good candidate for a subject of a clause, may or may not be the beginning of a main clause. As shown in (2b), it may be part of a deeply embedded phrase where *Taroo-ga* 'Taro' is the subject of a relative clause within other relative clauses. The initial hypothesis about the structure the parser makes in Japanese may be reanalyzed frequently, or the parser must posit numerous structures in parallel or delay commitment until much later than in English. All possibilities predict that Japanese sentences are processed with a significantly increased processing cost, which goes against the native intuition of processing Japanese with just as much ease as any other language (cf. Mazuka and Itoh 1995).

(2) a. Taroo-ga . . .
Taro-NOM
'Taro. . .'

b. [[ $e_i$ [[ $e_j$ [[ [[ Taroo-ga $e_k$ katte iru]$_S$ inu$_k$-ni]$_{NP}$
Taro-NOM keeps dog-DAT
oikakerareta]$_S$ kodomo$_j$ -o]$_{NP}$ dakiageta ]$_S$ otoko$_i$-ga ]$_{NP}$ . . . ]
was chased child-ACC lifted up man-NOM
'The man who lifted up the child who was chased by the dog Taroo keeps . . .'
(Mazuka and Lust 1987: 336)

The study by Mazuka and Lust (1987) was followed by both theoretical and experimental studies on processing in Japanese.[3] Inoue (1991) and Inoue and Fodor (1995) proposed models of HLP that account for language types such as English as well as Japanese. Inoue and Fodor (1995) proposed "Information-based Parsing", a parser that builds partial structures before definite confirmation but adjusts its degree of commitment based on the nature of the linguistic input the parser receives. A number of experimental studies reported that in Japanese, a head-final language, the parser seems to anticipate the type of forthcoming verb or structure while it processes a series of preverbal case-marked arguments. Information available to the parser in preverbal positions, such as case markers, seems to play a significant role in the processing of Japanese (Kamide and Mitchell 1999; Yamashita 2000; Miyamoto 2002; Uehara and Bradley 2002; Oishi and Sakamoto 2004; Sato et al. 2009). Yoshida, Aoshima, and Phillips (2004) showed that the parser utilizes classifier information to predict the forthcoming structure of the relative clause.[4]

With distinct linguistic characteristics from English, HLP studies in LCI languages either provided the same evidence for hypotheses made in English to support their universality or eliminated confounding factors in studies of English to support or modify the hypotheses. For example, the head-final nature and flexible word order of Japanese enabled researchers to further examine hypotheses in filler-gap/gap-filler dependencies (e.g., Sakamoto 1995; Miyamoto and Nakamura 2003; Ueno and Garnsey 2008). Arai and Nakamura (2016) successfully eliminated a possible confounding factor of number of lexical items, observed in studies of "digging-in

**3** See Nakayama (1999) for a review of the main theoretical proposals and discussion of other phenomena that contribute to the complexity of Japanese sentence processing.
**4** See also Yamashita, Hirose, and Packard (2011) for investigations of processing in head-final languages, including Basque, Chinese, German, Hindi, Japanese, and Korean.

effects" in English (e.g., Ferreira and Henderson 1991; Tabor and Hutchins 2004). In production, Yamashita and Chang (2001) and Kondo and Yamashita (2011) showed that, as opposed to the "short-before-long" length-based phrasal ordering preference in head-initial languages (e.g., Stallings, MacDonald, and O'Seaghdha 1998; Arnold et al. 2000), speakers of Japanese demonstrated a "long-before-short" preference (cf. Hawkins 1994).

Over the years, numerous studies in a variety of languages have accumulated data and expanded the knowledge base for HLP in general, that is, principles common across all languages. As the knowledge base continues to be built and advances theories of HLP in general, let us consider a wider range of phenomena that can be observed through the lenses of the study of HLP, and the implications of such observations.

# 3 Getting a fuller picture: Individual LCI language processing model as a goal

## 3.1 Null elements, pre-head processing, and reanalysis in Japanese

In conducting experiments, choices are made about best practices for collecting data with as little noise as possible. This includes careful controls of experimental tasks and environments and setting criteria for subjects. For sentence processing or production studies, this means controlling the linguistic properties of stimuli sentences so they allow researchers to focus on the aim of the study as much as possible.

Stimuli in the study of Japanese and other head-final languages often use a sequence of words that appear preverbally, such as the case-marked arguments below.

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (3) | a. | NP | -ga | NP | -ni | NP | -o | ... | | | | | |
| | | | -NOM | | -DAT | | -ACC | | | | | | |
| | b. | NP | -ga | NP | -ni | NP | -o | NP | -ga | NP | -ni | ... | |
| | | | -NOM | | -DAT | | -ACC | | -NOM | | -DAT | | |

The array of nominative-, dative-, and accusative-marked arguments are often presented in canonical order or scrambled order, with a ditransitive verb as a probe compared with other types of verbs (e.g., Yamashita 1997, 2000). Some studies that involve embedded clauses present more than three overtly case-marked arguments.

Note that, although sentences with all overtly case-marked arguments as in (3) are certainly possible, the occurrences in actual use are infrequent and its common use may be approached with caution. A corpus study in Kondo and Yamashita (2011) using the *Corpus of Spontaneous Japanese* (National Institute for Japanese Language and Linguistics, National Institute of Information and Communications Technology, and Tokyo Institute of Technology 2004) reports that the number of occurrences of ditransitive clauses with all three overt arguments of nominative-, dative-, and accusative-marked sentences in any order is 24 (21 in canonical order, 3 scrambled) out of 80,000 clauses. Likewise, the number of occurrences of a transitive sentence with both overt nominative- and accusative-marked arguments in any order followed by a transitive verb was only 706 (660 in canonical order, and 46 scrambled). One factor in limiting occurrences of clauses with all overt arguments is due to null elements, that is, trace(s) and null pronouns commonly used in Japanese.

In English pronouns must be overtly present. In contrast, in Japanese these pronouns may be null, both in spoken and written language. It is possible that a string of all overtly case-marked arguments embodies one or more null pronoun(s) or trace(s), and the parser posits them incrementally, that is, processes with partial information available at each point (Kamide and Mitchell 1999). One possible example is shown in (4).

(4) a. *e* NP -ga *e* NP -ni *e* NP -o *e* . . .

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| b. | [ *e* | [ Takashi | -ga | [[[ $e_i$ | Kana | -ni | [[ $e_j$ | |
| | | | -NOM | | | -DAT | | |
| | purin | -o | tabeta] | hannin$_j$] | -o | osieta] | hito$_i$] | -wa |
| | pudding | -ACC | ate | suspect | -ACC | informed | person | -TOP |
| | Shoko | -da] | to | itta] | to | omotta]. | | |
| | | -cop | comp | said | comp | thought | | |

‘(I) thought that Takashi said the person who let Kana know the suspect who ate the pudding was Shoko’.

Two questions on examining preverbal arguments in processing Japanese arise. The first is, assuming Japanese is processed incrementally, in a string of all overtly case-marked arguments, how likely it is that the parser posits any null pronoun(s) or the trace(s) of forthcoming relative clause(s) along the way, and, if it does, where those null elements are posited. Even though experimental and theoretical studies are bound to be built on a set of assumptions, a thorough investigation of the psychological reality of null elements in Japanese may facilitate establishing more accurate hypotheses or reducing results that are unaccounted for.

The second is what (or whether) Japanese language users hypothesize when they see a sentence starting with arguments other than nominative-marked arguments, as in (5a) or (5b). Do they hypothesize that the argument is a part of a scrambled sentence or a sentence with a null pronoun? When (or whether) do they change the initial hypothesis?

(5) a. NP -ni ...
-DAT
b. NP -o ...
-ACC

An event-related potential (ERP) study by Ueno and Kluender (2003) reports that when Japanese subjects saw an accusative-marked noun phrase as in (5b) as an initial part of a single transitive scrambled sentence, they observed slow anterior negative potentials followed by P600 and left anterior negativity, the ERP potentials associated with different types of processing difficulty in English (Ueno and Kluender 2003: 258). They interpret the effects as follows: the parser treats the sentence-initial accusative-marked noun phrase as a scrambled argument (a filler), stores it in the working memory, syntactically integrates filler-gap dependency as the readers proceed to read nominative-marked subject and an adverb immediately before the gap, and at the gap retrieves the information from the scrambled accusative-marked noun phrase for the gap. Their study suggests that a sentence-initial accusative-marked noun phrase in Japanese is not interpreted as an object of a sentence with a null subject but as a scrambled argument.

Study such as Ueno and Kluender (2003) offers a possible answer for a piece of the puzzle on pre-verbal processing in Japanese. It also leads to additional questions, such as whether the effects will be different if a preceding context is given, what happens at the verb if the sentence is indeed a canonical sentence with a null subject, and how expectations change as readers process more arguments and other lexical items emerge (Sato et al. 2009; Sakamoto 2015), among others.

Equally important in investigating processing in Japanese are theoretical and experimental studies of reanalysis cost in Japanese, the processing cost to revise the analysis. It has been suggested that reanalysis in Japanese may be less costly than in English (Inoue and Fodor 1995; Mazuka and Itoh 1995)[5] or that optional reanalysis may take place in Japanese (Yamada, Arai, and Hirose 2017), but the exact nature of reanalysis in a language like Japanese is still unknown. Examining the nature of reanalysis and its cost is essential in the further development of a processing model

5 Also note that in theoretical linguistics it has been reported that ungrammaticality of subjacency and the empty category principle (ECP) in Japanese are not as severe as those in English (Saito and Fukui 1998; Sprouse et al. 2011).

of Japanese. Moreover, the contrast of the Japanese processing model with that of English will lead to a fuller picture of HLP in general.

Note that not all studies on the processing of individual languages may entail direct implications for HLP of all languages. HLP studies in any language need to be on solid ground regardless of the goal, however. More detailed examination of processing and the development of processing models for each language is necessary for conducting studies of both HLP in general and in individual languages.

## 3.2 Factoring in language change

Setting HLP models for individual languages as a goal also entails an important aspect of observing and contouring to changings in languages. All languages change over time, and so are the uses by their users. Sometimes languages change due to a shift in language use. Even adult language users change the way they use their knowledge of language during processing, based on changes in their exposure to the language (Wells et al. 2009; Dabrowska 2018).

Recent development of social networking services (SNSs) and digital communications have changed a great portion of one's exposure to "written" language. Nishikawa and Nakamura (2015) note that language used in digital communication tends to be short and truncated, its discourse halted in the middle and continuing after a long interval (such as a reply sent the next day); multiple topics are simultaneously handled smoothly; and visual input such as stamps and emoji play a significant part of communication. Another characteristic of SNS or reading on digital devices is that sentences in Japanese are overwhelmingly written horizontally, unlike traditional vertical writing in hard copy books and newspapers. Furthermore, the scrolling of digital screens that enables the gaze to stay in the center of a screen may change eye movement compared to traditional media, in which eyes move over wider areas on the two-dimensional pages of hard copy books and newspapers. It is not surprising that a generation of people accustomed to reading and writing on digital platforms process sentences differently from pre-digital generations. Capturing such processing changes may be also part of the goal of studying HLP in individual languages.

Language change as reflected in case marking also needs close observation and may potentially be woven into a HLP model of an individual language. The most common use of the case marker *ga* in Japanese is as subject marking for all predicates, except for the limited case of predicates used for marking an object. Recently, there have been more cases observed in which *ga* does not mark subjects for normal predicates. See the utterances observed on TV below (Seito 2021; Nakamura 2021):

(6) a. (being asked what the department store basements mean to her)

| | | | | | | |
|---|---|---|---|---|---|---|
| Nanka | genki | -ga | kureru | basho | desu | ne. |
| sort of | mojo | -NOM | gives | place | cop | sp |

# '(It)'s a place that mojo gives me'.
'(It)'s a place that gives me mojo'.

b. (being solicited for an opinion in an interview)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| ?? | *e* | Kinkyuuzitaisengen | -ga | mada | haturee | site | -iru. |
| | | state of emergency | -NOM | still | issue | do | pr prog |

'(We) are still under a state of emergency'.

In (6a), *genki* 'mojo' is marked by a nominative marker *ga* and it appears to be the subject of the sentence. Reading as such, her utterance is interpreted as '(It)'s a place that mojo gives me'. It does not match the context, however, and the speaker said it without any disfluency nor a pause to signal it is an afterthought. Coupled with the preceding context, it suggests that she meant '(It)'s a place that gives me mojo'. For a sentence with such a meaning, *genki* should be marked by *o*. Likewise, the speaker in (6b) marked the object with *ga* and uttered the sentence without any disfluency. The use of *ga* in these utterances may belong to cases of production error, that is, the speaker was unable to correctly conclude the sentence with the proper verb.[6] Alternatively, due to the frequency of such errors, or for other reasons, the sentence-initial *ga* may be beginning to take on a different role in addition to the existing one of nominative-case marking.

HLP is the use of knowledge of language by humans. In order to accurately observe HLP in any language, it is necessary to have a thorough knowledge of the language, including its changes over time. These changes may be monitored and reflected in the updated model, if the HLP for each individual language is set as another goal of HLP study. Furthermore, compilation of such individual models and comparisons between them also may facilitate the development of a model of HLP in general.

**6** I thank Kentaro Nakatani for raising the possibility that *kureru* '(someone) gives (something) to me' in (6a) may be the result of a production error. The correct use of giving and receiving verbs is indeed difficult even for Japanese speakers, and the speaker may have not been able to retrieve the lexical item correctly. She may have failed to retrieve the correct verb, *moraeru* '(I) can receive'. At the same time, the fact that the verbs may elicit frequent errors could further facilitate a change in the role of *ga* marking.

# 4 Observations and applications: Mechanisms of HLP by a wider range of language users

Most studies of HLP have focused on processing in people with a typical cognitive system: young (college-age) adults with normal hearing, and normal or corrected vision. Let us consider another dimension that may be fully integrated with the study of HLP: investigation of HLP in people with an atypical cognitive system.[7] There are many cases of atypical use of language knowledge due to limitations in human memory, attention, perception, or speech motor control, among others. Atypical language use is often observed in people with dyslexia, dysgraphia, autism spectrum disorder (ASD), Williams Syndrome, or visual impairment, and in people who are deaf or hard of hearing (D/HOH), among others. While the number of such language users is small compared to the typical group, the investigation of their mechanism of HLP is crucial for assisting their language development and learning, among others.

Studies of HLP have employed or newly developed a variety of experimental methods that measure human reactions in using the knowledge of language online (self-paced reading, eye tracking, visual-world paradigm, cross-modal priming, ERP, fMRI, PET, NIRS, grammatical decision, lexical decision, among others) and off-line (acceptability/grammatical questionnaire rating, corpus analysis, priming, maze test, among others). Online tasks use computers or other mechanical equipment and enable researchers to collect responses in milliseconds as subjects read sentences. Corpus analyses enable the observation of occurrences of certain linguistic structures from a large amount of data in actual use.

These tasks commonly employed in the study of HLP in general may facilitate enhancing the understanding of HLP in atypical groups by adding data and observations obtained by different methods to complement existing studies. For example, tasks that do not involve human interactions, such as written or spoken dialogues framed as games, may solicit additional data from children with ASD. Investigations of online responses to written language might allow researchers to identify more fine-grained sources of difficulty in reading by the D/HOH population or students with dyslexia. Let us discuss such possibilities by taking the D/HOH as an example.[8]

7 Current section focuses on (young) adult atypical language users to highlight contrast with typical language user group. However, HLP in children with typical or atypical cognitive system, that in adults using second/foreign languages, and in the elderly, all add dimensions to understanding of how humans use the knowledge of languages. While HLP in children has been explored for decades, HLP in second language users and elderly still await in-depth investigation.

8 Many researchers in a variety of disciplines including medical, disability education, educational psychology, and psychology have been working on understanding the mechanisms of language

## 4.1 Observations from multiple angles: HLP in atypical language users

Studies in English report that reading comprehension skills by D/HOH often lag behind peers of the same age. A study by Newport and Meier (1985) reports that the children exposed to American Sign Language (ASL) as their first language demonstrate stages of language acquisition in ASL parallel to those of spoken language acquisition (also see Cheng and Mayberry 2019). The onset, type, and amount of input in English to D/HOH children in English-speaking countries varies. Although some speakers with residual hearing may acquire English orally, to others, learning English is similar to second language learning (Berent 1996; Piñar, Dussias, and Morford 2011). Due to diverse backgrounds in the acquisition of (a) language(s) by D/HOH, there is great variability in their comprehension of written language (see Bochner and Albertini 1988; Berent 1996 for review). How exactly written languages are processed by D/HOH remains unclear and a challenge. A study on deaf readers by Piñar and colleagues (2011) notes, "The extent to which Deaf readers rely on lexical, semantic, and syntactic information to process sentences remains poorly understood" (Piñar, Dussias, and Morford 2011: 695).

Processing of the Japanese language by the Japanese D/HOH population also remains to be investigated in depth (Sawa 2015). The general sentence comprehension skills of Japanese by D/HOH children tend to remain at the third-grade level. More than forty percent of high-school-age D/HOH students showed deficits in syntactic structures (Minamide and Shindo 1984). Commonly observed are strategized processing based on one's own experience or lexical meaning; difficulty in understanding syntactic structures of causatives, passives, and giving/receiving; lack of full understanding of elements of sentences that bear grammatical roles, such as subject, object, and verb; and poor distinction between transitive and intransitive verbs (see Agatsuma 2000 for review). Crucially, lack of full use of case-marking information and their integration with syntactic knowledge is commonly observed among D/HOH population. Also there is a strong tendency to use the *ga* marker as agent even in passives and giving/receiving sentences (Agatsuma, Sugawara, and Imai 1980).

Some of the experimental methods in HLP may offer possible avenues for investigating the reading process of D/HOH readers from different angles. A comparison of ERPs in hearing and D/HOH groups may reveal the role of syntactic and

use by atypical groups. The strategies commonly used in sentence processing and production may further facilitate the process of understanding these mechanisms in depth, as well as their applications to enhance the language acquisition or use by atypical groups.

semantic information in both groups. A self-paced reading passage with a variety of syntactic structures followed by simple content questions may reveal the overarching cause of processing difficulty, as well as where the difficulty arises within the experimental sentences.

Note that some studies of HLP in D/HOH populations have applicational or pedagogical implications. For example, in reading languages with multiple orthographic possibilities, self-paced or eye-tracking data of the same sentence with different orthography may reveal the optimal ways for readers with different linguistic backgrounds to process written material. The sentence in (7) may be written in multiple ways by combinations of Hiragana, Katakana, and Kanji characters and spacing between words. Some of the possibilities are shown in (8a–c).

| | | | | | | |
|---|---|---|---|---|---|---|
| (7) | Simin | -no | marason taikai | -wa | yokka | asa |
| | citizen | -GEN | marathon meet | -TOP | fourth | morning |
| | sitizi | -ni | hajimarimasu. | | | |
| | seven | at | start | | | |

‘The citizens’ marathon meet starts at seven in the morning on the fourth’.

(8) a. 市民のマラソン大会は4日朝７時に始まります。
b. しみんのまらそんたいかいはよっかあさ７じにはじまります。
c. しみんの　まらそんたいかいは　よっか　あさ　７じに
はじまります。
‘The citizens’ marathon meet starts at seven in the morning on the fourth’.

The sentences in (8a–c) are the same except for orthography and the space between *bunsetsu* (content words followed by a case marker or other morphemes). Detailed examinations of which pattern(s) is/are the most effective for different D/HOH readers may facilitate making written documents in Japanese more universally accessible.

A comprehensive study of mental grammar and processing models of atypical language users may facilitate understanding the mechanism of HLP in diverse groups of language users. Historically, hypotheses of different syntactic structures have been made to capture phenomena in both theoretical and experimental linguistics.[9] Bochner (1978) hypothesized a linear-sequential rather than hierarchical linguistic representation of English based on the grammatical judgment of multi-clausal sen-

9 For example, in order to capture the nature of scrambling in Japanese, the accounts using configurational and non-configurational syntactic structures were proposed and they enabled systematic examinations of the phenomenon (Farmer 1980; Hale 1980; Hoji 1987; Miyagawa 1989).

tences by D/HOH subjects at the college age. Discussions from such approaches may deepen the understanding of HLP in atypical language users.

Taking it further, investigations in the processes of learning/acquiring a language through individual patterns of input – be it entirely visual input in the case of D/HOH children or mixed inputs of both auditory and visual for hearing children – may deepen the understanding of not only how processing works for D/HOH language learners but also how humans with a variety of linguistic input may establish ways to process languages. Such understanding may facilitate exploration of the best ways to teach written language to each group. Likewise, investigations in language acquisition through assessing the input to ASD children, and understanding their process of HLP, might facilitate guiding their use of their first language.

## 5 Summary

This chapter proposed two dimensions that may enhance studies of HLP: an HLP model for individual languages and HLP in atypical language users. It was argued that revealing the HLP mechanisms of individual languages, and of typical and atypical language users, both feed into understanding HLP in general as well as benefiting the studied group.

There are challenges in both dimensions when one ventures into them. The academic communities must be receptive to such an approach, and stable venues for discussion and dissemination are necessary. Learning about new subjects – especially those with special needs – is essential if studies are to collect data respectfully and appropriately, yet such information may not be readily accessible to all researchers. The field of HLP, however, has always been interdisciplinary and made advances through collaborations beyond disciplines. Through cross-pollinations across disciplines, such goals would be made possible.

## References

Agatsuma, Toshihiro. 2000. Tyookakusyoogaizi-no bunrikainooryoku-ni kansuru kenkyuu-no dookoo [Review of the sentence processing abilities in deaf children]. *Tokushukyoikugaku Kenkyu* [The Japanese Journal of Special Education]38(1). 85–90.

Agatsuma, Toshihiro, Koichi Sugawara, & Hideo Imai. 1980. Tyookakusyoogaizi-no gengonooryoku <III> : Ukemi, yarimoraibun-no rikai. [Linguistic ability of hearing-impaired children <III>: comprehension of passive and giving/receiving sentences.] *Kokuritsu Tokushukyooiku Soogookenkyuujo Kenkyuu Kiyoo* [Bulletin of The National Institute of Special Needs Education] 7. 39–47.

Arai, Manabu & Chie Nakamura. 2016. It's harder to break a relationship when you commit long. *PLOS ONE* 11(6). e0156482.
Arnold, Jennifer, Anthony Losongco, Thomas Wasow, & Ryan Ginstrom. 2000. Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language* 76(1). 28–55.
Berent, Gerald. 1996. The acquisition of English syntax by deaf learners. In William C. Ritchie & Tej Bhatia (eds.), *Handbook of Second Language Acquisition*, 469–506. The Netherlands: Academic Press.
Bochner, Joseph. 1978. Error, anomaly and variation in the English of deaf individuals. *Language and Speech* 21(2). 174–189.
Bochner, Joseph & John Albertini. 1988. Language varieties in the deaf population and their acquisition by children and adults. In Michael Strong (ed.), *Language Learning and Deafness*, 3–48. Cambridge: Cambridge University Press.
Cheng, Qi & Rachel Mayberry. 2019. Acquiring a first language in adolescence: The case of basic word order in American Sign Language. *Journal of Child Language* 46(2). 214–240.
Dabrowska, Eva. 2018. Experience, aptitude and individual differences in native language ultimate attainment. *Cognition* 178. 222–235.
Farmer, Ann. 1980. *On the interaction of morphology and syntax.* Cambridge, MA: Massachusetts Institute of Technology dissertation.
Ferreira, Fernanda & John Henderson. 1991. Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language* 30. 725–745.
Hale, Ken. 1980. Remarks on Japanese phrase structure: Comments on the papers on Japanese syntax. In Ann Farmer & Yukio Otsu (eds.), *Theoretical Issues in Japanese Linguistics*, 185–203. (MIT Working Papers in Linguistics 2). Cambridge, MA: Department of Linguistics and Philosophy, Massachusetts Institute of Technology.
Hawkins, John. 1994. *A Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.
Hoji, Hajime. 1987. Weak crossover and Japanese phrase structure. In Takashi Imai & Mamoru Saito (eds.), *Issues in Japanese Linguistics*, 163–201. Dordrecht: Foris.
Inoue, Atsu. 1991. *A comparative study of parsing in English and Japanese.* Storrs, CT: University of Connecticut dissertation.
Inoue, Atsu & Janet Fodor. 1995. Information-paced parsing of Japanese. In Reiko Mazuka & Noriko Nagai (eds.), *Japanese Sentence Processing*, 9–63. Hillside, NJ: Lawrence Erlbaum Associates.
Kamide, Yuki & Don Mitchell. 1999. Incremental pre-head attachment in Japanese parsing. *Language and Cognitive Processes* 14(5–6). 631–662.
Kimball, John. 1973. Seven principles of surface structure parsing in natural language. *Cognition* 2. 15–47.
Kondo, Tadahisa & Hiroko Yamashita. 2011. Why speakers produce scrambled sentences: Analyses of spoken language corpus in Japanese. In Hiroko Yamashita, Yuki Hirose, & Jerry Packard (eds.), *Processing and Producing Head-Final Structures*, 61–65. Dordrecht: Springer.
Mazuka, Reiko & Kenji Itoh. 1995. Can Japanese speakers be led down the garden path? In Reiko Mazuka & Noriko Nagai (eds.), *Japanese Sentence Processing*, 295–329. Hillsdale, NJ: Lawrence Erlbaum Associates.
Mazuka, Reiko & Barbara Lust. 1987. Why is Japanese not difficult to process?: A proposal to integrate parameter setting in Universal Grammar and parsing. In Reiko Mazuka & Barbara Lust (eds.), *North East Linguistics Society* (*NELS*) 18, 333–356. Amherst, MA: the Graduate Linguistics Students Association, University of Massachusetts, Amherst.

Minamide, Yoshifumi & Hiroshi Shindo. 1984. Roogakooseito-no toogonooryoku-no hyooka-ni kansuru kennkyuu [A study of evaluation of syntactic ability by deaf students]. *Chookakugengo Shogai* [The Japanese journal of hearing and language disorders] 13(4). 165–172.
Miyagawa, Shigeru. 1989. *Structure and Case Marking in Japanese*. New York: Academic Press.
Miyamoto, Edson. 2002. Case markers as clause boundary inducers in Japanese. *Journal of Psycholinguistic Research* 31(4). 307–347.
Miyamoto, Edson & Michiko Nakamura. 2003. Subject/object asymmetries in the processing of relative clauses in Japanese. In Gina Garing & Mimu Tsujimura (eds.), *Proceedings of the 22nd West Coast Conference on Formal Linguistics*, 342–355. Somerville, MA: Cascadilla Press.
Nakamura, Kenji (dir.). 2021. *Shin Nihon Fudoki*: *Tokyo-no Chika* [Underground Tokyo]. Tokyo: Nippon Hoso Kyokai. (documentary)
Nakayama, Mineharu. 1999. Sentence processing. In Natsuko Tsujimura (ed.), *The Handbook of Japanese Linguistics*, 398–424. Boston: Blackwell.
National Institute for Japanese Language and Linguistics, National Institute of Information and Communications Technology & Tokyo Institute of Technology. 2004. *Corpus of Spontaneous Japanese*. https://clrd.ninjal.ac.jp/csj/en/index.html
Newport, Elissa & Richard Meier. 1985. The acquisition of American Sign Language. In Dan Slobin (ed.), *The Crosslinguistic Study of Language Acquisition. Volume 1: The Data*, 881–938. Hillsdale, NJ: Lawrence Erlbaum.
Nishikawa, Yuske & Masako Nakamura. 2015. *LINE komyunikesyon no tokusee no bunseki* [Analyses of characteristics of LINE communication]. *Journal of Information Studies* 16. 47–57.
Oishi, Hiroaki & Tsutomu Sakamoto. 2004. An ERP study on the timing of the parser's structural decision: Is the parsing performed in an incremental manner or in a delayed manner? *Cognitive Studies* 11(4). 311–318.
Oku, Satoshi. 1998. *A theory of selection and reconstruction in the minimalist program*. Storrs, CT: University of Connecticut dissertation.
Otaki, Koichi. 2014. *Ellipsis of arguments: Its acquisition and theoretical implications*. Storrs, CT: University of Connecticut dissertation.
Piñar, Pilar, Paola E. Dussias, & Jill P. Morford. 2011. Deaf readers as bilinguals: An examination of deaf readers' print comprehension in light of current advances in bilingualism and second language processing. *Language and Linguistics Compass* 5(10). 691–704.
Pritchett, Bradley. 1991. Head position and parsing ambiguity. *Journal of Psycholinguistic Research* 20(3). 251–270.
Saito, Mamoru. 2004. Ellipsis and pronominal reference in Japanese clefts. *Nanzan Linguistics* 1. 21–50.
Saito, Mamoru & Naoki Fukui. 1998. Order in phrase structure and movement. *Linguistic Inquiry* 29(3). 439–474.
Sakamoto, Tsutomu. 1995. Transparency between parser and grammar: On the processing of empty subjects in Japanese. In Reiko Mazuka & Noriko Nagai (eds.), *Japanese Sentence Processing*, 275–294. Hillsdale, NJ: Lawrence Erlbaum Associates.
Sakamoto, Tsutomu. 2015. Processing of syntactic and semantic information in the huan brain: Evidence from ERP studies in Japanese. In Mineharu Nakayama (ed.), *Handbook of Japanese Psycholinguistics*, 457–510. Berlin: De Gruyter Mouton.
Sakamoto, Yuta. 2016. Phases and argument ellipsis in Japanese. *Journal of East Asian Linguistics* 25(3). 243–274.
Sanz, Montserrat, Itziar Laka, & Michael Tanenhaus (eds.). 2015. *Language Down the Garden Path: The Cognitive and Biological Basis for Linguistic Structures.* Oxford: Oxford University Press.

Sato, Atsushi, Baris Kahraman, Hajime Ono, & Hiromu Sakai. 2009. Expectation driven by case-markers: Its effect on Japanese relative clause processing. In Yukio Otsu (ed.), *The Proceedings of the Tenth Tokyo Conference on Psycholinguistics*, 215–237. Tokyo: Hituzi Syobo.
Sawa, Takashi. 2015. Tyookakusyoogaizi-no bun-no rikai-ni kansuru kenkyuu dookoo: Bunrikaihooryaku-ni kansuru bunkenteki koosatsu [Strategies for sentence comprehension in deaf children: a review]. *Tokyo Gakugei Daigaku Kyoiku Jissen Kenkyuu Shien Sentaa Kiyoo* [Bulletin of Center for the Research and Support of Educational Practice] 11. 115–123.
Seito, Yasushi (dir.). 2021. Nyuusu 7 [News 7]. Tokyo: Nippon Hoso Kyokai. (broadcast on 9 September, 2021)
Sprouse, Jon, Shin Fukuda, Hajime Ono, & Robert Kluender. 2011. Reverse island effects and the backward search for a licensor in multiple *wh*-questions. *Syntax* 14(2). 179–203.
Stallings, Lynn, Maryellen MacDonald, & Padraig O'Seaghdha. 1998. Phrasal ordering constraints in sentence production: Phrase length and verb disposition in heavy-np shift. *Journal of Memory and Language* 39(3). 392–417.
Tabor, Whitney & Sean Hutchins. 2004. Evidence for self-organized sentence processing: Digging-in effects. *Journal of Experimental Psychology, Learning, Memory and Cognition* 30(2). 431–450.
Uehara, Keiko & Diane Bradley. 2002. Center-embedding problem and the contribution of nominative case repetition. In Mineharu Nakayama (ed.), *Sentence Processing in East Asian Languages*, 257–287. Stanford, CA: Center for the Study of Language and Information (CSLI).
Ueno, Mieko & Susan Garnsey. 2008. An ERP study of the processing of subject and object relative clauses in Japanese. *Language and Cognitive Processes* 23(5). 646–688.
Ueno, Mieko & Robert Kluender. 2003. Event-related brain indices of Japanese scrambling. *Brain and Language* 86(2). 243–271.
Wells, Justin, Morten Christiansen, David Race, Daniel Acheson, & Meryellen MacDonald. 2009. Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology* 58. 250–271.
Yamada, Toshiyuki, Manabu Arai, & Yuki Hirose. 2017. Unforced revision in processing relative clause association ambiguity in Japanese: Evidence against revision as last resort. *Journal of Psycholinguistic Research* 46(3). 661–714.
Yamashita, Hiroko. 1997. The effects of word-order and case marking information on the processing of Japanese. *Journal of Psycholinguistic Research* 26(2). 163–188.
Yamashita, Hiroko. 2000. Structural computation and the role of morphological markings in the processing of Japanese. *Language and Speech* 43. 429–459.
Yamashita, Hiroko & Franklin Chang. 2001. "Long before short" preference in the production of a head-final language. *Cognition* 81. B45–B55.
Yamashita, Hiroko, Yuki Hirose, & Jerry Packard (eds.). 2011. *Processing and Producing Head-Final Structures*. Dordrecht: Springer.
Yoshida, Masaya, Sachiko Aoshima, & Collin Phillips. 2004. Relative clause prediction in Japanese. Paper presented at the 17th annual City University of New York (CUNY) Conference on Human Sentence Processing, College Park, MD, 25–27 March, 2004.

Matthew Wagers

# Chapter 3
# Encoding interference in verb-initial languages

## 1 Similarity in syntactic processing

The idea that working memory constraints play some role in language processing is a foundational one in psycholinguistics. A usual starting point is the analysis of nesting constructions in Miller and Chomsky (1963), who asserted that "available memory (i.e., number of states) is clearly quite limited for real-time analytic operations . . . from these observations we are led to conclude that sentences of natural languages containing nested dependencies or self-embedding beyond a certain point should be impossible for (unaided) native speakers to understand." Perhaps the most pathological example of overloaded memory comes from the center self-embedded sentence, as in (1), which most English speakers would judge as unacceptable. In (1), three critical DPs are underlined which have to be correctly paired with its matching predicate.

(1) The chef$_1$ [ that the critic$_2$ [ that the artist$_3$ recommended$_3$ ] praised$_2$ ] was$_1$ grateful for the press.

As Bever (1974) and others have observed, it is possible to improve the parsability of sentences like (1) by using DPs that are more distinct. Each DP in (1) is similar syntactically and semantically: each combines the determiner *the* with an NP to form a definite description. If we improve upon the variety of DPs, however, we seem to improve the acceptability of the sentence:

(2) The chef$_1$ [ that [ everyone$_2$ [ that you$_3$ recommended$_3$ ] praised$_2$ ] was$_1$ grateful for the press.

The underlined DPs in (2) differ along many dimensions, including their internal syntactic structure and how they introduce referents into the discourse (cf. Warren and Gibson 2002, 2005). But the contrast between (1)/(2) suggests we need a theory of the parsing/memory interaction that is sensitive not just to order and constituency of encoded information, but to the kind of information encoded. As it turns out, the same contrast obtains in much simpler singly-embedded object relative clauses (ORCs), as in (3) v. (4).


https://doi.org/10.1515/9783110778946-003

(3) The banker$_1$ [ that the barber$_2$ praised$_2$ ] climbed$_1$ the mountain.

(4) The banker$_1$ [ that you$_2$ praised$_2$ ] climbed$_1$ the mountain.

In a series of influential papers (Gordon, Hendrick, and Johnson 2001; Gordon Hendrick, and Levine 2002; Gordon, Hendrick, and Johnson 2004; Gordon et al. 2006), Gordon and colleagues have shown that ORCs with two highly similar DPs are harder to comprehend, compared to subject relative clauses (SRCs) with the same DPs. But when the DPs are dissimilar, the difficulty associated with ORCs is considerably reduced or eliminated. Thus the ORC in (3) is harder to comprehend than the SRC in (5); but the ORC in (4) is not harder to comprehend than the SRC in (6):

(5) The banker$_{1/2}$ [ that praised$_2$ the barber ] climbed$_1$ the mountain.

(6) The banker$_{1/2}$ [ that praised$_2$ you ] climbed$_1$ the mountain.

An obvious and important question – to which we return shortly – is how similarity should be defined. But first, let us consider the account proposed by Gordon, Hendrick, and Johnson (2001) to explain the interaction between DP type and RC difficulty:

> The parsing and semantic interpretation of a sentence require that intermediate representations be held in memory and addressed during comprehension. Object-extracted constructions impose greater demands of this sort than do subject-extracted constructions because **they require that two NPs be stored and subsequently accessed while subject-extracted constructions do not**. The differing functions of those two NPs are specified by the order in which they appear in the sentence. **Memory for order information is impaired** when the items to be remembered are similar **because the similarity of the items causes interference in retrieving the order information** (Lewandowsky and Murdock 1989; Murdock and Vom Saal 1967; Nairne 1990). (1420) [Emphasis by MW]

The two highlighted passages provide a roadmap for our own inquiry here:

(i) Why does the similarity of two constituents impair the recovery of order information? Is it always an impairment?

(ii) Do similarity effects only impinge upon dependency formation when it involves constituents that are stored and later re-accessed? Or, more concretely: is there something especially pernicious about the N – N – V orders compared to N – V – N?

In response to question (i), we will turn first in Section 2 to two major sources of similarity-based interference that have been differentiated in language processing: retrieval interference and encoding interference. We will argue that similarity-based impairments in ORC processing are most plausibly conceived of as species of *encoding interference*. The underlying cause of this encoding problem is not storage and recovery of linear order per se, but rather the reanalysis that is often involved in ORC processing.

As a consequence – and in response to question (ii) – we will argue that N – V – N word orders are liable to the same encoding interference, if they implicate reanalysis. This is noteworthy, as N – V – N word orders are common RC word orders in languages with verb-initial clauses. We draw upon recent evidence from relative clause processing in Chamorro, Zapotec, and English to support this claim.

Finally, we conclude with a conjecture. Based on our reanalysis account, encoding interference should itself have both destructive and constructive effects. Whereas the effect on the reanalyzed N is predicted to be destructive, leading to a degraded encoding, the second N should benefit from what we call "constructive interference," leading to an enhanced encoding. We consider some ways to test this prediction.

## 2 Motivating encoding interference

Retrieval interference is by now a familiar tool for explaining why some dependencies are more difficult to form than others (Van Dyke and Lewis 2003; Lewis, Vasishth, and Van Dyke 2006; Badecker and Lewis 2007; Wagers, Lau, and Phillips 2009; Dillon et al. 2013). The basic idea is this: to form a dependency between two elements in a sentence, the right-dependent provides a retrieval context for recalling or reactivating the left-dependent. The properties of the right-dependent, combined with the systematic knowledge embedded in the parser/grammar, provide the cues for this retrieval. The optimal scenario is when the retrieval context provides strong, unique cues. The equation in (7) expresses this idea by defining the probability of sampling a particular item in memory based on the cue as strength of association *s*(•) between the cue and the memory; divided by the sum of the strengths of association between and *all* items in memory (see, e.g., Nairne 1990).

$$P(M_i|Q_j) = \frac{s(Q_j, M_i)}{\sum_i^N s(Q_j, M_i)} \propto \frac{Match}{Selectiveness} \qquad (7)$$

In other words, the probability of sampling a target item depends on both how much the cue *matches* the target in memory and how *selectively* it does so. To illustrates how this works, consider the contrast in (8), from Arnett and Wagers (2017).

(8) a. The explorer$_{[\text{NOM,MAT}]}$ who believed that the monster$_{[\text{NOM,EMB}]}$ was prowling the ruins **went insane** . . . S-COMP

b. The explorer$_{[\text{NOM,MAT}]}$ who believed the monster$_{[\text{ACC,EMB}]}$ to be prowling the ruins **went insane** . . . ECM

Observe that (8a) and (8b) contain the same matrix clause, but vary in what kind of clause is embedded inside the subject-attached relative clause. In (8a), the embedded clause is a full, tensed clause complement of *believe*. In (8b), the embedded clause is a non-finite complement of *believe* in an ECM construction, which crucially has a non-nominative subject. When the comprehender has processed the relative clause and must then link the matrix TP *went insane* with its corresponding external argument, assume that they attempt to reactivate potential external arguments in memory using cues to case [NOM] and clause membership [MAT]. In S-COMP sentence (8a), the [NOM] cue is associated with both *explorer* and *monster*, reducing its selectiveness. In contrast, in the ECM sentence (8b), the [NOM] cue is uniquely associated with the correct item, *explorer*. As a consequence, retrieval is expected to be more effective in (8b) compared to (8a). Consistent with that prediction, Arnett and Wagers (2017) found that readers took longer to read the critical region in (8a) compared to (8b) as measured in total times; the probability of a regressive saccade out of the critical region was higher in (8a) compared to (8b).

Could retrieval interference be responsible for the effects of similarity found in ORCs? For example, could it explain the greater ease in processing (4) compared to (3), (repeated below)?

(3) The banker$_1$ [ that the barber$_2$ praised$_2$ ] climbed$_1$ the mountain.

(4) The banker$_1$ [ that you$_2$ praised$_2$ ] climbed$_1$ the mountain.

To pursue an explanation along these lines would require us to identify the RC verb as a retrieval context, one which provides cues that are less selective in (3) compared to (4). This is conceivable, although there is a source of conceptual friction to doing so: nothing about the relationship between the verb and its subject or object dependents is easily characterizable in terms that differentiate pronouns and descriptions, or names, or quantifier phrases, or any of the other kinds of expressions that lessen the ORC penalty.[1] A striking example of the problem here comes

**1** This is not to deny the possibility of retrieval effects in ORCs, and the necessity of retrieving the RC-external DP. The question is whether that will be sufficient to explain the interference effects observed, given the wide-ranging nature of the similarity effects that have been discovered and the fact that the dimensions of similarity they fall along are not (generally) directly related to an ORC's from a recent self-paced reading experiment by Villata, Tabor, and Franck (2018). They showed that mismatches in grammatically-marked gender make Italian ORCs easier to understand, at the ORC verb, even though that verb does not bear any gender marking itself. In (9), two conditions from their experiment are shown. (9a) includes two masculine DPs, while (9b) includes one masculine DP and one feminine DP (cf. the determiners *il* v. *la*).

| | | | | | | |
|---|---|---|---|---|---|---|
| (9) | a. | Il ballerino$_{[\text{MASC}]}$ | che | il cameriere$_{[\text{MASC}]}$ | ha sorpreso | . . . |
| | | the dancer.M | that | the waiter.M | has surprised | |
| | | 'The dancer that the waiter has surprised . . .' | | | | |
| | b. | Il ballerino$_{[\text{MASC}]}$ | che | la cameriera$_{[\text{FEM}]}$ | ha sorpreso | |
| | | the dancer.M | that | the waiter.F | has surprised | |
| | | 'The dancer that the waiter has surprised . . .' | | | | |

ORCs with either two feminine DPs or two masculine DPs had longer reading times at the ORC verb, compared to feminine-masculine combinations in either order. It is challenging to accommodate these results by appealing to retrieval interference, because it requires postulating that the verb, as the retrieval context,[2] provides cues to DP gender despite there not being any gender marking on the verb.

Retrieval interference allows two DPs to interact only indirectly, via a kind of competition that occurs when retrieval cues are unselective. Encoding interference, on the other hand, allows us to hypothesize that the two DPs interact directly with one another when they are encoded in a position. Contemporary theories of encoding interference postulate that the features which comprise the representations of complex, compositional objects are themselves are fundamentally labile. That is to say, those features can migrate among items in the representation, be forgotten, or be hallucinated. Figure 1 presents a templatic scenario in which two items, X & Y, are encoded in an ordered representation, represented by positions 1 and 2.

syntactic derivation. A potential way to overcome this friction is to build cue sets that directly reference prominence hierarchies, and that explicitly, but defeasibly, target the optimal feature combinations for a given grammatical role (Silverstein, 1976; Ariel, 1990; Minkoff, 2000). We leave this as an area for further theoretical development, although the results of Villata, Tabor, and Franck (2018) discussed above remains a likely sticking point.

**2** The assumption that the verb is the retrieval context may not be correct, as Omaki et al. (2015) has argued.


**Figure 1:** Schematic representation of two items in an ordered representation.

In *feature overwriting* theories (Nairne 1990; Oberauer and Kliegl 2006) shared features between X & Y can get overwritten – weakening the content of either X or Y. In *superposition* theories (Farrell and Lewandowsky 2002), if X & Y are similar, then Y effectively contributes some of its features to the preceding position – strengthening the binding of X to position 1. Most research on similarity and order comes from list recall experiments, and there an important assumption is that items get associated – somehow – with preceding positions (cf. Polyn, Norman, and Kahana 2009). We will return to this assumption, and its relevance for the ORC problem, shortly. But to see how it could play out in a non-linguistic study, consider an experiment by Oberauer et al. (2012). In that study, participants memorized lists of words with the goal of recalling particular targets that were either phonologically similar or dissimilar to distractors that followed them. For example, a SIMILAR list in their study was, e.g., "**baff** *daff haff* **vame** *rame pame* **nidd** *jidd gidd*", where bold face marks the targets and italics the distractors. In contrast, a DISSIMILAR list was, e.g., "**baff** *jab maab* **vame** *zegg yegg* **nidd** *vipe yipe*." Oberauer et al. (2012) found that there was actually better recall of targets, in the right order, for SIMILAR lists – an apparent strengthening effect of phonological similarity. However, in SIMILAR lists, there were also more selective intrusions from the immediately adjacent distractor as well. Oberauer (2009) reported comparable findings for semantic similarity. These results illustrate why it is not possible to say, tout court, that similarity impairs recall of order. It can lead to both improved and degraded performance.

Now let us return to the assumption, mentioned above, that there are strong forward associations between items in a list. How should that apply to language processing and, specifically, the comprehension of ORCs? We hypothesize that it is reanalysis, and not linear order per se, which is the culprit in encoding interference between two DPs in an ORC. Figure 2, below, makes an analogy to the schema introduced in Figure 1. For the two DPs in sentence (3), we assume that it is $DP_1$, *the banker*, which is multiply associated.

(3) <u>The banker</u>$_1$ [ that <u>the barber</u>$_2$ praised$_2$ ] climbed$_1$ the mountain.

In the ordinary course of comprehending an English ORC, the comprehender will first link $DP_1$ to the subject position before reanalyzing it to the object position

(Staub 2010; Wagers and Pendleton 2016).[3] $DP_2$, the trigger for reanalysis, is solely associated to the subject position. Extending the superposition logic of Farrell and Lewandowsky (2002) leads to the prediction that $DP_2$'s representation will potentially be enhanced by this interaction. $DP_1$, on the other hand, is liable to be degraded as the features it shares with $DP_2$ remain bound to the subject position.

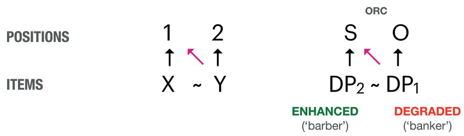



**Figure 2:** Schematic of two DPs in the representation of an RC.

In other words, when the two DPs are sufficiently similar, we can think of the effect of reanalysis as *destructive* to $DP_1$ and *constructive* to $DP_2$.

A recent study by Rich and Wagers (2020) provides some evidence of the destructive effect on $DP_1$ by probing the downstream effects of interference on subject-verb integration (see also Lowder and Gordon 2021). Consider the two sentences in (10). In (10a), an oblique RC (OblRC) is attached to the subject NP whereas in (10b), an SRC is attached. In both cases, there is a second DP inside the RC whose similarity to $DP_1$ is manipulated. In this example, *sword* is highly similar to *knife*, but in other conditions it was replaced by *stick* (medium similarity) or *shirt* (low similarity).

(10) a. The $knife_1$ that the $sword_2$ was placed near _ had been $recently_{Crit}$ sharpened.
b. The $knife_1$ that _ was placed near the $sword_2$ had been $recently_{Crit}$ sharpened.

The critical region was the sequence "had been recently": a series of non-lexical verbs and an adverb which should prompt attachment of the subject, but not provide any

**3** For our purposes in exploring sources of encoding interference, it is enough to assume that subject gaps are (stochastically) inserted early giving rise to (more) SRC parses before ORC parses. That assumption holds for English and, as we have discovered in our initial findings, for Chamorro and Zapotec. There are many hypotheses about the source of this preference, and whether it is universal or not (see discussion of alternatives in Wagers, Borja, and Chung 2018; and for some evidence from object-before-subject languages, see Yasunaga et al. 2015). What is critical for us is merely that there are effectively ordered analyses: if ORCs tend to be hypothesized before SRCs in some language, then it may be that, in that language, the deleterious effects of similarity affect the SRC parse and not the ORC parse. Likewise, the language may provide morphological cues that can effectively guide the comprehender to the correct analysis early, such as case on the two DPs.

discriminative retrieval cues about its contents. This design allows for the detection of a "damaged" $DP_1$ because any difficulty must stem from inherent properties of its representation, and not the (un)selectivity of the retrieval cues. In two self-paced reading experiments, Rich and Wagers (2020) found that the downstream reading times in the critical region were indeed function of $DP_1/DP_2$ similarity. Participants took longer to read the critical region when $DP_2$ was most similar to $DP_1$, as in *knife ~ sword* but not *knife ~ shirt*. However, this was only the case when reanalysis was implicated, i.e., in OblRC conditions like (10a) but not in SRC conditions like (10b). Thus encoding interference doesn't just affect the processing of the RC itself, but it has longer-term consequences for the RC head noun.

The reanalysis hypothesis leads to the prediction that encoding interference should not be the exclusive province of N – N – V orders. Any scenario where the two Ns interact via reanalysis should provide the setting for interference. In the next section, we turn to the ambiguous N – V – N orders of verb-initial languages like Chamorro and Zapotec to assess the plausibility of that prediction.

# 3 Encoding interference effects in verb-initial language processing

In transitive verb-initial clauses, head-initial RCs can create the order N – V – N. Depending on other constraints of the grammar, that order can be ambiguous between an SRC and an ORC. This provides a setting to test the reanalysis hypothesis. An RC with N – V – N order provides good positional cues to distinguish the first and second nouns, both of which occur along the "edges" of the sequence. If encoding interference has its deleterious effect only when it is difficult to recall the linear order of the two nouns, then we shouldn't (for that reason) predict any $DP_1/DP_2$ similarity interactions. If, on the other hand, the relevant configuration is induced by reanalysis, and reanalysis can occur with N – V – N RCs, then similarity-based encoding interference should be observable. In two recent studies on the verb-initial languages Chamorro (Wagers, Borja, and Chung 2018) and Santiago Laxopa Zapotec (Sasaki et al. 2022), we find just that: that the accessibility of the ORC parse in N – V – N orders depends on the similarity of the two nouns.

## 3.1 Chamorro

Chamorro is an Austronesian language of the Mariana Islands. It has relative clauses in two head-RC orders. The N – V – N configuration is found in head-initial relative clauses, where it is ambiguous between an SRC and ORC interpretation. (11a) gives an example of an ambiguous head-initial RC.

| | | | | |
|---|---|---|---|---|
| (11) | Ågang | atyu na biha$_1$ | i ha papaini | i palåo'an$_2$ |
| | call | that L old.lady | D is combing | the woman |

'Call that old lady who _ is combing the woman' *or*
'Call that old lady who the woman is combing _'

The ambiguity of head-initial RCs has been confirmed in fieldwork and in naturally occurring examples. However, despite their ambiguity, Wagers, Borja, and Chung (2018) found that comprehenders overwhelmingly preferred to interpret them as SRCs. In two picture-matching studies, the SRC interpretation rate for sentences like (11) was 94% (Exp. 1) and 97% (Exp. 2).

It is also possible to construct RCs with the N – V – N word order that are unambiguous, just like English RCs are. Chamorro has a system of *wh*-agreement wherein the verb's inflection signals the position of the gap (Chung 1998). The infix -*um*- signals a subject gap, as in (12); the infix -*in*-, with suffixal agreement, signals an object gap, as in (13).

| | | | | |
|---|---|---|---|---|
| (12) | Ågang | atyu na biha$_1$ | i p*um*apaini | i palåo'an$_2$ |
| | call | that L old.lady | D WH[SUBJ].combing | the woman |

'Call that old lady who _ is combing the woman' ~~*or*~~
~~'Call that old lady who the woman is combing _'~~

| | | | | |
|---|---|---|---|---|
| (13) | Ågang | atyu na biha$_1$ | i p*in*apaine-*ña* | i palåo'an$_2$ |
| | call | that L old.lady | D WH[OBJ].combing | the woman |

~~'Call that old lady who _ is combing the woman'~~ ~~*or*~~
'Call that old lady who the woman is combing _'

Even in unambiguous RCs, Wagers, Borja, and Chung (2018) found a strong advantage for SRCs. Participants made errors on unambiguous SRC sentences like (12) very infrequently, in only 2% of trials. However, they made many more errors in unambiguous ORC sentences like (13), in 22% of trials – a difference of an order of magnitude.

Chamorro, despite its rich morphology and its flexible word order, shows the familiar SRC > ORC advantage. However, just as in English, the advantage can be neutralized be making the two DPs less similar. One way to do this is by using a null-headed RC. (14) illustrates a null-headed RC, which includes the demonstrative *atyu* but no pronounced N. Like (11), (14) has both SRC and ORC interpretations.

(14) Ågang atyu$_1$ i ha papaini i palåo'an$_2$
call that [one] D is combing the woman
'Call that one who _ is combing the woman' *or*
'Call that one who the woman is combing _'

When participants listened to sentences like (14), the ORC interpretation suddenly became much more accessible. The SRC interpretation rate dropped to 68% for null-headed RCs. In unambiguous sentences with wh-agreement, the error rate asymmetry was neutralized. Participants made 2% errors on subject wh-agreement trials, the same as in the head-initial RCs with two full DPs; but they only made 6% errors on object wh-agreement trials, as long as one DP was null and the other was overt.

Chamorro thus shows an analogous contrast in (14) v. (11) to what English shows in (4) v. (3): an ORC parse is difficult to attain when two full DPs are involved, but not when the two DPs are distinct. In both instances, the distinction is conceivably one of reference, of phonology or, perhaps, of the size of the DP. It is not as clear in Chamorro, as it is in English, that reanalysis is implicated. Wagers, Borja, and Chung (2018) provide one argument for reanalysis by contrasting head-initial RCs with head-final RCs, which have V – N – N orders. (15) illustrates a head-final RC, which demonstrates the same SRC/ORC ambiguity as (11) and (14).

(15) Ågang atyu i ha papaini i palåo'an$_2$ na biha$_1$
call that D is combing the woman L the old.woman
'Call that old lady who _ is combing the woman' *or*
'Call that old lady who the woman is combing _'

In the same picture-matching experiments, head-final RCs proved to be much more neutral with respect to SRC/ORC interpretation. Participants interpreted ambiguous sentences like (15), with two full DPs, as SRCs 43% of the time in Exp. 1 and 54% of the time in Exp. 2. When head-final RCs were made unambiguous by wh-agreement, the error rates were essentially the same for subject wh-agreement (12%) and object wh-agreement (13%).

Wagers, Borja, and Chung (2018) argued that comprehenders tend to insert subject gaps whenever an RC is detected (all else equal). In head-initial RCs, comprehenders commit to this analysis early. This makes ORC parses harder to achieve, even when later morphology signals them unambiguously. In head-final RCs, on the other hand, the actual identity of the head remains undetermined until much later, which must somehow render the SRC parse more defeasible. In present terms, we could see the effects as ones based in similarity: in any ORC reanalysis, the RC-internal DP must be bound to the subject position which was first linked to the RC-external DP. When both DPs are full DPs, and thus highly similar to one another, this reanalysis is hard. When the the RC-external DP has fewer features to confuse it with the RC-internal DP (because it is either null or has not yet been encountered), then this reanalysis is easy.

In Chamorro, similarity was defined in terms of full versus reduced DPs. Now we turn to Zapotec, where similarity can be defined based on animacy-based noun classes.

## 3.2 Zapotec

Santiago Laxopa Zapotec (SLZ) is an Oto-Manguean language spoken in the Sierra Norte of Oaxaca. It is a rigidly V – S – O language but ambiguity arises in RCs and other movement constructions which create N – V – N configurations (Adler et al. 2018). (16) illustrates that, in a verb-initial clause, the arguments are interpreted rigidly as S > O. But, compare with (17), which incorporates an RC. Here both SRC and ORC interpretations are possible.

| | | | | | |
|---|---|---|---|---|---|
| (16) | Tsyill | bene’ | nu’ulhe=nh$_1$ | bene’ | xyage’=nh$_2$ |
| | Pinch | CL | woman=DEF | CL | man=DEF |

“The woman is pinching the man”
(NOT “The man is pinching the woman.”)

| | | | | | | |
|---|---|---|---|---|---|---|
| (17) | Shlhe’eyd=a’ | bene’ | nu’ulhe=nh$_1$ | tsyill | bene’ | xyage’=nh$_2$ |
| | see=1SG | CL | woman=DEF | pinch | CL | man=DEF |

“I see the woman that _ is pinching the man” *or*
“I see the woman that the man is pinching _”

It is possible to eliminate this ambiguity by using resumptive pronouns (RPs). For example, a subject RP can cliticize on the verb and force the SRC interpretation, as in (18); or an object RP can occur after DP$_2$ and force the ORC interpretation, as in (19)

| | | | | | | |
|---|---|---|---|---|---|---|
| (18) | Shlhe’eyd=a’ | bene’ | nu’ulhe=nh$_1$ | tsyill=e’ | bene’ | xyage’=nh$_2$ |
| | see=1SG | CL | woman=DEF | pinch=RP | CL | man=DEF |

“I see the woman that she is pinching the man” ~~*or*~~
~~“I see the woman that the man is pinching _”~~

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| (19) | Shlhe’eyd=a’ | bene’ | nu’ulhe=nh$_1$ | tsyill | bene’ | xyage’=nh$_2$ | le’ |
| | see=1SG | CL | woman=DEF | pinch | CL | man=DEF | RP |

~~“I see the woman that _ is pinching the man”~~ ~~*or*~~
“I see the woman that the man is pinching her”

Sasaki et al. (2022) investigated how SLZ comprehenders navigated the ambiguity of sentences like (17) and how they used RPs to disambiguate. In one experiment, SLZ speakers (n = 103) performed a picture-matching task while they listened to stimuli containing RCs. In ambiguous RCs, participants showed an SRC advantage, although a comparatively weak one of only 67%. Strikingly, however, sentences containing object RPs were very often incorrectly parsed, with an error rate of nearly 48%. In contrast, when subject RPs were present, participants mostly correctly selected the SRC picture, with an error rate of only 13%.

The asymmetry in error rates is reminiscent of the findings from wh-agreement conditions in Chamorro and it points to the familiar SRC > ORC advantage existing in SLZ. Like the head-initial Chamorro RCs, the SLZ RCs that Sasaki et al. tested in their first experiment contained two highly similar full DPs. In a second experiment (n = 105), they drew upon the existence of several noun classes, or genders, in SLZ to manipulate the similarity between the two DPs. SLZ makes animacy-based distinctions to generate these classes. It distinguishes inanimate referents, animals, non-elder humans and elder humans in a four-way split. For example, *bi'i nu'ule'nh,* "young girl", belongs to the non-elder human class and is referred to with non-elder human pronouns: *=ba'* (clitic) and *leba'* (strong). In contrast, *bene' gule'n,* "old person", belongs to the elder human class and is referred to with elder human pronouns: *=(n)e'* (clitic) and *le'* (strong).

In their second experiment, RCs always consisted of *mismatching* DPs, such as an elder human and a non-elder human referent; or a non-elder human and an animal. This led to considerable reductions in the proportion of SRC interpretations for ambiguous RCs: 52% in Exp. 2, compared to 67% in Exp. 1. As well, the error rate for Object RPs was reduced: 31% in Exp. 2, compared to 48% in Exp. 1.[4]

Although this study does not provide a direct comparison between similar and dissimilar DPs, as did the Chamorro experiments reported above, it does provide suggestive evidence of a similarity-based determinant of ORC accessibility. Moreover eye-movement data from Experiment 2 (not reported here) provides more direct evidence that participants are often engaging in reanalysis from an initial SRC parse toward an ORC parse.

**4** This reduction in error rates on Object RPs from Exp. 1 and Exp 2., while considerable, does leave us with the puzzle of why the error rate remains so high for Object RPs (compare to the 13% error rate for Subject RPs in Exp. 1). It may be that the type and position of the pronoun serves as a better or worse cue for the correct parse (independently of similarity of the DPs involved): the Subject RP is an early clitic pronoun that appears before the RC-internal DP, whereas the Object RP is a later independent pronoun that appears after the RC-internal DP.

# 4 Conclusions

We have argued that the relative inaccessibility of ORC parses is, in part, a function of encoding interference. In turn, we have argued that encoding interference is not itself dependent on the strict linear contiguity of two DPs in N – N – V word orders. Instead we appealed to the reanalysis that typically occurs as comprehenders transition from an SRC parse to an ORC parse. Reanalysis provides the environment in which two DPs interact and interfere with one another. In two verb-initial languages, we have shown that RCs with N – V – N word orders show the same hallmarks of encoding interference that English RCs do, which would be expected if those word orders can also involve reanalysis.

In linking encoding interference to reanalysis, we make an untested prediction about the quality of the $DP_2$ encoding: i.e., that of the eventual subject phrase. Some accounts of encoding interference, like Farrell and Lewandowky (2002), lead us to suspect that interference is sometimes *constructive*. The shared features between two DPs can strengthen the representation of their common associate, the subject position. This is a challenging prediction to test, because the relationship between $DP_2$ and the verb is established relatively quickly. One possibility is to test participants later in a trial, with comprehension questions that probe the thematic role of $DP_2$. Another possibility is to use a more indirect method. For example, agreement attraction designs (Wagers, Lau, and Phillips 2009) could be used to assess whether $DP_2$ is more resistant to attraction, when it is more similar to $DP_1$. Thus, we might expect (20) to show less agreement attraction than (21), because the representation of the (grammatical) controller of agreement, $DP_2$, is more likely to strengthened via constructive interference in (20) compared to (21).

(20) There was a professor$_1$ who the student by the bushes$_2$ was/were waving to . . .

(21) There was someone$_1$ who the student by the bushes$_2$ was/were waving to . . .

Finally, we recommend further research on how ambiguous N – V – N word orders are parsed to more firmly establish the presence of reanalysis.

**Acknowledgments** I thank the audience at IJPCP 2021 for their helpful feedback on this project. I warmly acknowledge my collaborators Sandra Chung, Manuel F. Borja, Maziar Toosarvandani, Ivy Sichel, and Stephanie Rich for helping me realize the research reported in this chapter, and for many illuminating conversations. Finally, I am grateful to the two anonymous reviewers who carefully read the original manuscript and provided incisive questions and observations, which I've tried to incorporate in this version. All errors are my own.

This work is supported by grants from the U.S. National Science Foundation to UC Santa Cruz (BCS #2019804: M. Toosarvandani, PI; I. Sichel and M. Wagers, Co-PIs; BCS #1251429: M. Wagers, PI; S. Chung, Co-PI) and, at UC Santa Cruz, from the Academic Senate Committee on Research, the Vice Chancellor for Research, and the Humanities Institute. I am grateful for their support.

## References

Adler, Jeff, Steven Foley, Jed Pizarro-Guevara, Kelsey Sasaki, & Maziar Toosarvandani. 2018. The derivation of verb initiality in Santiago Laxopa Zapotec. In Jason Merchant, Line Mikkelsen, Deniz Rudin, & Kelsey Sasaki (eds.), *A Reasonable Way to Proceed: Essays in Honor of Jim McCloskey*, 31–49. https://escholarship.org/uc/item/7z29n70x

Ariel, Mira. 1990. *Accessing Noun-Phrase Antecedents*. London: Routledge.

Arnett, Nathan & Matthew Wagers. 2017. Subject encodings and retrieval interference. *Journal of Memory and Language* 93. 22–54.

Badecker, William & Richard Lewis. 2007. A new theory and computational model of working memory in sentence production: agreement errors as failures of cue-based retrieval. Paper presented at the 20th Annual Conference on Human Sentence Processing, University of California, San Diego.

Bever, Thomas G. 1974. The ascent of the specious, or there's a lot we don't know about mirrors. In David Cohen (ed.), *Explaining Linguistic Phenomena*, 173–200. New York: Wiley.

Chung, Sandra. 1998. *The Design of Agreement: Evidence from Chamorro*. Chicago: University of Chicago Press.

Dillon, Brian, Alan Mishler, Shayne Sloggett, & Colin Phillips. 2013. Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language* 69. 85–103.

Farrell, Simon & Stephan Lewandowsky. 2002. An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin & Review* 9. 59–79.

Gordon, Peter C., Randall Hendrick, & Marcus Johnson. 2001. Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27. 1411–1423.

Gordon, Peter C., Randall Hendrick, & Marcus Johnson. 2004. Effects of noun phrase type on sentence complexity. *Journal of Memory and Language* 51. 97–114.

Gordon, Peter C., Randall Hendrick, Marcus Johnson, & Yoonhyoung Lee. 2006. Similarity-based interference during language comprehension: Evidence from eye tracking during reading. *Journal of Experimental psychology: Learning, Memory, and Cognition* 32. 1304–1321.

Gordon, Peter C., Randall Hendrick, & William H. Levine. 2002. Memory-load interference in syntactic processing. *Psychological Science* 13. 425–430.

Lewandowsky, Stephan & Murdock Jr., Bennet B. (1989). Memory for serial order. *Psychological Review* 96. 25–57.

Lewis, Richard L., Shravan Vasishth, & Julie A. Van Dyke. 2006. Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences* 10. 447–454.

Lowder, Matthew W. & Peter C. Gordon. 2021. Relative Clause Effects at the Matrix Verb Depend on Type of Intervening Material. *Cognitive Science* 45. e13039.

Miller, George A. & Noam Chomsky. 1963. Finitary models of language users. In R. Duncan Luce, Robert R. Bush, & Eugene Galanter (eds.), *Handbook of Mathematical Psychology, Vol 2*, 419–491. New York: Wiley.

Minkoff, Seth. 2000. Animacy hierarchies and sentence processing. In Andrew Carnie & Eithne Guilfoyle (eds.), *The Syntax of Verb-Initial Languages*, 201–212. Oxford: Oxford University Press.

Murdock Jr., Bennet B. & Walter Vom Saal. (1967). Transpositions in short-term memory. *Journal of Experimental Psychology* 74. 137–143.

Nairne, James S. 1990. A feature model of immediate memory. *Memory & Cognition* 18. 251–269.

Oberauer, Klaus. 2009. Interference between storage and processing in working memory: Feature overwriting, not similarity-based competition. *Memory & Cognition* 37. 346–357.

Oberauer, Klaus, Simon Farrell, Christopher Jarrold, Kazimir Pasiecznik, & Martin Greaves. 2012. Interference between maintenance and processing in working memory: The effect of item–distractor similarity in complex span. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 38. 665–685.

Oberauer, Klaus & Reinhold Kliegl. 2006. A formal model of capacity limits in working memory. *Journal of Memory and Language* 55. 601–626.

Omaki, Akira, Ellen F. Lau, Imogen Davidson White, Myles L. Dakan, Aaron Apple, & Colin Phillips. 2015. Hyper-active gap filling. *Frontiers in Psychology* 6. 384.

Polyn, Sean M., Kenneth A. Norman, & Michael J. Kahana. 2009. A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review* 116. 129–156.

Rich, Stephanie & Matt Wagers. 2020. Semantic similarity and temporal contiguity in subject-verb dependency processing. Paper presented at the 33rd Annual Conference on Human Sentence Processing, University of Massachusetts, Amherst.

Sasaki, Kelsey, Steven Foley, Jed Pizarro-Guevara, Fe Silva-Robles, Maziar Toosarvandani, & Matthew Wagers. 2022. Evidence for a universal parsing principle in Santiago Laxopa Zapotec. UC Santa Cruz manuscript.

Silverstein, Michael. 1976. Hierarch of features and ergativity. In R.M.W. Dixon (ed.), *Grammatical Categories in Australian Languages*, 113–171. Atlantic Highlands, NJ: Humanities Press.

Staub, Adrian. 2010. Eye movements and processing difficulty in object relative clauses. *Cognition* 116. 71–86.

Van Dyke, Julie A. & Richard L. Lewis. 2003. Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language* 49. 285–316.

Villata, Sandra, Whitney Tabor, & Julie Franck. 2018. Encoding and retrieval interference in sentence comprehension: Evidence from agreement. *Frontiers in Psychology* 9. 2.

Wagers, Matthew W., Manuel F. Borja, & Sandra Chung. 2018. Grammatical licensing and relative clause parsing in a flexible word-order language. *Cognition* 178. 207–221.

Wagers, Matthew W., Ellen Lau, & Colin Phillips. 2009. Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language* 61. 206–237.

Wagers, Matthew W. & Emily Pendleton. 2016. Structuring expectation: Licensing animacy in relative clause comprehension. In Kyeong-min Kim, Pocholo Umbal, Trevor Block, Queenie Chan, Tanie Cheng, Kelli Finney, Mara Katz, Sophie Nickel-Thompson, & Lisa Shorten (eds.), *Proceedings of the 33rd West Coast Conference on Formal Linguistics*, 29–46. Somerville, MA: Cascadilla Press.

Warren, Tessa & Edward Gibson. 2002. The influence of referential processing on sentence complexity. *Cognition* 85. 79–112.

Warren, Tessa & Edward Gibson. 2005. Effects of NP type in reading cleft sentences in English. *Language and Cognitive Processes* 20. 751–767.
Yasunaga, Daichi, Masataka Yano, Yoshiho Yasugi, & Masatoshi Koizumi. (2015). Is the subject-before-object preference universal? An event-related potential study in the Kaqchikel Mayan language. *Language, Cognition and Neuroscience* 30(9). 1209–1229.

Satoshi Shioiri, Rumi Tokunaga, and Ichiro Kuriki

# Chapter 4
# Cross-cultural comparison of lexical partitioning of color space

## 1 Introduction

Accumulation of information is essential for human knowledge production, and information technology has accelerated the speed of data accumulation (Smil 2019). The increase in quantity of information, however, does not promise high-quality knowledge production and possibly causes problems (Muraoka et al. 2017; Shioiri et al. 2021). One critical problem in information usage is information overload – that is, deterioration of productivity due to excessive information. It is well-known that decision accuracy decreases with increase in amount of information beyond a certain point, although it increases in the beginning. One approach to solve the problem of information overload could be to select only high-value information. For example, for regular communication, text data are often highly valuable despite their much smaller data size compared to videos or images. However, although most people will likely concur that text data are significant for communication, for the testimony of civilization, and as a bearer of culture, there is no method to compare the value of information conveyed via text and other mediums, such as videos and images. Therefore, it is necessary to develop knowledge of information values and technologies to evaluate them. In this study we consider the value of words as information tools, focusing on color terms and color perception. Color-naming is one of the simplest cases for language communications and accuracy of communications can be evaluated based on experiments to reveal the link between perceived colors and color terms.

The number of color terms varies among different languages, and most people can discriminate a much larger number of colors than terms used for colors (Lindsey and Brown 2009). As a communication tool, the number of color terms limits the variation of colors that can be transmitted between people via words. Although certain languages of technologically less advanced regions have fewer color terms (Berlin and Kay 1969), that does not pose a problem as long as there is no need to refer to many different colors using words. Nonetheless, partitioning of color space

**Acknowledgments:** This work was supported by a Cooperative Research Project (International type; No. H26/A25) of the Research Institute of Electrical Communication, Tohoku University, JP15H03460 to IK, and Yotta Informatics Project by MEXT, Japan.


https://doi.org/10.1515/9783110778946-004

is crucial in communication. A color term can be used for communication only when the term indicates the same region in color space among different people, or at least between the sender and receiver of the information – that is, partitioning of color space for each color term is a significant factor for color communications. Usage of color terms appears to be a purely linguistic issue; however, it is necessary to investigate the knowledge of color perception.

Investigating a number of languages, Berlin and Kay (1969) proposed two conjectures: (i) there exists a limited set of "universal" categories from which all languages draw their color lexicons (basic color terms, BCTs), and (ii) languages "evolve" by adding color names in a relatively fixed sequence. This has been supported by many studies (Lindsey and Brown 2006; Marlowe et al. 2011). This universality is based, at least partially, on physiological processes common for human species, which should contribute to universal color categories. Color vision starts with three types of photoreceptors for day-time vision: long-wavelength, middle-wavelength and short-wavelength sensitive cones, except for people with color deficiencies, such as dichromats in classical color vision models. Additionally, there is little known difference among people from different cultures in terms of following visual processes. As it is likely that the borders of color regions indicated by color terms are based on these physiological processes, similarity in categorization of colors, thus, in color terms, is expected among different cultures, at least at early visual stages. The common visual processes likely influence developments of color terms.

The linkage between perceptions and concepts/terms to express the perceptions is a significant issues in communication of sensory information. Describing what you see in words requires translation of information from visual perception to words. In the case of color, interactions between color terms and color perception is well-known as the Stroop effect (Stroop 1935). This effect results in longer response time to name ink color when there is a mismatch between the name of a color and the name of the color of ink used to spell that color – for example, the word "red" printed in blue ink. The Japanese experience slight confusion in terms of colors of traffic lights. The color term "blue" or "ao" is typically used for the go sign lamp, which is often considered green as color perception. If the green light were used for stop, shouting, "stop, the light is blue" could make the reaction slower compared to shouting, "stop, the light is green," as a consequence of the Stroop effect. It is fortunate that red is used for stopping and there are few similar cases in everyday life. The intimacy between colors and color terms is, in fact, beneficial for communication. There are, however, still essential problems related to color and color terms in communication. Categorization of colors and usages of color terms are useful when there is little confusion in identification of objects by color, as in the case there is only one blue pen in the scene and a person is asked to bring the blue pen; whereas, in case there are multiple blue pens, identification by color will not work. Of course, there is no reason to use the color term in such condition; other features could be added to the description to identify the object. However, the colors categorized as blue may differ among individuals, and a person may think there is only one blue pen, whereas another may imagine a multiplicity of blue pens (See Figure 1).


**Figure 1:** Color is useful to identify an object by words. Using a color is intuitive and one of the best ways to identify an object (top). However, using a color term may not be sufficient to identify an object by more than one person when there are many items with similar colors. The same color can be named differently by different people: one may call it “blue,” another may call it “green” – especially in the case of colors around borders (bottom).

Categorization of color in terms of perception indicates which region in color space a color term indicates. In this study, we introduce color-naming experiments to investigate lexical partitioning of color space and compare the results among different languages: Japanese (Kuriki et al. 2017), Mandarin Chinese spoken in Taiwan – or Taiwanese Chinese in short (Hsieh et al. 2020) – and American English (Lindsey and Brown 2014). This is to confirm the similarity reported between Japanese and American English (Uchikawa, Uchikawa, and Boynton 1989) and to expand the comparisons with Taiwanese Chinese, using advanced techniques of data analyses. We used *k*-mean clustering for categorization, and color name data are summarized as binary vectors (Lindsey and Brown 2006, 2009). These methods realized the objective judgments with mathematical criterion for number of color terms, overcoming individual variation of color term usages within each language group.

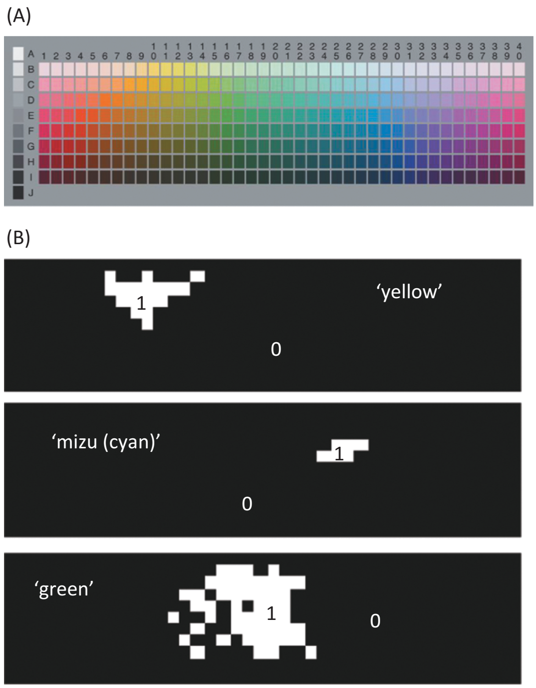



**Figure 2:** (A) The 330 color chips used in the present study. (B) Example vectors of three color terms. Each binary vector was derived from each participant's color category. Assign "1" for color chips that were named with a color term and "0" for others; the resultant values are shown for three colors, being arranged to correspond to the map of stimulus color chips. Word label on each category was removed afterward.

# 2 Experiment

We conducted experiments in Japanese (Kuriki et al. 2017) and Taiwanese Chinese (Hsieh et al. 2020), where experimental procedures followed studies by Lindsey and Brown (2014). Here, we describe the experiment in Japanese. The experiment in Taiwanese Chinese used essentially the same procedure and stimuli. The same experimental procedure with the same stimuli in different languages facilitates direct comparison of results among different languages. Appropriate control of color stimuli is difficult, and such attempts are crucial for detailed analysis of color-related studies, including that of color terms.

## 2.1 Methods

### Informants

Fifty-seven native Japanese speakers (30 male and 27 female) participated in this study. All informants had normal or corrected-to-normal visual acuity, and their color vision was confirmed to be normal using Ishihara pseudoisochromatic plates.

### Color samples and illuminant

The color samples, illuminant, and background color papers were similar to those used in the World Color Survey (WCS). The 330 color chips used in the present study were from the *Munsell Book of Color* glossy (X-Rite, Inc.). The chips were chosen to match the WCS samples with respect to hue, chroma, and value (although the WCS samples were from the matte edition) (Figure 2 (A)). Each chip was mounted on a cardboard square 5 cm x 5 cm, covered with gray matte paper approximating N5/ (in Munsell notation). Experiment was performed under an illuminance of 2,713 lx with color temperature of approximate 6000 K.

### Procedure

Informants used a single, monolexemic color term to name each sample. They were not allowed to use compound color terms such as *ki-midori* (*yellow-green*) or modifier words such as *usu-murasaki* (*pale purple*). However, they could use the name of a substance if they felt it was generally accepted as a representation of a color and could be generalized to identify the color of any type of object. Similar restrictions were applied for other languages for comparison (Hsieh et al. 2020; Lindsey and Brown 2014).

### Cluster analysis

Analysis of the color-naming data was performed in two steps, both of which involved *k*-means cluster analysis. The first step was to extract two entities from the raw data sets: (a) an estimate of the number of chromatic color categories in Japanese and (b) the extent of each of these categories in color space. The second step was to analyze the color-naming patterns (motifs), that is, patterns of color categories in color space used by informants. The analysis estimates the number of motifs and determines their categorical structures (partitioning pattern of color space by different color terms). We used custom programs, which had been previously used by researchers (Brown, Isse, and Lindsey 2016; Lindsey and Brown 2006, 2009, 2014). Here we present an overview of the methodology.

## 2.2 Clustering for color categories

First, *k*-means cluster analysis was performed to classify feature vectors representing the sets of color samples associated with each *chromatic* color term deployed by each informant (Figure 2 (B)). A chromatic color term was defined as a term used by an informant to name one or more of the 320 chromatic colors in the WCS chart, but never used by that informant to name any of the 10 achromatic colors (achromatic color terms were handled separately). Each chromatic-term feature vector comprised 320 elements, each of which was set to a value of 1 or 0, depending on whether (or not) the chromatic color term was used by the informant to name the WCS color sample. The resulting 828 binary feature vectors obtained from the chromatic words used by all informants (i.e., the sum of the chromatic color terms used by each informant) were then sorted into *k* clusters by *k*-means cluster analysis. This first *k*-means cluster classifies responses solely on the basis of how color terms are deployed across the 320 WCS chromatic colors, as embodied in the patterns of color-term deployment encoded in the binary feature vectors, without regard for the actual terms used by the informant. In Japanese, for example, *k*-means analysis showed that *sora* (sky) and *mizu* (water) were synonymous.

We determined the number of clusters as $k_{L,opt}$, using an index called gap statistic (Tibshirani, Walther, and Hastie 2001). After performing *k*-means analyses for values of *k* from 1 to 25, we performed the gap-statistic analysis on these 25 separate cluster formed by comparing the tightness of clustering of the data to the tightness of the same clustering analysis of reference null distributions. By design, the expected value of $k_{L,opt}$ for a reference distribution is 1. Thus, as the value of $k$ increases from 1 to $k_{L,opt}$, the tightness of clustering of the data is expected to improve relative to that obtained for the reference null distributions. Beyond $k_{L,opt}$, increasing $k$ should not lead to any further improvement. The number of clusters that explains the data, $k_{L,opt}$, was determined via a step-by-step computational framework. The values of $k_{L,opt}$ are 16 for Japanese, 8 for Taiwanese Chinese, and 17 for American English as candidates of new BCTs (see later).

We chose common color terms in each language as names for each of the clusters: *aka* (*red*), *ao* (*blue*), *ki* (*yellow*), and so on (see later for details). For example, if the feature vector for the color term *moegi*, as deployed by a particular informant, falls into the *midori* (*green*) cluster, then we say that "*moegi* glosses to *midori*."

## 2.3 Clustering for motifs

We next performed a second *k*-means/gap-statistic analysis to determine $k_{M,opt}$, the number of statistically significant motifs, and the structures of these motifs. In the case of Japanese, the analysis clustered 57 motif feature vectors, each vector representing all 330 color-naming responses of a single Japanese informant. Each feature vector comprised 19 elements corresponding to the 19 color categories: 16 chromatic categories derived from the first cluster analysis plus three achromatic color categories (white, gray, and black). Each of the 19 elements was assigned a value between 0.0 and 1.0, which was the proportion of samples (out of 330 WCS samples) a given informant named with the glossed color term. For example, if the informant used the word *aka* to name three samples, the value of the *aka* element in that informant's motif feature vector would receive the number 3/330 = 0.0091. The 57 feature vectors were then sorted into *k* clusters using the *k*-means method. Gap-statistic analysis was again performed to determine $k_{M,opt}$ based on the *k*-means results obtained for $k = 1, \ldots, 5$. Glossed color-naming patterns of individual informants were compiled within motifs, and aggregate results were displayed as consensus diagrams (see later).

## 2.4 Color-term popularity diagrams

The number of informants using each term (the term's "popularity") as a function of the sorted rank order of that term's popularity is analyzed. The frequency of word used usually follows a power law – that is, the logarithms of the frequency with which words occur is linearly related with the logarithm of their rank order. Double-power law behavior – that is, two lines predict well the relationship between logarithms of rank order and logarithms of popularity or frequency – is common in language corpora. Two exponents are thought to divide words in two different sets: (i) a kernel lexicon formed by a certain number of versatile words or a finite number of core words and (ii) an unlimited lexicon for specific communication or the remaining virtually infinite number of non-core words (Ferrer i Cancho and

Solé 2001; Gerlach and Altmann 2013). This analysis is, therefore, used to classify terms into one of the two groups of word sets.

# 3 Results

## 3.1 Histograms of the number of informants and popularity diagrams

We summarized results of three studies (Hsieh et al. 2020; Kuriki et al. 2017; Lindsey and Brown 2014). Figure 3 shows histograms of the number of informants with a number of color terms used by informants in the Japanese, Taiwanese Chinese, and American English experiments. The largest number of color terms used by an informant is 50 each for American English and Japanese, and 15 for Taiwanese Chinese, and the average, as well as median and mode, number of color terms used was the largest for American English speakers, followed by Japanese and Taiwanese Chinese speakers. The total number of color terms used also varied: 122 from 51 American English speakers, 93 from 57 Japanese speakers, and 23 from 41 Taiwanese Chinese speakers.

Note that the basic idea of the analyses here allow us not to classify color category simply by color terms. Even when different people use different terms to name the same color we did not interpret the terms referring different color categories. If one term is used by an informant and the other is used by another for the same color region, their color vectors are the same, which indicates that the two terms can be regarded to have the same meaning, at least for perceiving colors. Conversely, if an informant uses two terms for different color regions with a border in between, they could be different color terms for the informant. A clustering analysis for the data from all the informants determines if they are different or not for the informant group. In actual data analysis, however, we used one color term for two colors, if one was the translation of the other. There were three such pairs in Japanese: *momo* and *pink*, *daidai* and *orange*, and *hai* and *gray*.

Figure 4 shows popularity diagrams of color terms for each of the three languages. For all cases, there was a ceiling effect, as the BCTs were used by all the informants and are, therefore, perfectly fitted by a constant function. Popularity decreases with ranks larger than 11. If two lines fit the data, these lines divide words in two different sets: a finite number of core words and the remaining virtually infinite number of non-core words, as words popularity in general. In addition to lower popularity, the slope of the fitted line is steeper for non-core words. For American English results, two sets of words: core and non-core word groups are

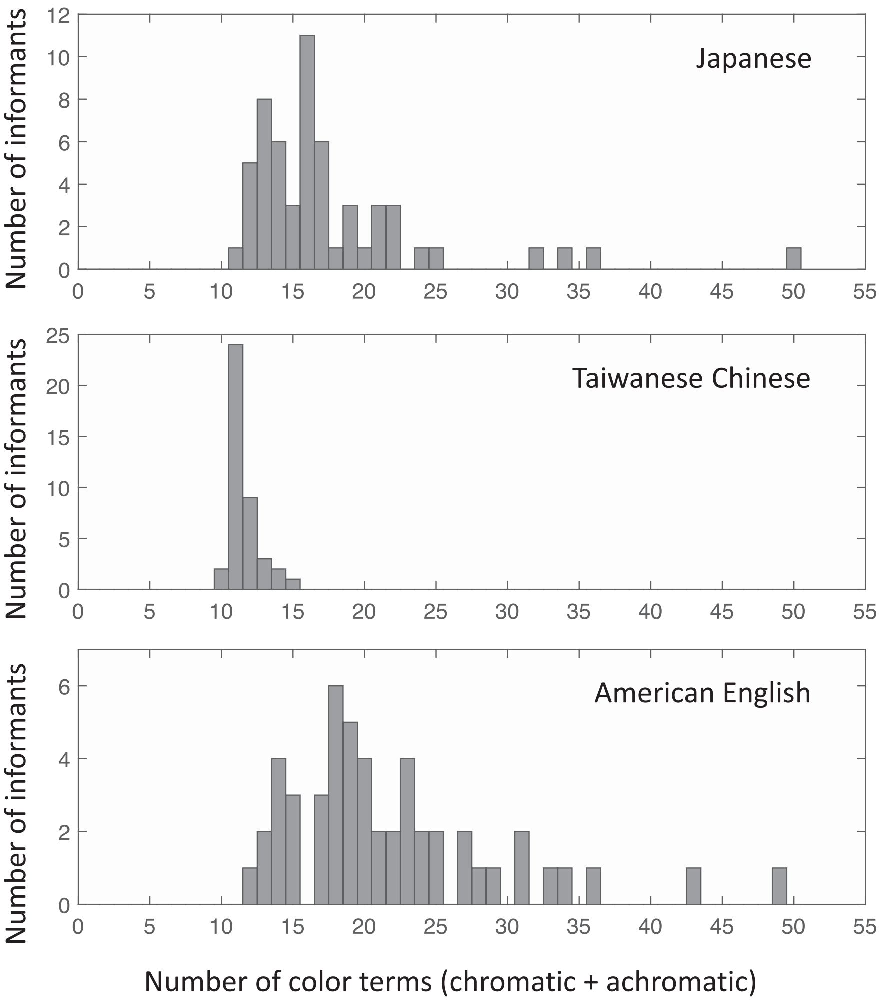



**Figure 3:** Histograms of informants for number of color terms used for each of the three languages studied.

suggested (separated by the arrow at around a log rand of 1.2). The usage of color terms is similar to words of the rest of the categories. Although no such effect is seen for Japanese and Taiwanese, there is a gap between *oudo* and *kon* in the Japanese data, which may correspond to the difference in the usage of terms as slope difference. It is possible that the number of color terms for Japanese and Taiwanese experiments is too small to classify into two groups with difference usages.

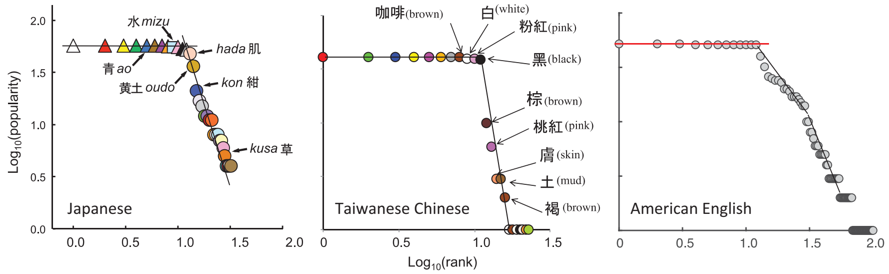



**Figure 4:** Popularity diagrams for Japanese, Taiwanese Chinese, and American English.

## 3.2 Cluster analysis

Consensus maps for the chromatic color terms revealed by the *k*-means with the gap statistic are shown in Figure 5 for each language. Numbers of chromatic color terms to satisfy a criterion for consensus maps are 16, 8, and 17 for Japanese, Taiwanese Chinese, and American English. The color terms determined by the clustering are candidates for new BCTs, and we call them clustered color terms (CCTs). CCTs include several non-BCTs in addition to Berlin and Kay's (1969) BCTs for Japanese and American English, whereas the number of CCTs is identical to that of BCTs for Taiwanese Chinese. CCTs of non-BCTs are *mizu* (light blue), *hada* (peach), *oudo* (light brown), *kon* (dark blue), *cream, matcha* (yellow green), *enji* (maroon), and *yamabuki* (gold) for Japanese, and *peach, teal, lavender, maroon, gold, beige, magenta, lime, and olive* for American English. There are similarities and differences in CCTs. The number of non-BCTs are different among languages, although a few of them are similar between Japanese and American English, such as *hada* and peach or *enji* and maroon. More detailed comparisons are described later in Discussion, with advantages of analyzing data obtained with the same stimuli and procedures.

## 3.3 Motif analysis

Two color-naming motifs were identified in Japanese (Figure 6). Each map in Figure 6 shows color terms used by at least 80% of the informants in each motif. Black indicates less than 80% consensus for the corresponding color patch. Whereas the non-BCTs of *mizu* and *hada* appear in both of the Japanese motifs, consensus of only *mizu* is above the 80% threshold at several color chips (other visualization techniques show the *hada* regions, although small, in the motifs). The major difference between the two motifs is the size in color region for *mizu*. Motif 1 shows

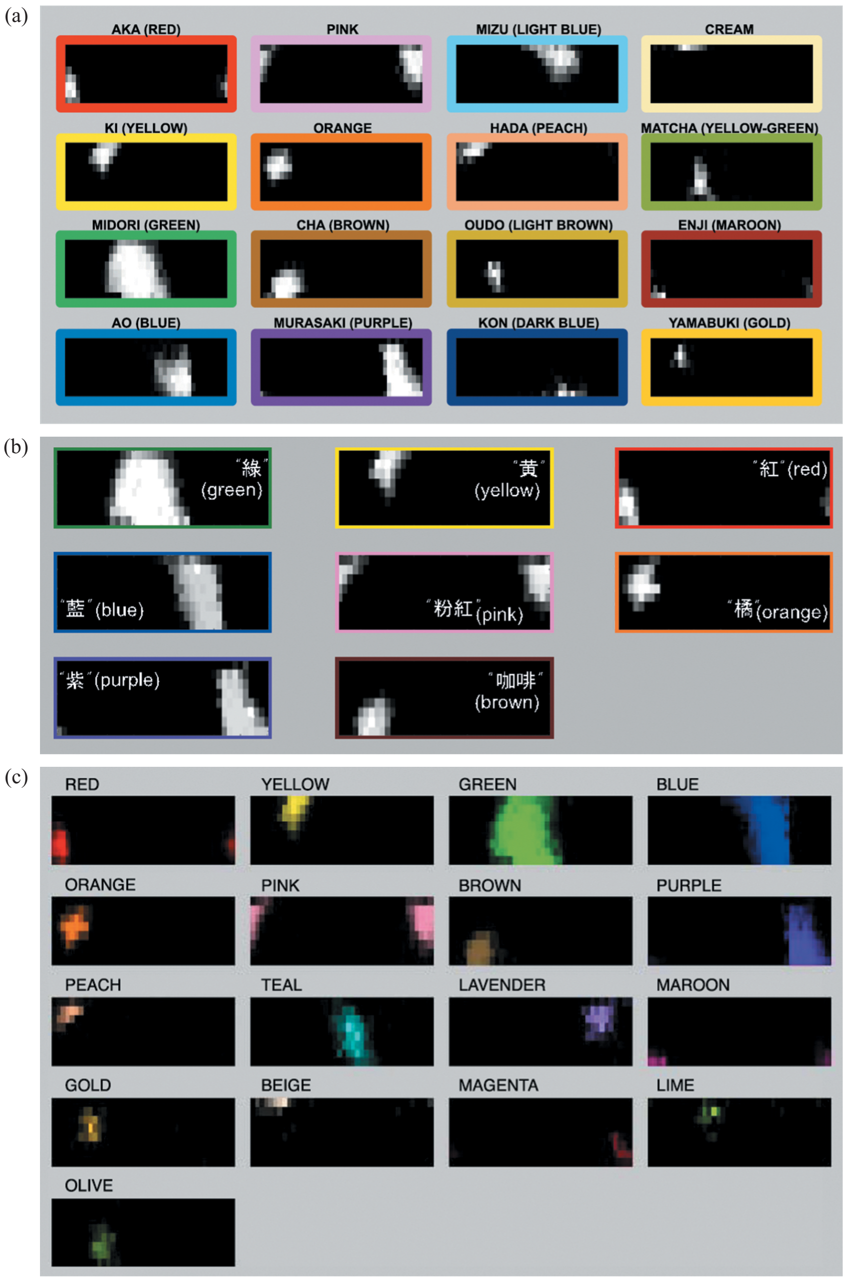



**Figure 5:** Consensus maps for the chromatic color terms revealed by the *k*-means for Japanese (a), Taiwanese Chinese (b) and American English (c). Note the differences in arrangements among the three figures.

larger region of *mizu* than Motif 2. The number of informants classified to motif with larger *mizu* region is 14 out of 51. Only one motif was identified in the Taiwanese Chinese as can be speculated from the smaller number of color categories identified by *k*-means than in Japanese. As the number is eight, which is consistent with chromatic BCTs, no difference is expected in dividing the map into more than one with smaller numbers of color terms. The results of motif analysis for American English showed two motifs with clear difference in the blue region: one with a region for teal and the other without it (Lindsey and Brown 2014). The American English motif with teal also has maroon, peach, and lavender. The variation of consensus color map (different motifs) suggests that there are sub-populations of a language community. Both Japanese and American English show that one motif appears to be more variant from the common 11 BCTs map, which suggest that the number of BCTs is increasing in the US and Japan. This is supported by the comparison between two Japanese color term studies from 1987 (Uchikawa and Boynton 1987) and 2017 (Kuriki et al. 2017). Although both studies found frequent use of *muzu* for light blue, Uchikawa and Boynton (1987) concluded that *mizu* is not a basic term. They did so primarily because 77% of the color chips named *mizu* by a few of the informants were named *ao* (blue) by other informants, and also 80% of the chips named *sora* (sky) were sometimes called *ao*. Based on this fact, they judged that *mizu* and *sora* were subsets of ao. In contrast, Kuriki et al. (2017) that concluded *mizu* is a basic color term, that is different from *ao* (blue), showing the number of the color chips named both as *mizu* and *ao* by different informants (fraction to ones named either *mizu* or *ao*) is much smaller than shown in the study by Uchikawa and Boynton (1987) 30 years ago. Kuriki et al. (2017) suggested that the Japanese color lexicon had somewhat changed in the last 30 years and *mizu* is likely emerging as a new basic color term.

# 4 Discussion

## 4.1 Analysis of pooled data

Cluster analysis for color terms and motifs were performed for pooled data of three experiments in Japanese, Taiwanese Chinese, and American English (Hsieh et al. 2020). This can be performed by analyzing color-naming data as binary vectors, ignoring actual color names. Figure 7 (a) shows the consensus map obtained, and 16 chromatic color terms are extracted as candidates for new BCTs. American English and Japanese have color terms for each of the color terms that are not Berlin and Kay's (1969) basic colors, except for the corresponding Japanese word for magenta


**Figure 6:** Two motifs obtained for Japanese.

in American English: peach/hada, beige/cream, olive/oudo, sky/mizu, maroon/enji, teal/emerald, and lavender/fuji. This similarity in CCTs of non-BCTs suggests that there are rules for color categorization beyond Berlin and Kay's (1969) BCTs for different languages.

Figure 7 (b) compares consensus maps clustered under the same condition as the Taiwanese Chinese data, that is, with a number of clusters of eight, which corresponds to eight chromatic BCTs (Hsieh et al. 2020). The agreement of each of the color regions is remarkable among the three languages. Correlations between the categories of the three languages explain more than 88% of the variance, leaving little room for variation due to cultural differences. This indeed suggests a strong contribution of physiological processes for color perception to developments of color lexicons.

For motif analysis, Hsieh et al. (2020) found three motifs, as shown in Figure 8. The most obvious difference among them is the sub-regions in blue area in color space. Motif 2 has a region for light blue (*mizu*), and Motif 3 has a region of greenish-blue (teal). No such sub-region is seen in Motif 1. Additionally, there are differences in the number of color teams among the three: eight, 10 and 12 for Motifs 1, 2, and 3, respectively. If the number of participants classified as users of each motif is considered, it is natural to summarize that Motif 1 is common for many of the Taiwanese-Chinese speakers, Japanese speakers, and American-English speakers; Motif 2 is unique to a few of the Japanese speakers; and Motif 3 is unique to a few of the American-English speakers. All Taiwanese-Chinese speakers, a few Japanese speakers, and approximately half of the Americans-English speakers are Motif 1 users. The majority of the Japanese speakers are Motif 2 users. Less than half of the American-English speakers are Motif 3 users. Motif analysis is useful to explore

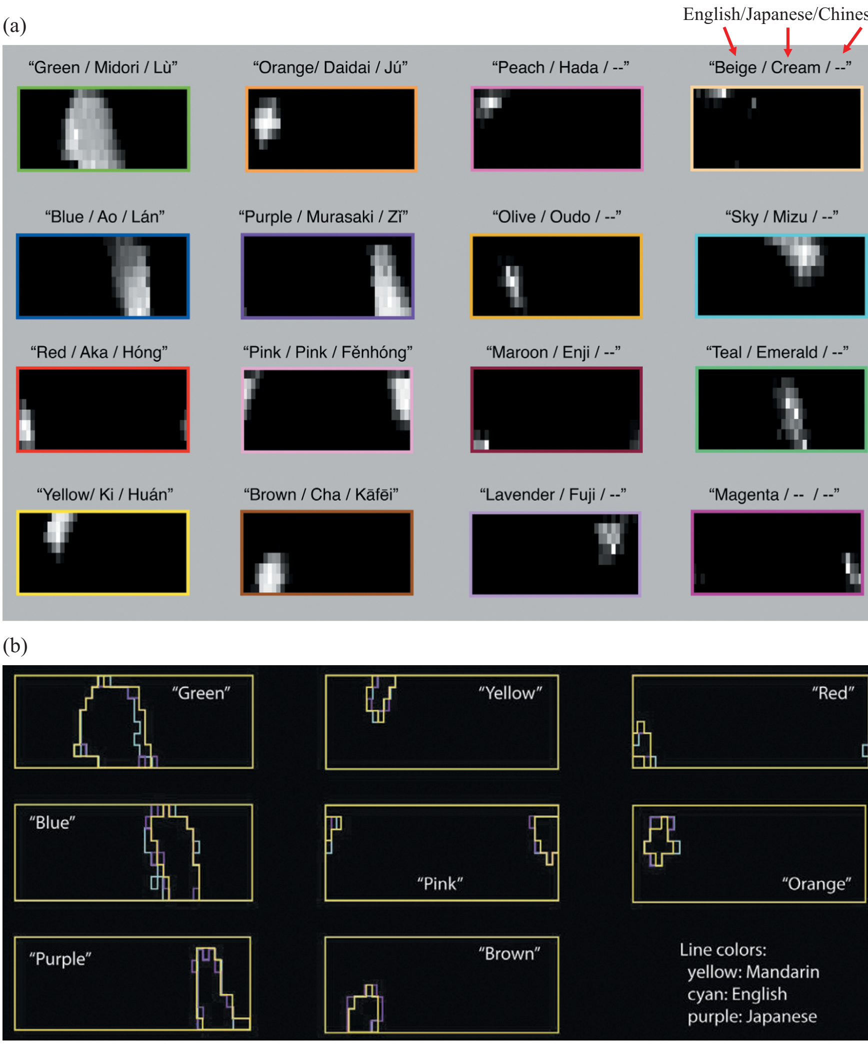



**Figure 7:** Consensus map with the k determined by gap statistic (a) and consensus map with k of eight obtained from pooled data of experiments in the three languages.

individual variations considering universal factors, which are perhaps related to physiological processes independent of cultural differences.

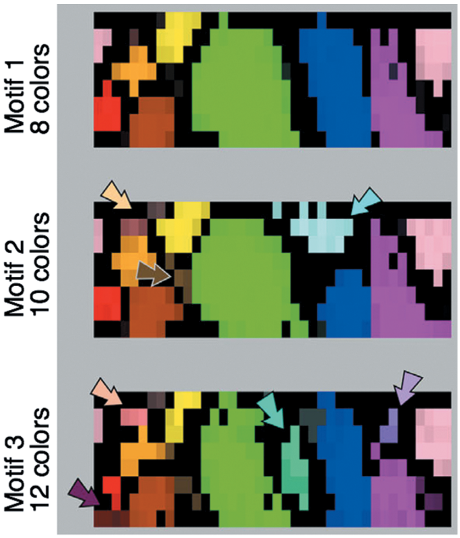



**Figure 8:** Three motifs obtained from the pooled data of the three languages. Arrows in Motif 2 (Peach/Hada, Olive/Oudo, and Sky/Mizu) and Motif 3 (Maroon/Enji, Beige/Cream, Teal/Emerald, and Lavender/Fuji) show clusters that are unique in each motif. This analysis revealed that Motif 1 with eight chromatic categories best represented the Taiwanese's color-naming structure (100% Taiwanese informants were classified to show this motif). The majority (86.0%) of the Japanese informants were classified as showing Motif 2 (12.3% and 1.75% for Motifs 1 and 3, respectively). The American informants were mostly split between Motif 1 (56.8%) and Motif 3 (37.3%).

## 4.2 Similarity of lexical partitioning of color space

Genetic variation for color perception has been studied for decades. One of the variations comes from color deficits, such as dichromacy (Neitz and Neitz 2011), where only two – instead of three – types of cone photoreceptors contribute to color perception. Another may come from possible tetrachromacy (Jordan et al. 2010). Approximately 12% of female are thought to have four types of cone instead of three. This suggests that they may be tetrachromacy instead of trichromacy. Although behavioral tests do not support tetrachromacy for most candidates, one woman has been reported to have passed the tetrachromacy tests. Although detailed analyses have been reported for these issues, to the best of our knowledge, no study has reported its relationship with cultures. The visual processes for color perception at stages that follow cone output are not as clearly known as the process at the photoreceptor level. Therefore, it is possible that cultures have an influence at a later stage.

Previous studies have shown the relationship between languages and/or cultures and color perception. The Stroop effect is a phenomenon that suggests the influence of the color category on color perception. When a person is asked to name the color of the ink of a word, he or she can respond faster with the color term to indicate the ink color, but responds slower with the term to indicate the different color. Additionally, it is easier to read a color term when the term is printed with the ink of the color that the term indicates, compared to the ink of a different color. As this effect cannot be shown for someone who does not read the word used, categorization of color links perception with concept, where perhaps cultures contribute. Another line of study of the relationship between languages and perception relates to the effect of color terms in different languages on visual functions (Thierry et al. 2009; Winawer et al. 2007). Thierry et al. (2009) demonstrated that two Greek color terms distinguishing light blue and dark blue lead to greater and faster responses for discrimination of these colors in Greek speakers compared to English speakers. This suggests that the difference in language influences perceptual ability and the same study also shows similar effects in brain activity.

Naming colors by categories or terms is a basic function of communication. Colors are often used to indicate an object, such as to say "take the blue cap," "get the red pen," "find the person in green shirt," and so on, and therefore, color categorization by color terms has been investigated (Yaguchi et al. 2004; Yokoi and Uchikawa 2005). These studies focus on common responses from participants, used without consideration of cultural differences and based on the assumption of the common processes among people from different cultures, as Berlin and Kay (1969) suggested. Similarities and differences among different languages described in the present study would facilitate future studies in the field of lexical partitioning of color space, particularly by CCTs in addition to Berlin and Kay's (1969) BCTs.

The three studies introduced in this research provided crucial information of possible influences of cultures on the relationship between color terms and color perception, or of the relationship between language and perception in general. There are color terms frequently used by many users of a language beyond the 11 BCTs proposed by Berlin and Kay (1969) for Japanese and American English. A question here is how these color terms, that is CCTs, are related to language or cultural differences. On the one hand, the answer to the question whether culture influences color term or its usage is "yes" according to the motif analyses. The difference between Japanese and American-English informants is seen at the blue region. Although both Motif 2 of Japanese and Motif 2 of American English have a sub-region at blue region, the sub-regions are not similar. Japanese Motif 2 has a light-blue region to represent *mizu,* but American-English Motif 2 has a greenish-blue region for both light and dark to represent *teal.* On the other hand, the answer to the question may be "no" according to the similarities of CCTs between

Japanese and American English. In addition to identical results of BCTs (Figure 7), clustering with the pooled data from Japanese, Taiwanese Chinese, and American English shows coherent results with eight and nine CCTs for Japanese and American English, respectively. Color terms given above panels in Figure 7(a) indicate corresponding color terms between Japanese and English for all the CCTs, including those in addition to BCTs, except for one: the region for Magenta is clear for English but not for Japanese. Even for *mizu* and *teal*, which motif analysis suggested as unique for Japanese and English, there are corresponding color terms: *sky* in American English and *emerald* in Japanese. The usage of color terms is likely under the influence of physiological processes even beyond the 11 BCTs. This is supported by studies of other languages. In contrast to American English, the color term for light blue has been reported for several other languages, such as Russian (Winawer et al. 2007), modern Greek (Thierry et al. 2009), Italian (Bimler and Uuskula 2014), and Turkish (Ozgen and Davies 1998), although these studies did not focus on precise classification of color regions. Furthermore, a recent study of Thai with the same procedure stimuli as in the studies introduced in this study, shows a color term for light blue in Thai (Panitanang 2019). Based on these discussions, we suggest that physiological processes of color perception influence usage of color terms more than cultural differences do.

## 4.3 Communications with colors

Similar usages of color terms among different languages are essential for verbal communications. Even words differ between languages, as *aka* and red, almost perfect communication is possible via translation. Translation is easily realized by a computer in the case where corresponding terms of different languages indicate a large overlap in color space. In other words, even when two groups use the same word for particular colors, communication fails if there is only partial overlap of regions in color space. We estimated efficiency of color communicate among the Japanese, Taiwanese, and Americans based on color-naming experiments. As the method and visual stimuli are virtually identical for three studies with the Japanese, Taiwanese, and Americans, reliable estimations are possible.

To compare accuracy in communicating colors among languages, we calculated the group mutual information, GMI (Hsieh et al. 2020) as follows:

$$GMI = \sum_{s,r} p(s,r) \log_2 \left( \frac{p(s,r)}{p(s)p(r)} \right) \tag{1}$$

where *p(s,r)* is a 11 x 11matrix of the joint probability distribution, which is obtained from the number of color patches named with each color term. *p(s,r)* indicates joint probability, which shows the percentage of patches named with the term by the sender and receiver, and *p(s)* and *p(r)* indicate probability of each color term used by the sender and receiver. The GMI is a summation of the combination of *p(s), p(r),* and *p(s,r)* for each of the 11 BCTs, which are determined by the color map of 11 BCTs for each individual informants.

The calculated GMIs are shown in Table 1. The values are similar for all combinations among the three languages, including pairs of people from the same language groups. The maximum GMI value, that is perfect communication, for 11 terms is 3.46; therefore, the values around 2 is not very high. However, the imperfect communication is likely due to individual variations rather than any factor related to language or cultural differences. This is another support that less contribution of culture differences to color term usages than factors common for all humans such as physiological ones.

**Table 1:** Group mutual information (GMI) between two languages.

| GMI (bits) | Japanese | Taiwanese Chinese | American English |
|---|---|---|---|
| Japanese | 2.196 | | |
| Taiwanese Chinese | 2.038 | 2.038 | |
| American English | 2.051 | 2.002 | 2.108 |

The investigation of color communications among people with different languages provide significant insight for human communication in general. Communication between two independent systems (two computers, two brains, or a computer and a brain) requires common knowledge of signals sent and received as, for example, Morse code. In this sense, it is puzzling how successful communication occurs among humans, whose brains are independently developed. Perhaps, we should assume that common knowledge and/or rules come from factors common for all people, such as genetic factors. We believe that studying relationships with perceptual and cognitive processes will contribute to understanding significant aspects of language development.

## 4.4 Number of color terms and definition of basic color terms

Although similarities are discussed in the previous sections, there are, of course, differences among different languages in color terms. A unique result based on the Taiwanese data is the smaller number of chromatic color terms used and smaller number (8) of color categories identified by the clustering method, compared to Japanese and American English (16 and 17, respectively). Hsieh et al. (2020) speculated that the monolexemic constraint applied could be one of the reasons. The Mandarin language that the Taiwanese use, in general, does not allow for strict evaluation of the monolexemic criterion because most color words comprise at least two characters, which represent single lexemes in Mandarin. The variety of color naming in modern Mandarin Chinese exhibits better in compound color terms instead of single-worded color terms. However, this does not imply there are not many monolexemic color term in Mandarin Chinese. Sun and Chen (2018) demonstrated that their participants were capable of selecting proper color chips to match historical color terms. They were asked to select color chips whose color was the one indicated by color terms, which was different form color naming of color chips in the experiments introduced here. Therefore, the issue of the color term numbers can be related to how informants are asked to link color and color terms. Active selections of monolexemic color terms may force the selection of limited terms for informants in Hsieh et al.'s (2020) study. This can be a problem related to definition or determination of BCTs. Although the monolexemic property is a widely accepted requirement for BCTs, there could be cases where the criterion is not strictly appropriate, as in the case of Mandarin Chinese. Certain color terms of a few languages are not monolexemic, but satisfy other criteria for BCTs. The clustering technique described here can be applied to solve the problem with color-naming without restriction of monolexemic terms. Other objective measures, such as reaction time for responses (Boynton and Olson 1990; Uchikawa and Boynton 1987) and brain activity measurements (Yang et al. 2016) may be applied to solve future problems.

# 5 Conclusions

Studies for Japanese, Mandarin Chinese spoken in Taiwan, and American English show similar lexical partitioning of color space, which is likely related to physiological properties of color vision. Additionally, there are differences among these different languages, which could be attributed to cultural influences, such as *mizu* for light blue in Japanese and *teal* for (light or dark) greenish-blue in American English, both in blue regions; however, further investigation is necessary before conclusion are drawn. Investigation of color lexicon assisted by advanced data analyses is promising for future linguistic sciences as well as color sciences.

We discussed the significance of ascertaining the value of information to avoid the problem of information overload. This study showed that visual information – color information in the present case – can be expressed in a universal manner even among people with different cultural backgrounds. Although the value of information depends on the people who use it and its purpose, yet there are factors that are common for all humans. Finding such factors for a variety of functions helps build knowledge to estimate the value of information and prioritize information accordingly to avoid information overload.

## References

Berlin, Brent & Paul Kay. 1969. *Basic color terms: Their universality and evolution*. Berkeley, US: University of California Press.

Bimler, David & Mari Uuskula. 2014. "Clothed in triple blues": Sorting out the Italian blues. *Journal of the Optical Society of America A* 31. A332–A340.

Boynton, Robert. M. & Conrad. X. Olson. 1990. Salience of chromatic basic color terms confirmed by three measures. *Vision Research* 30.1311–1317.

Brown, Angela M., Abdirizak Isse, & Delwin T. Lindsey. 2016. The color lexicon of the Somali language. *Journal of Vision* 16(5). 14.

Ferrer i Cancho, Ramon & Ricard V. Solé. 2001. Two regimes in the frequency of words and the origins of complex lexicons: Zipf's Law Revisited. *Journal of Quantitative Linguistics* 8. 165–173.

Gerlach, Martin & Eduardo G. Altmann. 2013. Stochastic Model for the Vocabulary Growth in Natural Languages. *Physical Review X* 3(2). 021006.

Hsieh, Tracy T., Ichiro Kuriki, I-Ping Chen, Yumiko Muto, Rumi Tokunaga, & Satoshi Shioiri. 2020. Basic color categories in Mandarin Chinese revealed by cluster analysis. *Journal of Vision* 20(12). 6.

Jordan, Gabriele, Samir S. Deeb, Jenny M. Bosten, & J. D. Mollon. 2010. The dimensionality of color vision in carriers of anomalous trichromacy. *Journal of Vision* 10(8). 12.

Kuriki, Ichiro, Ryan Lange, Yumiko Muto, Angela M. Brown, Kazuho Fukuda, Rumi Tokunaga, Delwin T. Lindsey, Keiji Uchikawa, & Satoshi Shioiri. 2017. The modern Japanese color lexicon. *Journal of Vision* 17(3). 1.

Lindsey, Delwin. T. & Angela. M. Brown. 2006. Universality of color names. *Proceedings of the National Academy of Sciences* 103(44). 16608–16613.

Lindsey, Delwin. T. & Angela. M. Brown. 2009. World Color Survey color naming reveals universal motifs and their within-language diversity. *Proceedings of the National Academy of Sciences* 106(47). 19785–19790.

Lindsey, Delwin. T. & Angela. M. Brown. 2014. The color lexicon of American English. *Journal of Vision* 14(2). 17.

Marlowe, Frank W., J. Colette Berbesque, Clark Barrett, Alexander Bolyanatz, Michael Gurven, & David Tracer. 2011. The 'spiteful' origins of human cooperation. *Proceedings of the Royal Society B Biological Science* 278. 2159–2164.

Muraoka, Hiroaki, Giro Gyoba, Yoiti Suzuki, Satoshi Shioiri, Mitsuyuki Nakao, Mariko Nihei, & Hiroki Shoji. 2017. Gigantic amount information and storage technology: Challenge to yotta-byte-scale informatics. *IEICE Technical Teport* 116. 27–32.

Neitz, Jay & Maureen Neitz. 2011. The genetics of normal and defective color vision. *Vision Research* 51. 633–651.

Ozgen, Emre & Ian Davies. 1998. Categorical effects in colour perception: cross-language differences and category learning. *Perception* 27. 177.

Panitanang, Nischanade, Chanprapha Phuangsuwan, Ichiro Kuriki, Rumi Tokunaga, & Mitsuo Ikeda. 2019. Thai basic color terms and new candidate nomination. *Proceedings of the 5th Asia Color Association Conference, ACA2019 Nagoya*, 164–169. Nagoya: Asia Color Association & Color Science Association of Japan.

Shioiri, Satoshi, Yoshiyuki Sato, Yuta Horaguchi, Hiroaki Muraoka, & Mariko Nihei. 2021. Quali-Informatics in the society with yotta scale data. Paper presented to the IEEE International Symposium on Circuits and Systems (ISCAS), 2021.

Smil, Vaciav. 2019. Data world: Racing toward yotta. *IEEE Spectrum* 56. 20.

Stroop, John, R. 1935. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology* 18. 643–662.

Sun, Vincent C. & Chien-Chung Chen. 2018. Basic color categories and Mandarin Chinese color terms. *PLoS One* 13. e0206699.

Thierry, Guillaume, Panos Athanasopoulos, Alison Wiggett, Benjamin Dering, & Jan-Rouke Kuipers. 2009. Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences* 106. 4567–4570.

Uchikawa, Hiromi, Keiji Uchikawa, & Robert M. Boynton. 1989. Influence of achromatic surrounds on categorical perception of surface colors. *Vision Research* 29(7). 881–890.

Uchikawa, Keiji & Robert M. Boynton. 1987. Categorical color perception of Japanese observers: Comparison with that of Americans. *Vision Research* 27(10). 1825–1833.

Yaguchi, Hirohisa, Nanako Endoh, Takayoshi Moriyama, & Satoshi Shioiri. 2004. Categorical color rendering of LED light sources. *CIE Expert Symposium on LED Light Sources*. 20–23.

Hajime Ono, Takuya Kubo, Manami Sato, Hiromu Sakai, and Masatoshi Koizumi

# Chapter 5
# Word orders, gestures, and a view of the world from OS languages

## 1 Introduction

Gestures are one of human beings' basic forms of non-verbal communication. Investigating gestures may sound like something linguists do not have to pay attention to because it does not involve language. Unlike exploring sign language, which is a study of the grammar of human language, gestures are typically not a topic found in linguistics courses. However, studies of gestures have asked a wide range of questions, some of which are relevant to linguistics.

Languages with either SOV or SVO word order outnumber languages with other possible word orders (Haspelmath et al. 2005). The distribution of word orders is skewed. Assuming that syntax is responsible for determining word order, it is relevant for linguists to ask why specific word orders are found so frequently. Even though we know that syntax plays a role in such skewed bias, the extent to which it does is still unclear. Syntax is essential, but the biases for SOV and SVO may have partially emerged from the properties of (non-linguistic, possibly universal) human cognition. The prevalence of SOV and SVO is also found in historical changes in languages (e.g., SOV to SVO) (Gell-Mann and Ruhlen 2011) and the emergence of (sign) languages (see some discussion on Al-Sayyid Bedouin Sign Language in Sandler et al. 2005). This chapter summarizes the results of our gesture study, which provides data to illuminate the close connection between human cognition and grammar.

The prevalence of SOV word order in the world's languages and its connection to human cognition are part of the central focus of gesture studies. Goldin-Meadow et al. (2008) observed that native speakers of English, Turkish, Spanish, and Chinese frequently used SOV gestures to describe events involving an agent (actor), a patient, and an action.[1] Interestingly, in their study, there was no language effect; regardless of the basic word order properties, the participants frequently used SOV

1 In this chapter, we use "S," "V," and "O" instead of thematic terms such as "agent (actor)," "action," and "patient" even when we talk about gesture orders. Although "S" and "O" should be used for referring to particular positions in the syntactic structure and not for gesture orders, the use of SVO, SOV, and others for identifying gesture orders allow us to follow the patterns quite easily.


https://doi.org/10.1515/9783110778946-005

gestures. Goldin-Meadow et al. (2008) argue that SOV is a natural sequence for representing events. Assuming that gesture production is independent of grammar, the SOV sequence has a cognitive advantage.

Because Goldin-Meadow et al.'s (2008) claim is influential, it is essential to further examine whether SOV prevalence in gesture production can be observed universally. The languages studied in the literature are quite limited, most being the SO type, where the subject comes before the object in the basic word order. Here, we aim to test an OS-type language (Kaqchikel; Mayan, Guatemala) in which the object comes before the subject in its basic word order, and examine the relationship between the language(s) the participants use and gesture patterns. The connection between language and gesture bears on questions in the Sapir-Whorf hypothesis (Sapir 1921; Whorf 1956). Goldin-Meadow et al.'s (2008) claim about the representation of events implies that we view the world (through entities, actions, and events) in a particular way regardless of the languages we speak. In gesture studies with native speakers of Kaqchikel, we can ask whether they view the world as native speakers of SO-type languages do. Such an inquiry may provide us with a key to the event comprehension process of Kaqchikel speakers because gesture production reflects (at least) part of the processes of how people interpret events.[2] Therefore, in this study, we asked native speakers of Kaqchikel to perform a gesture task.

Previous gesture studies have already identified some factors that affect gesture order. For instance, participants produced more SVO gestures for events involving two human entities than for events involving a human and an object (Gibson et al. 2013; Hall, Mayberry, and Ferreira 2013; Hall, Ferreira, and Mayberry 2014). This observation is known as the reversibility effect of events. We will review this effect and the two major approaches proposed in recent studies. According to the noisy-channel hypothesis (Gibson et al. 2013), SOV gesture order may not be the best way to convey a message for events with two human entities because it is potentially ambiguous concerning the thematic role when one of the gestures for the entity is lost. On the other hand, the SVO gesture order is more durable because the relative order in relation to the action gesture can indicate which entity is the agent/patient. This increases the proportion of SVO gesture production when the event involves two human entities.

Hall, Mayberry, and Ferreira (2013) and Hall, Ferreira, and Mayberry (2014) have a different idea about the reversibility effect. They argued that when participants perform action gestures, they do so as if they are the agent doing the action

2 It is also true that there is no transparent relationship between gesture production and event comprehension. Various properties, such as attention, focus, novelty, etc., have an influence on gesture production in addition to sequencing participants and actions in a linear manner.

(imagine a situation in which you are making a gesture for pushing or kicking). From this perspective, the participants perform gestures for S and V from an agent's point of view but not for O. Producing an SOV gesture, the participants perform an S gesture from an agent's point of view, then a gesture for O, and a gesture for V, again from an agent's point of view. Here, there is a switch of the point of view from S to O, and O to V. However, in making an SVO gesture, they can keep the perspective of the agent's role while making gestures of S and V. In other words, they prefer the SV gesture sequence (rather than the OV). Both accounts explain a gesture bias for an event with two human entities, but we can examine another event type so that those two accounts result in different predictions. In Experiment 2, we used an event in which an inanimate agent caused some action upon a human entity (e.g., the boat pulling the boy, the big tire bumping against the man). This provides critical insight into factors that play a crucial role in determining constituent order in gesture production.

In addition to event properties, using a predetermined action gesture encourages participants to use SVO gestures (Hall, Ferreira, and Mayberry 2014; Marno et al. 2015). This situation arises when the participants, somewhat intensively, practice gestures using (at least some) actual stimulus pictures beforehand. There are some suggested reasons for the increase in SVO gesture production. With practice, gestures start behaving like lexical items in language, inviting the involvement of grammar (Marno et al. 2015); the intensive practice creates a language-like setting (Hall, Ferreira, and Mayberry 2014). The exact role of practice remains unsettled and needs to be examined further to uncover the underlying mechanisms for such observations.

In this study, we explore the possibility that practicing action gestures alters what the action gestures represent. Discussing the prevalence of SOV gestures in their experiments, Goldin-Meadow et al. (2008) noted a close cognitive tie between the patient and the action (see also Goldin-Meadow (2003) for a relevant discussion). In particular, we suggest that predetermined action gestures break this tie between the patient and the action. We call this suggestion "patient-boundedness," which is similar to what is known as the symbol-grounding problem (Harnad 1990; Imai 2017). The symbol-grounding in lexical learning is a process of abstraction from each instance. In learning vocabulary, we initially label (e.g., throwing) a particular action (e.g., an event of throwing a ball) and may use another label for a different type of patient of the event (e.g., throwing a man). Eventually, we end up using the same label for a wide range of patient types after encountering various situations. Through abstraction, we now have a label of throwing that represents various things. Similarly, in a typical impromptu gesture situation, the action gesture for throwing may take various forms. The precise manual movement differs depending on the patient's shape, size, or weight. However, through practice, participants develop a general throwing gesture that is applicable to many kinds of patients. Also, such a gesture would take a perspective of agent and contribute to an increase of the SVO gesture. We examined the practice effects and how patient-boundedness affected the gestures.

The fundamental question addressed in this study is whether and how speakers of Kaqchikel, an OS-type language, produce gestures. A study based on such a typologically unique language is crucial to determine the extent to which the SOV gesture prevalence found in past research is genuinely universal. In Experiment 1, we tested gesture production for events involving different patients (human, object, or animal). In Experiment 2, we asked the participants to practice gestures before they performed the task to see how the practice affects their gesture patterns. Using the measure of action properties based on complexity and patient-boundedness, we looked for cues of the underlying mechanism for an increase in SVO gestures. We also included a condition for events in which an inanimate entity has an effect on a human entity. As we describe below in more detail, the existing hypotheses make different predictions for various conditions.

This study targeted native speakers of Kaqchikel, a Mayan language spoken in Guatemala (population of approximately 410,000) (Eberhard, Simons, and Fennig 2019). We introduce some basic grammatical properties of Kaqchikel below. It notably allows word-order alternations between VOS and SVO, as shown in (1) and (2), making this language very distinct from the set of languages often investigated in the field. Kaqchikel is a head-marking language; noun phrases do not carry case markers. However, the verb has agreement markers with the subject and object in the ergative-absolutive case alignment in addition to a morpheme for representing aspect.

| | | | | | | |
|---|---|---|---|---|---|---|
| (1) | X-∅-u-chöy | ri | chäj | ri | ajanel. | (VOS) |
| | COMPL-ABS.3SG-ERG.3SG-cut | DET | pine.tree | DET | carpenter | |

| | | | | | | |
|---|---|---|---|---|---|---|
| (2) | Ri | ajanel | x-∅-u-chöy | ri | chäj. | (SVO) |
| | DET | carpenter | COMPL-ABS.3SG-ERG.3SG-cut | DET | pine.tree | |
| | “The carpenter cut the pine tree.” | | | | | |

In Kaqchikel, VOS word order has been claimed to have a basic syntactic structure; SVO is a derived word order (England 1991; Koizumi et al. 2014; Rodríguez Guaján 1994: 200). In an (auditory) sentence-plausibility judgment task, Koizumi et al. (2014) found that the VOS sentences were responded to faster than the SVO sentences, suggesting that the structural complexity plays a role in determining the processing cost. In contrast, in an (oral) picture-description task, Kubo et al. (2015) showed that SVO was produced most frequently, and VOS was produced in about 20% of the trials, indicating an interesting comprehension-production discrepancy.

# 2 Experiment 1

## 2.1 Methods

### 2.1.1 Participants

We gathered 32 native speakers of Kaqchikel from towns and villages around Antigua, Guatemala. They gave written consent and were paid to participate in the experiment. Participants completed a linguistic background questionnaire before the experiment. According to the questionnaire, they also use Spanish at work and at school daily, but Kaqchikel seemed to be their primary language.

### 2.1.2 Materials

The stimuli were 18 line-drawn pictures depicting a transitive action (e.g., catching, pushing, breaking, folding; see the sample picture of stimuli in Figure 1). These stimuli consisted of three types of events based on the animacy of the patient in the event (human, inanimate object, or animal; the agent is always human, so the three types are Human-Human, Human-Object, and Human-Animal). There were six pictures of stimuli for each type. Actions in the events were different for each type. Events for the Human-Human type were reversible (e.g., hitting, kicking, helping) in that the entity depicted as the patient can be the agent of the same action. On the other hand, not all events in the Human-Animal type are reversible (e.g., holding, washing), but only some of them (e.g., chasing, pulling).


**Figure 1:** Sample picture stimuli for each condition.

### 2.1.3 Design and procedure

Participants were shown a picture and instructed to produce gestures and show them to a confederate (a native speaker of Kaqchikel) sitting in front of the participant, who was supposed to guess what was depicted in the picture. During the introduction, the participants saw some sample picture stimuli and pictures of human characters (old man, old woman, lady, man, boy, and girl). As a result, they became familiar with the characters and could differentiate between different human characters performing the gestures. All instructions were provided in Kaqchikel by another experimenter.

Eighteen picture stimuli, individually printed on a sheet of paper, were presented to the participants in pseudo-random order. The experimenter showed each picture from behind the confederate sitting in front of the participant so that the picture was visible only to the participant. The experimental session was video-recorded to code the data.

#### 2.1.4 Coding and analysis

Due to some mechanical problems with the video recording, data from six participants were not available. Data from the remaining 26 participants were analyzed. Data from two pictures in the Human-Object condition were removed from the analysis because the participants produced many gestures with just the agent and action (i.e., SV, for those pictures, they omitted gestures for the patient), resulting in the Human-Object condition having data from only four pictures.

Gesture sequences were coded by the second author according to constituent orders (e.g., SOV, SVO, and OVS). Vague gestures (no clear transition from one to another) or completion with just one gesture (e.g., just an action) were categorized as “uncodable” and eliminated from further analyses (14/416; 3.4%). Repetitive gestures (e.g., VSVO, SOSV), gesture productions with a two-gesture sequence (e.g., SV, OV), and other gesture sequences that were too few to analyze were categorized as “others.”

The data were analyzed using logistic mixed-effect models (Jaeger 2008) with the lme4 package (Bates et al. 2015b) for R software (R Core Team 2019). Statistical models were built for the gesture productions of either SVO or SOV order (i.e., removing data from the “other” category and a few other gesture productions, such as OSV), to evaluate the effects of the animacy of the patient on the rates of SVO and SOV gesture production. The SVO gesture data were coded as 0, and the SOV gesture data were coded as 1 and used as a dependent variable. As a fixed factor, three event types were Helmert coded (Human-Human condition = [0, –2]; Human-Object condition = [1, –1]; Human-Animal condition = [–1, –1]), so that the Human-Human condition was compared to the other conditions, and then the two non-Human conditions were compared to each other. In the models, random intercepts and random slopes were estimated for the participants, while only random intercepts were estimated for the items. The models were initially fit with the maximal random effects structure, and the backward selection procedure determined the maximal model, in which no correlation between parameters was assumed to be the optimal one (Bates et al. 2015a).

## 2.2 Results

### 2.2.1 Group data

Figure 2 shows the overall distribution of the constituent order proportions for each event type averaged across the participants. There were many gesture productions of SVO and SOV orders across all event types and a few OSV orders. There were very few VOS gestures (only four instances), even though VOS is an available constituent order in Kaqchikel sentences.

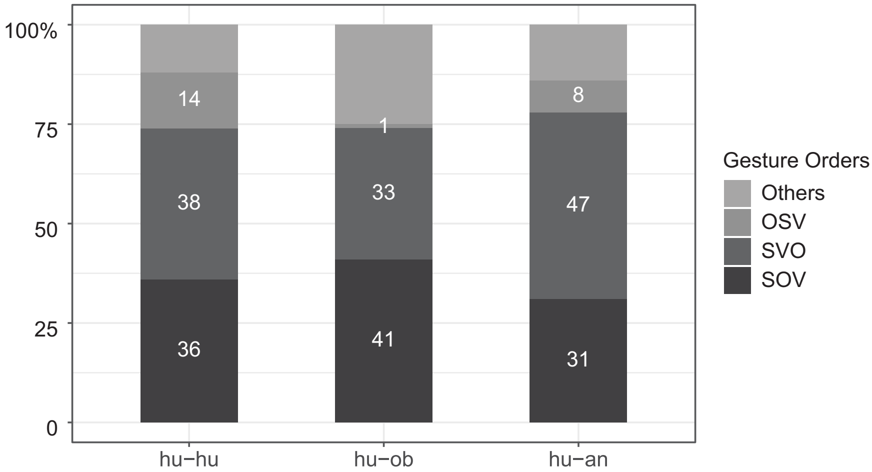



**Figure 2:** Distribution of constituent orders.

**Table 1:** Summary of the statistical analysis in Experiment 1.

| | Estimate | Standard error | *z* value | *p* value |
|---|---|---|---|---|
| (Intercept) | 0.20 | 0.54 | 0.38 | |
| human-animate vs. human-object | 0.50 | 0.25 | 2.03 | < 0.05 * |
| (animate + object) vs. human | 0.05 | 0.21 | 0.23 | 0.82 |

The mean proportions of SOV gestures were significantly higher in the Human-Object than the Human-Animal conditions (see Table 1). On the other hand, no difference was found between the non-Human conditions and the Human-Human condition. Also, as Figure 2 indicates, the SOV prevalence found in Goldin-Meadow et al. (2008) was not seen with native Kaqchikel speakers as a group.

#### 2.2.2 Individual data

A close inspection of the individual data indicated a sizable between-participant variation in the results. Figure 2 shows that the participants produced SVO gestures about half the time and SOV the other half. Still, while some participants produced many SVO gestures but not SOV gestures; others produced the opposite pattern. We therefore calculated the SVO-biased score for each participant, which identifies the production bias between SVO and SOV. Scores for each participant were calculated by subtracting the number of SOV gesture productions from the number of SVO gesture productions separately for the Human-Human and the Human-Animal conditions (we set aside the Human-Object condition from this analysis because the data from the two picture stimuli were eliminated from the analysis). There were six trials for each condition, so the score ranges from +6 to –6. A positive score means that the participant produced more SVO gestures than SOV gestures for that condition, and a negative score means that the participant produced more SOV than SVO. Figure 3 shows the distribution of the SVO-biased scores. Notice that the score distribution is widespread, meaning that there are participants who produced many SOV gestures and participants who produced many SVO gestures.

In previous studies, the participants produced more SVO gestures than SOV gestures for reversible events in which both entities were human (Gibson et al. 2013; Hall, Mayberry, and Ferreira 2013; Hall, Ferreira, and Mayberry 2014). In our experiment, the Human-Human condition involved reversible events, while not all of the stimuli in the Human-Animal condition were reversible. Figure 4 is a scatter plot showing the SVO-biased scores from each participant in the Human-Human and the Human-Animal conditions. The data points on the diagonal line from the lower left to the upper right indicate that these participants had an equal bias against SVO gestures and SOV gestures in these conditions. The data points below this line indicate the participants who showed a stronger SVO bias in the Human-Human condition. In contrast, the data points above this line indicate the participants who showed a stronger SVO bias in the Human-Animal condition. The figure shows a large variation. If there are reversible event effects, the distribution should cluster toward the lower right quadrant, but the distribution itself is spread out. One possible source of the weak reversible effects is that some actions in the Human-Animal condition are also reversible, leading the participants to produce a relatively large number of SVO gesture sequences, even in the Human-Animal condition.

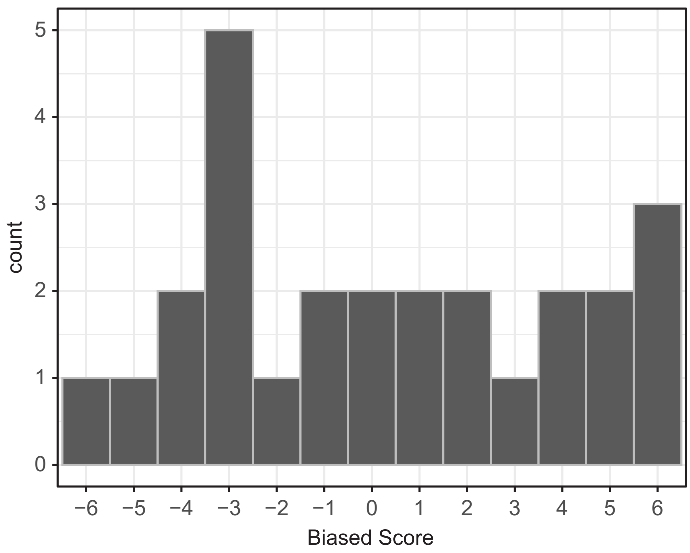



**Figure 3a:** Distribution of SVO-biased score in the Human-Human condition.

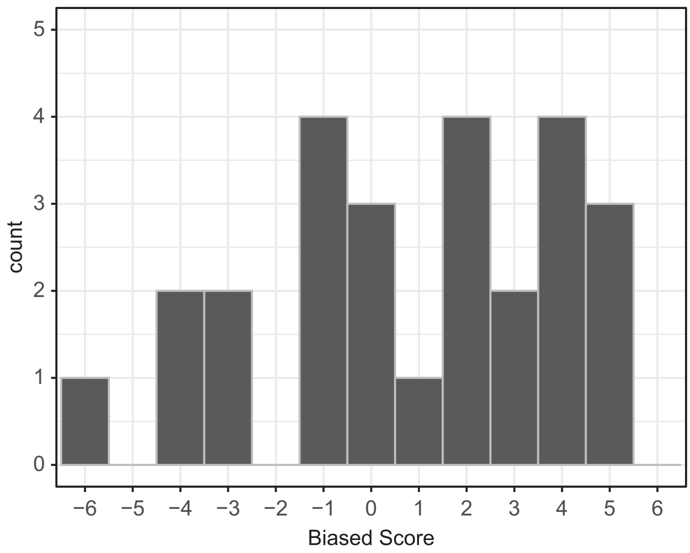



**Figure 3b:** Distribution of SVO-biased score in the Human-Animal condition.

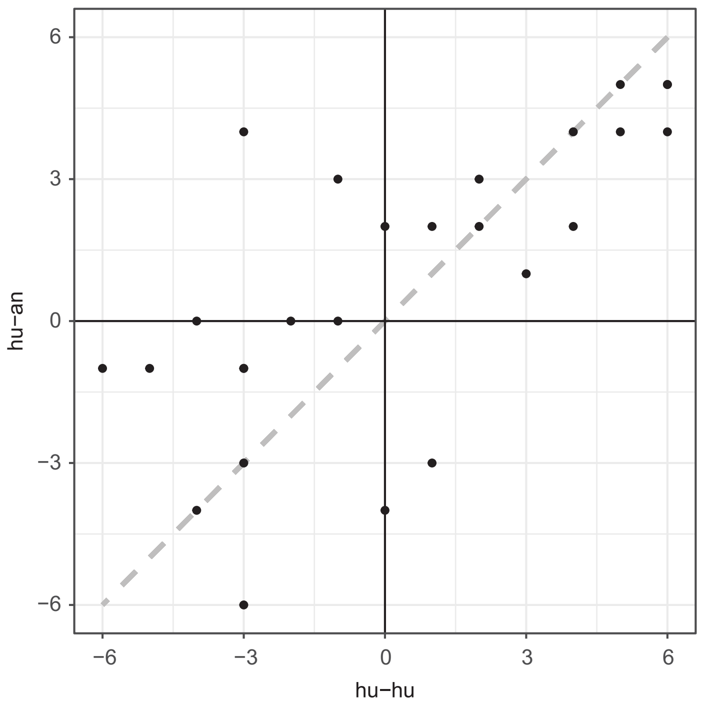



**Figure 4:** Scatter plot between the SVO-biased scores for the Human-Human condition and the Human-Animal condition for each participant.
Vertical axis: SVO-biased score for the Human-Animal condition.
Horizontal axis: SVO-biased score for the Human-Human condition.

## 2.3 Discussion

One observation from this experiment was that there was no strong SOV bias in native speakers of Kaqchikel as a group. Their mean proportion of SOV gestures, in general, was not as large as in previous gesture studies, as seen in (3) below. Each study listed in (3) used their own set of picture stimuli, so there is a possibility that the pictures (and whatever properties in the events) in our experiment somehow led the participants to avoid SOV gesture sequences. Still, the mean SOV proportions in most of the previous studies were much greater than those observed with native Kaqchikel speakers.

| | | |
|---|---|---|
| (3) | Goldin-Meadow et al. (2008): | more than 80% SOV (Turkish, English, Spanish, Chinese) |
| | Langus and Nespor (2010): | more than 80% SOV (Italian, Turkish) |
| | Gibson et al. (2013): | about 30–70% SOV (English), almost all SOV (Japanese, Korean) |

The finding that SVO gestures were produced about as often as SOV gestures appears to be related to the considerable between-participant variation. As we saw from the distribution of the SVO-biased scores, about a half of the participants produced a number of SOV gesture sequences (along with those who produced a number of SVO gestures). This does not oppose the claim that SOV is a natural sequence representing an event (Goldin-Meadow et al. 2008), but the results suggest that SOV may not be the only one. At this point, the exact source of the between-participant variation is not entirely clear.

Some may wonder whether participants produced many SVO gestures because they performed the task as if they were orally forming Kaqchikel sentences. A previous oral picture-description study (Kubo et al. 2015) showed that Kaqchikel speakers produced SVO sentences about 70–80% of the time. This may explain the large proportion of SVO gestures; however, Kaqchikel speakers produced VOS sentences approximately 20% of the time in the oral picture-description task. Because we did not see a comparable number of VOS gesture productions, it is unlikely that the participants who produced many SVO gestures did so because they performed the task as if they were performing an oral picture-description task. In addition, this does not help us answer why there was a significant variation among participants; if there was a group of participants who performed gestures like the oral picture-description task, and there was another who did not, the distribution of the SVO-biased score should have been a bimodal shape.

Another possible reason for seeing so many SVO gesture productions by Kaqchikel participants may be related to the accessibility of the action. Goldin-Meadow et al. (2008) claimed the universality of the SOV gesture order, suggesting that entities are cognitively more basic than actions. Imagine a gesture for a box depicting a shape. In contrast, a gesture for pushing would involve manual movements and be more complicated. Looking at the grammatical properties of Kaqchikel, the availability of VOS word order may allow Kaqchikel speakers to be more sensitive to action, increasing the accessibility of action (verb), even in impromptu gesture production. Note that the rich morphology of Kaqchikel verbs may also affect the accessibility of action because verbs in Kaqchikel have agreement markers for both subject and object. If accessibility is strongly related to the position of the verb in gesture production, and if cognitive properties such as accessibility are affected by the language the participants speak, even in a non-verbal task, Kaqchikel speakers may place the action gesture in an early position and produce SVO, in contrast to speakers of other languages investigated in the literature to date.

There are some further suggestions concerning the choice between SVO and SOV. Hall, Ferreira, and Mayberry (2014) and Marno et al. (2015) observed that the number of SVO gestures increased when the participants practiced the gesture before the task. The reason that practicing gestures affects the proportion of SVO gesture production is not yet fully understood. They suggest that the practiced gestures have some characteristics of lexical items in languages and that the grammatical system prefers the SVO order by default (Langus and Nespor 2010). Building on the literature's previous suggestions, we suggest that using practiced gestures will unleash action gestures from being bound by the patient (cf. Goldin-Meadow et al. 2008). When we think about a throwing gesture, the exact form of the action in the event "the woman throwing a ball" would be different from the event "the woman throwing a man." In other words, the exact form of the action gesture is typically dependent on the entity (patient). This would explain why the action gesture tends to follow gestures for entities. However, when the participants practice gestures, they set up a general gesture for throwing that works regardless of the patient's type, shape, or size. This removes the necessity of placing an action gesture after the patient. The effect of practicing gestures was examined in Experiment 2.

In Experiment 1, there was no significant SVO increase in the reversible Human-Human condition (compared to the Human-Object and the Human-Animal conditions), unlike in previous studies (Gibson et al. 2013; Hall, Mayberry, and Ferreira 2013; Hall, Ferreira, and Mayberry 2014). For instance, Gibson et al. (2013) argued that participants use SVO gestures in reversible events to avoid ambiguity. According to their noisy-channel hypothesis, in reversible events where both agent and patient are human, SOV gesture production is vulnerable to information loss. In the SOV gesture, when the information of one entity, either the agent or patient, is somehow lost, the resulting gesture (just one gesture for an entity followed by an action gesture) is ambiguous because the remaining gesture for an entity could be for an agent or patient. In contrast, even when a gesture for an entity is missing, the (thematic) role of the remaining element is evident in the SVO gesture. The perceiver could guess that the gesture before the action is an agent and the gesture after the action is a patient. There is no thematic role ambiguity in events where the patient is an inanimate entity (e.g., a box). The performer does not have to rely on relative ordering in relation to the action.

Hall, Mayberry, and Ferreira (2013) suggested an alternative explanation. Their approach suggested that people perform action gestures as if they were agents of events. From this perspective, SOV gestures are not an ideal way to order the elements in reversible events because the performers (namely, experimental participants) make a gesture for an agent, then a gesture for a patient, and then perform an action gesture in the role of the agent again. In contrast, SVO gestures do not require this role switch. According to their account, SVO gestures are preferred because people want to have an SV gesture chunk to maintain their role in the performance. In a non-reversible event with an inanimate entity, this role-taking is not an issue; the performer does not take the perspective of a box, glass, or another object.

In Experiment 1, there was a slight increase in OSV in the Human-Human condition (though it was not statistically examined). Under the role-conflict approach, an increase in SVO is expected in reversible events due to the preference for clarifying who is doing the action by placing an agent immediately before the action. Based on this approach, an increase in OSV is also relevant to marking the agent of the action, assuming that native Kaqchikel speakers have some familiarity with the OS order. Nevertheless, the fact that the number of SVO gesture productions does not differ between the Human-Human and other non-Human conditions is unexpected. The influence of event types, especially reversibility, should be examined further.

Given that the events with two human entities are potentially ambiguous and problematic because the entities involved in the event are equally likely to be an agent, events with a human and an inanimate entity are also problematic when the agent (or cause) is an inanimate entity. For example, an alarm clock awakens a girl, and a large heavy tire knocks down a man. It is fair to assume that there is a natural bias to consider that the agent is a human, so from the gesture producers' perspective, they must make sure to indicate which is an agent and which is a patient in this kind of event (i.e., Object-Human event) to convey a message in an intended way. Examining the gesture production for this kind of event is crucial because the two accounts mentioned above make different predictions (see some related discussion in Meir et al. 2017). Assuming that the ambiguity is quite high in this event, Gibson et al.'s (2013) noisy-channel hypothesis predicts an increase in SVO gestures to clarify which entity is the agent. In contrast, it is unlikely that the participants will take the role of inanimate entities. The role-conflict approach by Hall, Mayberry, and Ferreira (2013), therefore, does not predict an increase in SVO gestures because there is no strong motivation to have the SV gesture sequence. Experiment 2 examined their predictions.

# 3 Experiment 2

## 3.1 Methods

### 3.1.1 Participants

As in Experiment 1, we gathered 53 native Kaqchikel speakers from towns and villages around Antigua, Guatemala. Some of them participated in both experiments. Because Experiments 1 and 2 were conducted about one and a half years apart, there is no concern about the experience of participating in Experiment 1 influencing the results of Experiment 2. The participants gave written consent and were paid to participate in the experiment. They completed a linguistic background questionnaire before the experiment.

### 3.1.2 Materials

Stimuli were 30 line-drawn pictures depicting transitive events. These stimuli consisted of three event types: Human-Human, Human-Object, and Object-Human. There were 10 pictures of each type. Pictures of the Human-Human and Human-Object conditions were prepared using the same actions. As for the actions for the Object-Human condition, five were shared with other conditions, and the other five were unique to this condition.

### 3.1.3 Design and procedure

The experiment was conducted similarly to Experiment 1. However, the pictures were presented on a computer monitor, and the participant's eye movements were measured while they were producing gestures (Tobii T120, Tobii Technology, Sweden). Due to space constraints, unfortunately, eye-tracking data will not be presented here. Eight lists were prepared with different presentation orders and left-right counter-balanced positions. Before the task, there was a practice session in which the participants saw the clipped pictures depicting just the part of actions and entities and could decide how to gesture for individual actions. The experimenter helped the participants during the practice session so that the gestures they used were different for each entity and action, and they were easy to remember. There were no specific predetermined action gestures that the experimenter led the participants to use. Throughout this practice, the participants became familiar with each entity and action, but they did not know how those items were combined into a single event. During the task, each picture appeared on the screen after a fixation cross (1.5 s). Images remained on the screen until the participants finished their gestures and pressed the spacebar of the keyboard. The entire session was video-recorded to code the data.

### 3.1.4 Coding and analysis

The gathered data were coded in the same fashion as in Experiment 1. Due to some mechanical problems with the video recording and eye tracking, data from 14 participants were not available. Data from the remaining 39 participants were analyzed. As in Experiment 1, the SVO gesture data were coded as 0, and the SOV gesture data were coded as 1 and used as the dependent variable. Three event types were Helmert coded (Human-Human condition = [0, –2]; Human-Object condition = [1, –1]; Object-Human condition = [–1, –1]) so that the Human-Human condition was compared to the other conditions, and then the two conditions with an inanimate entity were compared to each other.

Regarding the factors that influence the proportion of SVO/SOV gesture production, the discussion from Experiment 1 suggests that at least two hypotheses, in addition to Gibson's noisy-channel account and Hall's role-conflict account, should be examined further. Goldin-Meadow et al. (2008) suggested that the SOV gesture sequence is a natural sequence for representing events because the action gesture is more complex than the gestures for entities. To measure the complexity of the action in the event, we administered a questionnaire that asked the participants to rate the complexity of the action in the event. Native speakers of Japanese (N = 33) saw a list of 15 verbs that corresponded to the actions used for the stimuli in Experiment 2 and rated their gesture complexity (how much they had to move their hands) using a 7-point Likert scale (1 = very simple to 7 = very complex). The raw scores were z-score transformed, and the mean complexity scores were calculated for each verb (action). These scores were used to examine whether the complexity of an action influences, more specifically decreases, the proportion of SVO gestures, as suggested by Goldin-Meadow et al. (2008).

We also administered another questionnaire to measure the "patient-boundedness" of actions. We have suggested that practicing the gesture before the task breaks a tie between the action and the patient (cf. Goldin-Meadow et al. 2008) by creating an action gesture that is not dependent on the shape/form of the patient. This implies that the extent to which the action gesture by itself is bound to the patient should correlate with the proportion of SVO gestures. A new group of Japanese speakers (N = 32) saw a list of 15 verbs (as in the complexity questionnaire). They rated how much the gesture for the action changed depending on the patient, again using a

7-point Likert scale (1 = no change to 7 = much change). The raw scores were z-score transformed, and the mean "boundedness" scores were calculated for each verb (action). If patient-boundedness plays a role in influencing SVO gestures, the proportion of SVO gestures should negatively correlate with the boundedness score.

The models were initially fit with the maximal random effects structure, including complexity and patient-boundedness scores as covariates. The backward selection procedure determined the model without any random slope was assumed to be the optimal one (Bates et al. 2015a).

## 3.2 Results

Figure 5 shows the overall distribution of constituent orders by condition in Experiment 2, averaged across participants. There were more SVO gestures in general than SOV gestures across the conditions. A number of gesture productions in the Object-Human condition were categorized as "others" because they had just two gestures, and some had repetitive gestures.

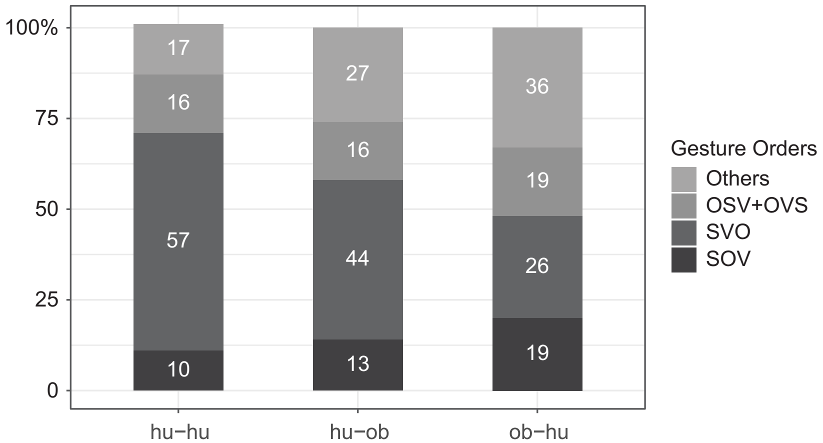



**Figure 5:** Distribution of constituent orders.

The mean SVO proportion in the Human-Object condition was significantly larger than the one in the Object-Human condition ($\beta = 0.32$, $z = 2.00$, $p < .05$) (see Table 2). The SVO gestures were produced significantly more often in the Human-Human condition than in the two conditions with an inanimate entity ($\beta = 0.39$, $z = 4.32$, $p < .001$). The pattern of increased SVO production in the Human-Human condition is quite similar to that observed in previous studies.

**Table 2:** Summary of the statistical analysis in Experiment 2.

| | Estimate | Standard Error | *z* value | *p* value | |
|---|---|---|---|---|---|
| (Intercept) | 1.06 | 0.35 | 3.01 | | |
| obj-hu vs. hu-obj | 0.32 | 0.16 | 2.00 | < 0.05 | * |
| obj.involved vs. hu | 0.39 | 0.09 | 4.32 | < 0.001 | *** |
| Complexity | −0.31 | 0.28 | −1.11 | 0.27 | |
| Boundedness | −0.54 | 0.31 | −1.73 | < 0.09 | . |

We were also interested in whether the proportion of SVO gesture production in general correlates with the gesture complexity score or the patient-boundedness score. A separate multiple regression model was built using only these two scores as dependent variables. The model indicated that while the patient-boundedness score significantly affected the SVO proportions negatively ($\beta$ = –0.19, SE = 0.05 $t$ = –3.54, $p$ < .005, see Figure 6a), the complexity score did not ($\beta$ = –0.09, SE = 0.05 $t$ = –1.44, *ns*, see Figure 6b). This indicates that the SVO proportions decreased as the patient-boundedness score increased; the participants tended to avoid using SVO gestures when the gesture for the action may have changed considerably depending on the properties of the patient.

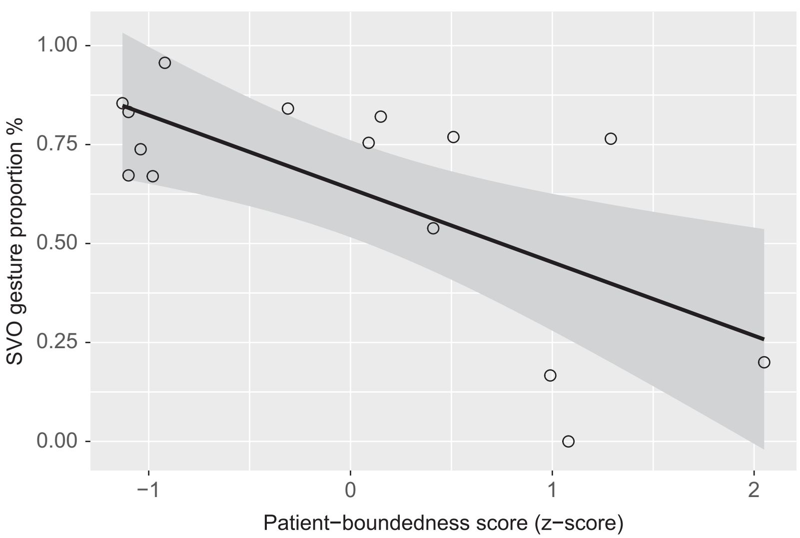



**Figure 6a:** Correlation between SVO gesture production rates and the patient-boundedness score.

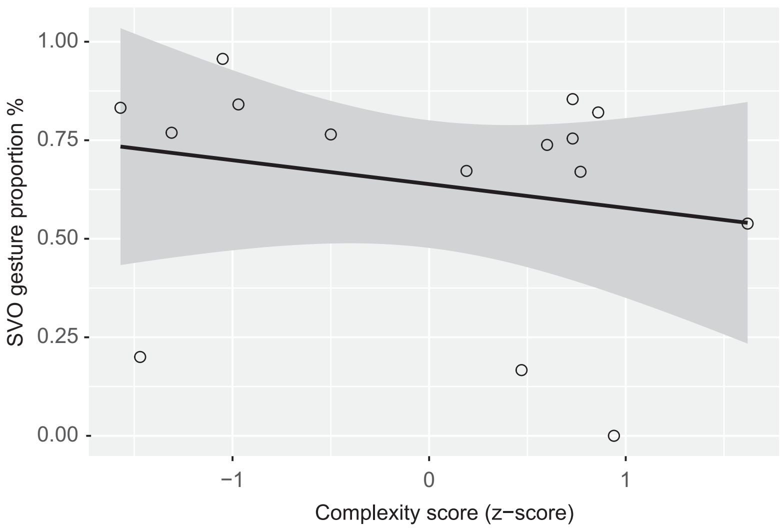



**Figure 6b:** Correlation between SVO gesture production rates and the complexity score.

Concerning individual variation, Figure 7 suggests that the variation among participants in Experiment 2 was much smaller than in Experiment 1. The SVO-biased scores were calculated as in Experiment 1, and the distribution of the SVO-biased scores in Experiment 2 was narrower than those in Experiment 1. Twenty-six participants produced more SVO in the Human-Human condition than in the Human-Object condition. However, only nine participants produced more SVO in the Human-Object condition than in the Human-Human condition. In sum, there were generally more SVO gesture productions in the Human-Human condition than in the Human-Object condition, and this observation applied to most of the participants.

## 3.3 Discussion

In contrast to Experiment 1, the participants practiced their gestures, especially the actions used in the stimuli before the task. Their overall SVO gesture production rates increased, similar to those observed in previous research (Hall, Ferreira, and Mayberry 2014; Marno et al. 2015). In previous studies, only SO-type languages were examined (e.g., Italian and Turkish in Marno et al. (2015)). Our results show that the observation holds even for native speakers of Kaqchikel, an OS-type language, who also speak Spanish. Regarding the connection between the SVO increase and conducting the gesture practice, Langus and Nespor (2010) suggested

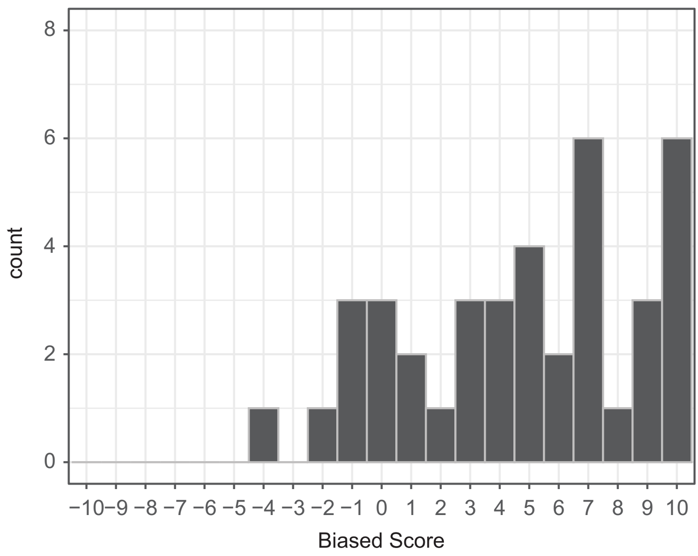



**Figure 7a:** Distribution of SVO-biased score in the hu-hu condition.

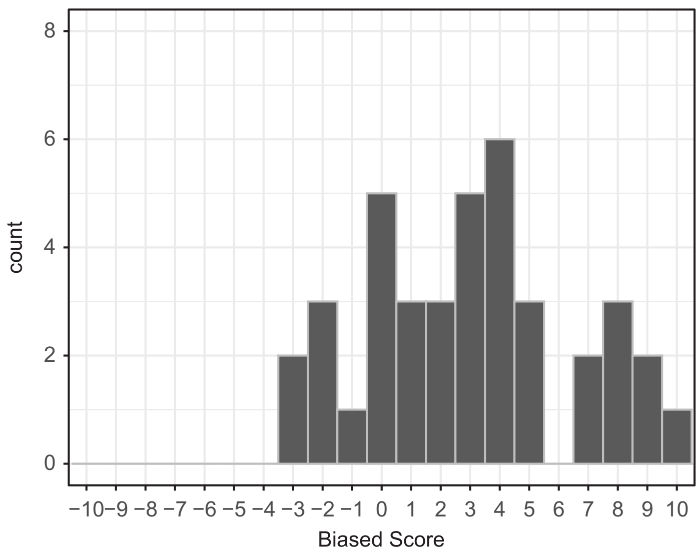



**Figure 7b:** Distribution of SVO-biased score in the hu-ob condition.

that practicing makes the gestures instances of lexical items, inviting involvement by the computational system of grammar. They argued that the grammatical system prefers the SVO order, contributing to the increase in SVO in gesture production.

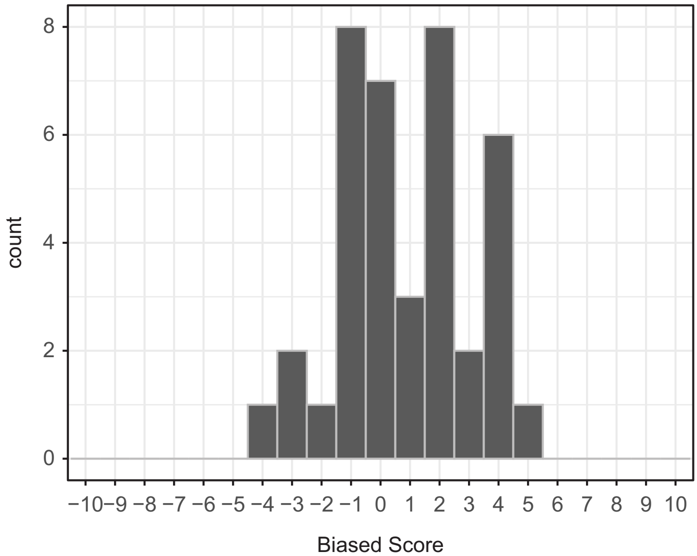



**Figure 7c:** Distribution of SVO-biased score in the ob-hu condition.

Our results indicate that the preference for SVO gesture order was visible in the Human-Human and Human-Object conditions, but not in the Object-Human conditions. If there was a strong connection between practicing and the involvement of the computational system of grammar, we should have observed the preference for SVO gesture order, even in the Object-Human condition.

Although we do not have any solid evidence to reject their hypothesis, we alternatively suggest that by practicing the gestures before the task, the action gestures become independent of the patient's type, shape, or size. This patient-boundedness account predicts that, when an action has a high degree of patient-boundedness, the participants tend to divert from using SVO gestures for the events with such actions. Concerning the general SOV preference for gesture production, Goldin-Meadow et al. (2008) suggested that the action gesture is placed after gestures for entities (S and O) because the action is cognitively more complex. They predicted that the degree of action complexity correlates with the proportion of SVO gestures. More specifically, they expected that there would be more SVO gestures when the action gesture is viewed as less complex.

Our multiple regression model indicates that the complexity of an action does not correlate with the overall proportion of SVO gestures. Instead, a high degree of patient-boundedness of action leads to reduced SVO gesture production. This finding implies that practicing gestures breaks the tie between the action and the patient. The throwing gestures, for instance, can take various forms, in principle.

Still, through practice, the participants decided to use a general form of throwing gesture, which takes a form independent of the patient's type, shape, or size, resulting in less motivation to place the action gesture after the patient. One may suspect that there was no complexity effect because the practice in this experiment removed the influence of the complexity of the action gestures. It is difficult to eliminate this possibility without manipulating this practice factor in one carefully controlled experiment. However, if the effect of the practice is so strong, we think it is quite odd to continue to observe the patient-boundedness effect. Thus, it seems fair to conclude that patient-boundedness plays a more decisive role than complexity of action in determining the choice between SVO and SOV gesture orders.

The Object-Human condition is one crucial addition in Experiment 2, where an inanimate entity causes some effect on a human entity. Compared to the other conditions in the experiment, where a human entity is an agent, the Object-Human condition did not increase SVO gesture production. Under the noisy-channel hypothesis, this is unexpected. Their account claims that SVO gestures increase when an event is ambiguous, particularly when the participants need to encode, utilizing relative ordering against the action gesture, of which the entity is an agent (and the other one is a patient) in the event.

Such ambiguity typically arises for events in which both entities are animate (e.g., human). Still, the Object-Human condition in our experiment also requires the participants to identify the agent of the event because the human entity, which is higher in an animacy hierarchy than the inanimate entity, is actually the patient. Therefore, the noisy-channel hypothesis predicts that there should be many SVO gesture productions, but that was not what we observed. The proportion of SVO was lowest in the Object-Human condition compared to the other conditions.

In contrast, the results are compatible with the role-conflict approach. According to Hall, Mayberry, and Ferreira (2013), SVO gestures increase because participants prefer to take the agent's perspective and make action gestures from the same perspective. Note that actions in the Object-Human condition are (at least partially) not the ones that need hand-gestures from an agent's perspective (e.g., (a tire) bumping against someone, (a clock) waking up someone, (a bike) knocking down someone, (wind) blowing someone)). These gestures are also different from those with manual motions (e.g., pulling by hand, pushing by hand, poking with fingers). Because actions caused by an inanimate entity do not motivate the participants to take the inanimate entity's perspective, the role-conflict approach predicts that the participants will not produce many SV chunks and that SVO gestures will not increase. Moreover, this approach indicates that the participants take a patient's perspective because it is human and may produce action gestures from their point of view (i.e., they may have produced not a pulling action but a "being pulled" action). If so, this contributes to an increase in SOV gesture production. The results from the Object-Human condition support their prediction. Future research should further investigate the effects of action types.

## 4 General discussion and summary

This study investigated gesture production processes with native speakers of Kaqchikel, a typologically quite distinct language. Unlike previous studies, there was no strong SOV preference. Individual data, however, indicated that there was sizable between-participant variation concerning the SVO/SOV bias. A certain number of participants showed SOV bias as seen in previous studies, but at the same time, others displayed a strong SVO bias. Unfortunately, the source of between-participant variations remains for future studies.

Concerning the claim that SOV is a natural sequence of representing events (Goldin-Meadow et al. 2008), it is noteworthy that numerous gesture sequences with S-before-O were found even in speakers of a language whose basic word order is VOS. This finding shows a powerful influence based on the agent advantage in representing an event. In addition, since Experiment 2 employed an extensive gesture-practice, inviting some characteristics of the grammar, a certain number of gestures with O-before-S were found. The number of these gesture sequences was not as plentiful as S-before-O, but they were not observed in previous research using SO-type languages. Therefore, it seems that the advantage of S over O is natural but not entirely free from the influence of language. It would be very interesting to examine the extent to which, for OS-type languages, gestures with O-before-S are also a viable option for naturally representing events. The mismatch between basic word order and gesture production in native Kaqchikel speakers may indicate that OS-type languages are constantly exposed to conflicting pressures from human cognition and grammar, possibly making this language fragile and typologically rare.

The placement of the action gesture in the SOV order is another point that needs explanation. Goldin-Meadow et al. (2008) suggested that the complex nature of action gestures is responsible for their placement. One prediction from their account is that the complexity measure of the action in the event should correlate with the proportion of SVO/SOV gesture production. Our results, however, indicate that there was no such correlation. We also measured patient-boundedness for actions and found a significant correlation; when the form of the action gesture can change substantially depending on the various natural properties of the patient, the action gesture tends to be placed at the end of the gesture sequence, even when participants practiced gestures before the task. This suggests that the placement of the action gesture in the SOV order is strongly affected by the immediately preceding patient. We argue that this is why the proportion of SVO gestures increases with practicing gestures in general. When the participants practiced and determined how to perform action gestures, their gestures became independent from the properties of the patient and had a general form. This idea is somewhat similar to a well-known "symbol-grounding" problem (Harnad 1990; Hirsh-Pasek and Golinkoff 1996; Imai 2017, among others), in the sense that when children learn a label for an action, they initially assign a label to particular events with certain entities. Later, they revise the label through abstraction, and it becomes a function that works with various agents and patients. Practicing gestures seems similar to this process of abstraction in lexical learning.

This discussion implies that participants prefer to produce an SVO gesture sequence, but when the action gesture is very much dependent on the patient, they are forced to make an SOV gesture sequence. Previous research has shown that a particular event property (i.e., having two human entities) triggers a preference for an SVO gesture sequence. The two accounts, the noisy-channel hypothesis (Gibson et al. 2013) and the role-conflict approach (Hall, Mayberry, and Ferreira 2013), provide different explanations for the effects observed in reversible events. Experiment 2 had a condition where an inanimate entity was an agent (or cause), and a human entity was a patient. The results showed that there was no SVO increase in that condition. This pattern is compatible with the prediction of the role-conflict approach. Further studies should investigate the relationship between the perspective taking of the participants and the various event properties.

Taken together, this study showed that investigating typologically different languages is crucial for understanding the relationship between language and human cognition. Gesture studies provide a compelling testing ground for the intricate nature of participants' event comprehension and the potentially rich influence of language. Therefore, we strongly advocate the view that we should increase the number of languages examined in the field (Anand, Chung, and Wagers 2011; Henrich, Heine, and Norenzayan 2010).

**Acknowledgments:** The present study (at some stage) has been presented at various conferences and workshops. We are genuinely grateful to the audience for their helpful comments and encouragement. In particular, we would like to thank Shota Momma for his suggestions about measuring complexity and patient-boundedness for events, which provided a crucial argument for this study. We would like to thank the two anonymous reviewers for their thoughtful comments, which greatly helped us clarify the various ideas and descriptions in this chapter. We are also grateful to Franklin Chang and Itsuki Minemi, who gave us feedback and suggestions about how we analyzed the data. We thank Lolmay Pedro Oscar García Mátzar, Juan Esteban Ajsivinac Sian, and Filiberto Patal Majzul for their extensive support in conducting the experiments in beautiful Antigua, Guatemala. We also thank Yuko Nakatani for drawing fantastic pictures for our research. Part of this work was supported by Grant-in-Aid for Scientific Research (S) titled “A Field-Based Cognitive Neuroscientific Study of the Processing of OS-type Languages” (No. 22222001, 19H05589, PI: Masatoshi Koizumi), and Grant-in-Aid for Scientific Research (C) (No. 19K00586, PI: Hajime Ono) from the Japan Society for the Promotion of Science.

## References

Anand, Pranav, Sandra Chung, & Matthew Wagers. 2011. Widening the Net: Challenges for Gathering Linguistic Data in the Digital Age. NSF SBE 2020: Future research in the social, behavioral and economic sciences. Retrieved from https://www.nsf.gov/sbe/sbe_2020/ (accessed 26 May 2021)

Bates, Douglas, Reinhold Kliegl, Shravan Vasishth, & Harald Baayen. 2015. Parsimonious mixed models. Available online at: https://arxiv.org/abs/1506.04967.

Bates, Douglas, Martin Maechler, Ben Bolker, & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67. 1–48.

Eberhard, David M., Gary F. Simons, & Charles D. Fennig. 2019. *Ethnologue: Languages of the World*. 22nd edn. Dallas, Texas: SIL International.

England, Nora C. 1991. Changes in basic word order in Mayan languages. *International Journal of American Linguistics* 57. 446–486.

Gell-Mann, Murray & Merritt Ruhlen. 2011. The origin and evolution of word order. *Proceedings of the National Academy of Sciences of the United States of America* 108(42). 17290–17295. https://doi.org/10.1073/pnas.1113716108

Gibson, Edward, Steven T. Piantadosi, Kimberly Brink, Leon Bergen, Eunice Lim, & Rebecca Saxe. 2013. A noisy-channel account of crosslinguistic word-order variation. *Psychological Science* 24(7). 1079–1088. https://doi.org/10.1177/0956797612463705

Goldin-Meadow, Susan. 2003. *The Resilience of Language: What Gesture Creation in Deaf Children Can Tell Us about How All Children Learn Language*. New York: Psychology Press.

Goldin-Meadow, Susan, Wing Chee So, Aslı Özyürek, & Carolyn Mylander. 2008. The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences of the United States of America* 105(27). 9163–9168. https://doi.org/10.1073/pnas.0710060105

Hall, Matthew L., Victor S. Ferreira, & Rachel I. Mayberry. 2014. Investigating constituent order change with elicited pantomime: A functional account of SVO emergence. *Cognitive Science* 38(5). 943–972. https://doi.org/10.1111/cogs.12105

Hall, Matthew L., Rachel I. Mayberry, & Victor S. Ferreira. 2013. Cognitive constraints on constituent order: Evidence from elicited pantomime. *Cognition* 129(1). 1–17. https://doi.org/10.1016/j.cognition.2013.05.004

Harnad, Stevan. 1990. The symbol grounding problem. *Physica D: Nonlinear Phenomena* 42(1–3). 335–346.

Haspelmath, Martin, Matthew S. Dryer, David Gil, & Bernard Comrie. 2005. *The World Atlas of Language Structures*. Oxford: Oxford University Press.

Henrich, Joseph, Steven J. Heine, & Ara Norenzayan. 2010. The weirdest people in the world? *Behavioral and Brain Sciences* 33(2–3). 61–83. http://dx.doi.org/10.1017/S0140525X0999152X.

Hirsh-Pasek, Kathy & Roberta M. Golinkoff. 1996. *The Origins of Grammar: Evidence from Early Language Comprehension*. Cambridge, MA: MIT Press.

Imai, Mutsumi. 2017. The "symbol grounding problem" reinterpreted from the perspective of language acquisition. In Jordan Zlatev, Göran Sonesson, & Piotr Konderak (eds.), *Meaning, Mind and Communication: Explorations in Cognitive Semiotics*, 145–159. Frankfurt am Main: Peter Lang. https://doi.org/10.3726/978-3-653-04948-0

Jaeger, T. Florian. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59(4). 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Koizumi, Masatoshi, Yoshiho Yasugi, Katsuo Tamaoka, Sachiko Kiyama, Jungho Kim, Juan Esteban Ajsivinac Sian, & Lolmay Pedro Oscar García Mátzar. 2014. On the (non)universality of the preference for subject-object word order in sentence comprehension: A sentence-processing study in Kaqchikel Maya. *Language* 90(3). 722–736. https://doi.org/10.1353/lan.2014.0068

Kubo, Takuya, Hajime Ono, Mikihiro Tanaka, Masatoshi Koizumi, & Hiromu Sakai. 2015. Kaqchikel-go VOS-gojun-no sansyutu mechanism: Yuuseesee-ga gojun-no sentaku-ni ataeru kooka-o toosite [A production mechanism for Kaqchikel VOS word order: Animacy effects toward word order selection]. *Cognitive Studies* 22(4). 591–603. https://doi.org/10.11225/jcss.22.591

Langus, Alan & Marina Nespor. 2010. Cognitive systems struggling for word order. *Cognitive Psychology* 60(4). 291–318. https://doi.org/10.1016/j.cogpsych.2010.01.004

Marno, Hanna, Alan Langus, Mahmoud Omidbeigi, Sina Asaadi, Shima Seyed-Allaei, & Marina Nespor. 2015. A new perspective on word order preferences: the availability of a lexicon triggers the use of SVO word order. *Frontiers in Psychology* 6. 1–8. https://doi.org/10.3389/fpsyg.2015.01183

Meir, Irit, Mark Aronoff, Carl Börstell, So-One Hwang, Deniz Ilkbasaran, Itamar Kastner, Ryan Lepic, Adi Lifshitz Ben-Basat, Carol Padden, & Wendy Sandler. 2017. The effect of being human and the basis of grammatical word order: Insights from novel communication systems and young sign languages. *Cognition* 158. 189–207. https://doi.org/10.1016/j.cognition.2016.10.011

R Core Team. 2019. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, https://www.R-project.org/.

Rodríguez Guaján, José Obispo. 1994. *Rutz'ib'axik ri Kaqchikel: Manual de Redacción Kaqchikel*. Guatemala City: Cholsamaj.

Sandler, Wendy, Irit Meir, Carol Padden, & Mark Aronoff. 2005. The emergence of grammar: Systematic structure in a new language. *Proceedings of the National Academy of Sciences of the United States of America* 102(7). 2661–2665. https://doi.org/10.1073/pnas.0405448102

Sapir, Edward. 1921. *Language: An Introduction to the Study of Speech*. New York: Harcourt, Brace and Company.

Whorf, Benjamin Lee. 1956. *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*. In John B. Carrol (ed.). Cambridge: Technology Press of Massachusetts Institute of Technology.

Takuya Kubo

# Chapter 6
# Factors affecting the choice of word order in Kaqchikel: Evidence from discourse saliency

## 1 Introduction

The act of producing a sentence has the effect of conferring an appropriate linguistic form on a nonverbal message. However, some languages have multiple corresponding linguistic forms for a single event. The speaker must instantly recognize one linguistic form from among several possibilities. For example, in Kaqchikel, in which word order alternation is relatively free, the same event can be produced in subject-verb-object (SVO) and verb-object-subject (VOS) word orders. What principle motivates the choice of word order in Kaqchikel? Is it similar to the principles of word order selection observed in other languages?

Gundel (1988) proposes the Given Before New Principle and the First Things First Principle as rules for word order selection from the viewpoint of discourse analysis. The former requires given information to precede any new information. The latter is characterized by requests for urgent or important information first. These principles are described as independent but occasionally contradictory. For example, if the "topic" in the preceding context is continued in the current sentence, the word order of the topic and the "comment" (= the state or activity of the topic) becomes problematic. In this case, the topic is given information, so the Given Before New Principle requires that the topic precede the comment. In contrast, the First Things First Principle requires that the comment, which is more important information, precede the topic – a word order that is more informative.

**Acknowledgments:** I would like to express our deepest gratitude to Lolmay Pedro Oscar García Mátzar, Juan Esteban Ajsivinac Sian, Filiberto Patal Majzu, and Melanio Cuma Chávez for their cooperation in recruiting the participants, conducting the experiment, and analyzing the data. We also thank Yuko Nakatani for providing line drawings for the experiments. This work was supported by JSPS KAKENHI Grant Numbers 22222001 and 19H05589 (PI: Masatoshi Koizumi) and the JSPS Grant-in-Aid for JSPS Fellows awarded to Takuya Kubo. An earlier version of this chapter has appeared as Kubo (2014) in Japanese. I would like to thank the Graduate School of Education, Hiroshima University, for their permission for reproduction.


https://doi.org/10.1515/9783110778946-006

While the tendency to obey the Given Before New Principle has been observed in previous sentence production studies (Arnold et al. 2000; Bock and Irwin 1980; Ferreira and Yoshita 2003), no studies have reported a tendency to follow the First Things First Principle. This is because the First Things First Principle is not treated as a central issue in the first instance. The two principles may behave differently depending on the language. For example, Herring (1990) argues that the order of the subject (S) and the verb (V) in basic word order determines the preference for the order of topic and comment. Specifically, in sentences with the same topic as in the preceding context, the word order in SV languages tends to be "topic-comment," whereas in verb-subject (VS) languages word order tends to be "comment-topic." In other words, SV languages follow the Given Before New Principle, while VS languages follow the First Things First Principle.

The languages that have been experimentally investigated to date include Japanese, whose basic word order is SOV (Ferreira and Yoshita 2003), and English, whose basic word order is SVO (Arnold et al. 2000; Bock and Irwin 1980). Both Japanese and English are considered SV languages. In contrast, VOS is the basic word order of Kaqchikel (Rodríguez Guaján 1994; Ajsivinac Sian et al. 2004), although the SVO word order has been reported frequently in recent years (Brown, Maxwell, and Little 2006; England 1991). Therefore, Kaqchikel is a VS language, and its VOS word order may be motivated by the First Things First Principle.

In this chapter, we aimed to experimentally determine which principle – the Given Before New Principle or the First Things First Principle – plays a greater role in the choice of VOS and SVO word order in Kaqchikel. An experimental approach has the advantage of allowing us to control for factors that may be implicated in word order selection. Through experimentation, we further identified that the choice of word order in Kaqchikel involves the First Things First Principle, a finding not previously observed.

# 2 Previous studies

## 2.1 The Given Before New Principle and the First Things First Principle

Gundel (1988) proposed the Given Before New and the First Things First Principles, because all languages include sentence structures that satisfy these principles. The Given Before New Principle predicts that elements that are given are produced prior to elements that are new, where "given" is defined by Prince (1981) as the listener's knowledge of or familiarity with their referents (Gundel 1988). Therefore, givenness does not necessarily correspond to the distinction of whether or not an element has been mentioned in the preceding context but is established if the element has already been mentioned. We obtained data supporting the Given Before New Principle through experimental investigation, the details of which are discussed in Section 2.2.

Givón (1991) also proposed the Pragmatic Principle of Linear Order (1), a rule similar to the First Things First Principle.

(1) Pragmatic Principle of Linear Order
  a. More important or more urgent information tends to be placed first in the string.
  b. Less accessible or less predictable information tends to be placed first in the string.

Data supporting the Pragmatic Principle of Linear Order are derived from studies that have analyzed word order by topicality (Fox 1983; Givón 1983b, 1983c). Topicality explains noun phrases not from the binary viewpoint of whether or not noun phrases are topics or not but as a continuous relation and is measured by how the same referent is distributed in the preceding and following contexts. For example, one of the indicators of topicality is the distance from the occurrence of the same referent in the preceding context, such that the shorter the distance, the higher the topicality while the longer the distance, the lower the topicality. The relationship between topicality and word order is as follows (Givón 1983a: 20):

(2) The relationship between topicality and word order

| | | | |
|---|---|---|---|
| COMMENT > | COMMENT-TOPIC > | TOPIC-COMMENT > | TOPIC |
| (zero topic) | | | (zero comment) |

From left to right, the order is "comment only," "comment-topic order," "topic-comment order," and "topic only," where the left end of the scale indicates more topical and the right end less topical. For example, if the topic of a sentence is highly topical, such as when it is mentioned in the previous sentence, it is omitted or follows the comment because it is predictable and less urgent. In contrast, when the topic is distant in the preceding context or less topical – such as when the topic is different from the one in the immediately preceding sentence – the topic tends to be mentioned again before the comment because it is less predictable to listeners.

In the following text, the term "urgency principle" is used as a cover term for Gundel's (1988) First Things First Principle and Givón's (1991) Pragmatic Principle of Linear Order. In parallel, we refer to Gundel's (1988) Given Before New Principle as the "givenness principle."

## 2.2 The givenness principle and a psycholinguistic approach

The givenness and urgency principles are related to the production of sentences. However, only the tendency to follow the givenness principle has been observed in previous studies of sentence production. In addition, the givenness principle is perceived as depending on the listener's knowledge in Gundel's (1988) definition, while the principle is explained from the speaker's perspective in sentence production research. In several studies, sentence production is assumed to be incremental and proceeds in parallel with available elements without waiting for memory retrieval of all sentence elements (Kempen and Hoenamp 1987; Levelt 1989). The advantages of such a processing method are that the production process is performed efficiently and the load on limited working memory capacity is evenly distributed (Branigan, Tanaka, and Pickering 2008; Slevc 2011).

The accessibility (speed of memory retrieval) of elements in a sentence is involved in determining the linguistic form under such a process. In particular, elements with high accessibility tend to be assigned higher grammatical roles (e.g., subjects) than elements with low accessibility, and are produced earlier in the sentence (Bock and Werren 1985; Feleki and Branigan 1999; McDonald, Bock, and Kelly 1993; Prat-Sala and Branigan 2000; Tanaka et al. 2011; Yamashita and Chang 2001). Givenness is also a measure of accessibility, with given elements more accessible than new elements (Bock and Warren 1985). Therefore, the givenness principle is considered the result of accessibility, as given elements are processed more rapidly than new elements under incremental processing (Arnold et al. 2000; Bock and Irwin 1980; Ferreira and Yoshita 2003).

For example, Ferreira and Yoshita (2003) examined the effect of givenness on word order selection in Japanese. Participants were asked to listen to a target sentence (3a or 3b) containing a direct and indirect object, and to subsequently verbalize the sentence. In the target sentences, only the word order of the direct and indirect objects was different. Just before playback, an eliciting sentence (4a or 4b) containing either a direct or indirect object was presented to manipulate the givenness.[1]

(3) Target sentences

a. okusan-ga otetsudaisan-ni purezento-o okutta
housewife-NOM housekeeper-DAT present-ACC gave
'The housewife gave the housekeeper a present.'

b. okusan-ga purezento-o otetsudaisan-ni okutta
housewife-NOM present-ACC housekeeper-DAT gave
'The housewife gave a present to the housekeeper.'

**1** The following abbreviations are used in this chapter. COMPL; completive; ABS: absolutive; ERG: ergative; 3: third person; sg: singular; DET: determiner; NOM: nominative; ACC: accusative; DAT: dative.

(4) Eliciting sentences

a. okusan-ga otetsudaisan-ni kanshashiteita
housewife-NOM housekeeper-DAT was grateful
'The housewife was grateful to the housekeeper.'

b. okusan-ga purezento-o katta
housewife-NOM present-ACC bought
'The housewife bought a present.'

This task requires the process of reconstructing the linguistic form based on the meaning of the target sentence stored in memory. The effect of givenness on word order was measured by the percentage of misplays. The results of the experiment illustrated that after (4a), (3b) was misplayed as (3a) at a higher rate than (3a) was misplayed as (3b), and the opposite was true after (4b). This finding indicates that the given information is more likely to be produced before new information is available. In addition, Prat-Sala and Branigan (2000) showed that the tendency for given information to precede new information is due to the relative discourse saliency of noun phrases rather than to the binary relationship between given and new information. The target language was Spanish, which alternates between the SVO and object-verb-subject (OVS) word orders in active sentences. In the experiment, a picture description task was performed while the preceding context was manipulated. For each picture, two contexts were created in which both an agent and a patient appeared. The agent was given greater prominence in one context while the patient was given importance in the other. Specifically, the salient entity was first introduced by an existential sentence and further modified by several adjectives and participles. A non-salient entity was introduced after the salient entity, and no modifying elements were added. The findings demonstrated that OVS word order was more frequently used in the condition in which the patient's saliency was increased than in the condition in which the agent's saliency was increased. Prat-Sala and Branigan (2000) explained this result in terms of accessibility. Specifically, the authors asserted that, because of the greater accessibility of the patient, the patient was produced earlier in the sentence when the patient's saliency was increased than when the agent's saliency was increased.

## 2.3 Remaining issues and special features of this study

As mentioned earlier, experimental studies to date have not demonstrated a tendency for language to follow the urgency principle. Notably, however, the urgency principle might have been in play in the experimental design of both studies summarized above. In other words, the elements mentioned or elevated to prominence in the preceding context were less urgent information, and therefore, the urgency principle required that these elements be produced further back in the sentence. Needless to say, the urgency principle assumes a listener, while an actual listener was absent in the study above. However, the principle might nevertheless have applied, given that the speaker is also a potential listener (Gibson et al. 2013).

Why have we not observed a tendency to follow the urgency principle in these studies? While Givón (1991) presents the urgency principle (the Pragmatic Principle of Linear Order) as universal to human languages, disagreement with this broad generalization exists. Myhill (1992) classifies languages in terms of verb and object word order, and states that the only languages that follow this principle are those with free verb-object (VO) word order and those with strict VO word order, in which the subject appears mainly at the end of a sentence. Moreover, as mentioned in Section 1, Herring (1990) argues that SV and VS languages have different preferences for the order of topic and comment. VS languages tend to follow the urgency principle when the subject is continuous from the preceding context. Therefore, a possible reason for the lack of observation to date of the urgency principle is that the languages under consideration have been SV languages and the word orders considered have not been fundamentally motivated by the urgency principle.

The language of interest in this chapter is Kaqchikel. Consistent with languages that follow Myhill (1992) and Herring's (1990) urgency principle, the basic word order of Kaqchikel is VOS. The VOS word order in Kaqchikel may be motivated by the urgency principle when the predictability of the subject is low. However, since SVO word order is also used frequently (England 1991), the actual situation is unclear. Therefore, this study aims to clarify how the givenness and urgency principles are involved in the choice of VOS and SVO word order in Kaqchikel.

## 3 Characteristics of the Kaqchikel language and experimental design

### 3.1 Linguistic features of Kaqchikel

Kaqchikel is one of the 21 Mayan languages spoken in Guatemala, with an estimated 450,000 speakers (Brown, Maxwell, and Little 2006). Like many other Mayan languages, Kaqchikel is considered to have a basic VOS word order (Rodríguez Guaján 1994; Ajsivinac Sian et al. 2004). The SVO word order in Kaqchikel is believed to have been in use since ancient times (England 1991). The following are example sentences of the VOS (5a) and SVO (5b) word orders:[2]

(5) 'The girl hit the boy.'

a. X-Ø-u-ch'äy ri ala' ri xtän
   COMPL-ABS3SG-ERG3SG-hit DET boy DET girl

b. Ri xtän x-Ø-u-ch'äy ri ala'
   DET girl COMPL-ABS3SG-ERG3SG-hit DET boy

Similar to other Mayan languages, Kaqchikel is a head-marking and morphologically ergative language. Subjects and objects are not overtly case-marked for grammatical relations; rather, grammatical relations are obligatorily marked on the predicate with two sets of person/number agreement morphemes, one denoting a transitive subject and the other denoting a transitive object and an intransitive subject. The order of the morphemes is [Aspect-ABS-ERG-Verb stem] for transitive verbs. Kaqchikel is also a pro-drop language, which signifies that noun phrases are not normally pronounced if they are obvious in the context, and verbs alone have sentence-equivalent functions. The examples above illustrate that the alternation of word order in Kaqchikel does not involve alternation of forms or the addition of special morphemes. However, the alternation of word order is not always possible and may be constrained by the definiteness or animacy of the noun phrases. For example, in Patzicia, Guatemala, the SVO word order is obligatory, and the VOS word order is reported to be ungrammatical when the subject is an indefinite noun phrase (Broadwell 2000). In Patzún, Guatemala, VOS word order is obligatory when the subject and object are definite noun phrases with equal animacy or when the subject and object are both indefinite noun phrases (Kim 2011).

## 3.2 Experimental design

In the experiment, we determined which principle played a greater role in the choice of SVO and VOS word order in Kaqchikel in an environment in which both the givenness and urgency principles could be involved. Specifically, by increasing the accessibility of the agent or patient in the antecedent context, we established the context for testing our hypothesis. We predicted that elements with increased accessibility in the antecedent context were more likely to be produced earlier in the sentence under the givenness principle, while we expected these elements to

2 [Ø] in the Kaqchikel examples indicates a phonetically null morpheme.

be produced later in the sentence under the urgency principle because they were redundant elements with high predictability.

In addition to the above, as the experiment in this study explores the principle of word order selection, we considered the constraints described in Section 3.1. Specifically, in an environment that obligatorily allows only SVO or VOS word order, determination of the governing principle becomes impossible, and such an environment must first be eliminated. For this reason, the study employed a picture description task that manipulates discourse saliency using the approach of Prat-Sala and Branigan (2000). In this method, both the agent and the patient are introduced in the preceding context to ensure that the subject and object of the sentence appear as definite noun phrases. In addition, we fixed the agent as animate and the patient as inanimate and established an environment in which SVO and VOS word orders could alternate. As in Prat-Sala and Branigan (2000), discourse saliency, a measure of accessibility, was also manipulated by the order of introduction of agents and patients in the context and the presence or absence of modifiers. Prat-Sala and Branigan (2000) further discussed the increase or decrease of OVS word order by manipulating the discursive congruence between agents and patients. The problem with this method is the lack of a baseline condition, so the condition in which the production of OVS word order is promoted or suppressed becomes unclear. Therefore, a neutral condition in which neither the agent nor the patient appears was established as the baseline of this study. In the analysis, we examined the effect of the neutral condition by comparing the neutral condition with the agent-salient and patient-salient conditions.

While we controlled discourse saliency in the experiment, the manipulation was not strictly a measure of topicality as described by Givón (1983a). However, this operation appeared suitable for testing the urgency principle in two respects. First, saliency is correlated with topicality and elements with higher saliency are more likely to be topicalized (Levelt 1989). Since the manipulation of discourse saliency affects the accessibility of the speaker (or listener), discourse saliency may also correlate with predictability.

If the givenness principle plays a larger role in the choice of word order in Kaqchikel, we would expect the elements with the highest intelligibility to be produced earlier in the sentence, as in Prat-Sala and Branigan (2000). In other words, the production of VOS word order is expected to be accelerated in the patient-salient condition, while the production of SVO word order is expected to be accelerated in the agent-salient condition. On the other hand, if the urgency principle plays a major role, more accessible elements are expected to be produced later in the sentence. Thus, VOS word order is more likely to be produced in the agent-salient condition, while SVO word order is more likely to be produced in the patient-salient condition.

# 4 Experiment

## 4.1 Participants

Forty Kaqchikel speakers (18 male and 22 female; mean age: 35.8 years) living in Guatemala participated in the experiment.

## 4.2 Materials

We created 24 line drawings that could be expressed with transitive sentences as target stimuli and 18 line drawings that could be expressed with intransitive sentences as filler stimuli. The agent was animate for all target stimuli while the patient was inanimate. In addition, six pictures of human, animal, and inanimate subjects were used as filler stimuli for diversity in animacy.

For each target stimulus, we created three types of contexts by manipulating saliency: an agent-salient condition, a neutral condition, and a patient-salient condition. Although both the agent and the patient appear in the context of the agent-salient and patient-salient conditions, the saliency was manipulated as follows: The salient element was introduced first by the existential construction, and the non-salient element was introduced after the salient element. In addition, the salient element was modified by three or more adjectives and participles, but no modifiers were added to the non-salient element. Neither the agent nor the patient of the target sentence appeared in this context in the neutral condition.

The following (6–8) are examples and their English translations of contexts corresponding to the “girl throws a stone” event.

(6) Agent-salient condition
K’o jun ko’öl chuqa’ k’aqät **xtän** (‘girl’) nib’iyin chunaqaj jun *ab’äj* (‘stone’), xukanoj jun wachinäq richin nretz’ab’ej qa. ¿Achike xk’ulwachitäj?
‘A small, inquisitive girl was walking near a stone. She is looking for something to play with. What happened?’

(7) Patient salient condition
K’o jun ko’öl chuqa’ qoloqöj **ab’äj** (‘stone’) pa b’ey chuwäch jun *xtän* (‘girl’), ri tz’il chuqa’ nojinäq chi ulew. ¿Achike xk’ulwachitäj?
“There was a small, rough stone on the road near the girl. The surface is stained with dirt. What happened?”

(8) Neutral condition
K'o jun li'aj chuqa' silan k'ayibäl akuchi' k'o jujun taq ch'akat. ¿Achike xk'-ulwachitäj?
'There was a large, quiet square. There are some chairs in it. What happened?'

Contexts were also created for the filler stimuli. In two-thirds of the filler stimuli, the saliency of the referent to be the subject in the target sentence was increased in the context, and in the remaining one-third of the stimuli, the context was neutral. These contexts were read out loud at a natural speed by a male native speaker of Kaqchikel and recorded on an integrated circuit recorder for use in the experiment. We created three presentation lists, one for each of the three contexts. Each list contained all the target and filler stimuli, but the contexts corresponding to each target stimulus were arranged differently in each list. The stimuli in the list were presented randomly to each participant.

## 4.3 Procedure

The experiment was conducted individually with native speakers of Kaqchikel in the following order: instruction, practice trials, and main trials. The instructions included the following four points: 1) "Produce them in natural Kaqchikel language as you would normally speak," 2) "There is no single sentence that is the correct answer," 3) "Utter the sentence that comes to mind without thinking too much," and 4) "Mention all the characters." Given the experiment's goal of examining the effect of context on the choice between SVO and VOS word orders, the fourth instruction point was intended to prevent the omission of the agent or the patient.

In the practice trials, we employed three pictures similar to the target stimuli and three pictures similar to the filler stimuli, and each context was presented in the same proportion as in the main trial. Stimulus pictures in the practice trials were not employed in the main trial. In the experiment, we first presented the context auditorily with the fixation cross, and the stimulus picture was automatically presented when the context ended. Participants were instructed to press the space bar to proceed to the next trial after completing their verbal description of the stimulus picture. The experiment lasted approximately 20 minutes per person.

## 4.4 Analysis

Native speakers of Kaqchikel performed the transcription and coding of speech data. All utterances in active sentences other than those in the SVO or VOS word order were excluded from the analysis. Even if utterances contained these word orders, statements that did not properly express the event were excluded from the analysis. Excluded utterances accounted for 37.8% (363 of 960) of utterances.[3]

We conducted tests for differences in the proportion of VOS word order produced in each condition using mixed logistic regression analysis (Jaeger 2008), including context type as a fixed effect and participant and stimulus picture as random effects.[4]

## 4.5 Results

Table 1 illustrates the appearance of SVO and VOS word orders in each condition. The overall trend was that SVO word order (75%) appeared more often than VOS word order (25%). The main effect of context on the percentage of VOS word order produced was not significant ($\beta = 0.23$, SE = 0.40, $z = 0.60$, $p > .1$) when comparing the patient-salient and neutral conditions. On the other hand, when we compared the percentage of VOS word order produced for the agent-salient condition to that for the neutral condition, the main effect of context was significant ($\beta = -0.95$, SE = 0.40. $z = -2.40$, $p < .05$). In other words, the result supports that the production of VOS word order is facilitated in the agent-salient condition.

**Table 1:** Appearance of SVO and VOS word orders in each condition.

| | **agent-salient** | **neutral** | **patient-salient** |
|---|---|---|---|
| SVO | 137 (68.9%) | 153 (79.3%) | 157 (76.6%) |
| VOS | 62 (31.1%) | 40 (20.7%) | 48 (23.4%) |
| total | 199 (100%) | 193 (100%) | 205 (100%) |

**3** The excluded data consisted of 2 utterances in OVS word order, 7 utterances in VSO word order, 61 utterances in intransitive sentences, 50 utterances in passive sentences, 45 utterances with subject omission, 32 utterances with object omission, 4 utterances with verb only, 10 utterances in reverse passive sentences, 2 utterances in split sentences, 145 utterances in other sentences, and 5 utterances with missing data.

**4** The statistical software R (ver. 3.0.2) was used for the analysis, and the package lme4 (ver. 1.1.6) was used to analyze the mixed logistic model.

# 5 Discussion

## 5.1 Involvement of the givenness and urgency principles in Kaqchikel

To clarify how the givenness and urgency principles are involved in the choice of SVO and VOS word order in Kaqchikel, we conducted a picture-description task that manipulated discourse saliency. The results of the experiment reveal that Kaqchikel speakers produced more VOS word order in the condition with increased saliency of the agent (i.e., the subject). If the givenness principle plays a major role in the choice of word order in Kaqchikel, then the agent-salient condition should promote the production of SVO word order, in which the salient subject is produced at the beginning of a sentence. However, the results supported the opposite tendency: Kaqchikel speakers produced more accessible and predictable agents at the end of sentences. The results therefore reveal that the production of the VOS word order in Kaqchikel is motivated by the urgency principle.

On the other hand, neither SVO nor VOS word order was facilitated in the patient-salient condition. While the reason for this finding is unclear at present, we state two possible explanations. First, the accessibility of the patient may not be involved in these word order choices. The alternation between SVO and VOS word order is seen as a choice to produce the subject at the beginning or end of the sentence, with the object always placed posterior to the verb. Therefore, the characteristics of the agent and not those of the patient appear to largely influence the choice of word order.

Another possibility is the asymmetry of accessibility between the agent and patient in the stimulus pictures. Agents and patients were animate and inanimate entities, respectively, in the stimulus pictures. However, animate entities have higher accessibility than inanimate entities, and agents have higher accessibility than patients (Bock and Warren 1985). In other words, the agent had higher accessibility than the patient even before the manipulation of discourse saliency. Therefore, the reason we did not observe a significant difference between the patient-salient and neutral conditions may be that the patient's accessibility did not exceed that of the agent. In any case, we assert that the impact of patient characteristics on word order choice in Kaqchikel must be further investigated in the future.

Using the same method as in previous studies, ours is the first study to demonstrate the psychological reality of the urgency principle, which has been overlooked in previous production studies. In addition, Gundel (1988) and Givón (1991) state that the urgency principle relates to accessibility and predictability for the listener. In this study, the task was to verbally describe stimulus pictures presented on the computer, not to make the pictures comprehensible to listeners. The results therefore suggest that the urgency principle, like the givenness principle, may be involved in situations in which a listener is absent.

Importantly, this study does not deny the involvement of the givenness principle in Kaqchikel. The experiment established a context in which both the givenness and urgency principles could be involved, and the environment was such that the two principles conflicted with one another. Therefore, the reason for the observed tendency to follow only the urgency principle may be that the effect of this principle was greater than that of the givenness principle, although both principles were involved.

## 5.2 Incrementality of sentence production in Kaqchikel

Research predicts that sentence production is incremental (Levelt 1989) and that elements with high accessibility are not only more likely to be processed earlier than those with low accessibility, but are more likely to be produced early in the sentence (Ferreira and Engelhardt 2006). In Kaqchikel, the frequency of SVO word order exceeded that of VOS word order, suggesting that sentence production in Kaqchikel is also based on incremental processing. As mentioned in Section 5.1, agents had higher overall accessibility than patients in the stimulus pictures used in the experiment. Therefore, the production of SVO word order can be considered the result of the incremental production of elements with high accessibility early in the sentence. On the other hand, the effect of the urgency principle of producing more instances of VOS word order in the agent-salient condition appears to be a phenomenon that violates the incrementality of sentence production because it produces highly accessible elements at the end of sentences. However, we note that the incrementality of sentence production and the urgency principle have different origins. The incremental nature of sentence production is based on general human cognitive features such as efficient processing and reduced load on working memory (Branigan, Tanaka, and Pickering 2008; Slevc 2011). In contrast, the urgency principle is determined by the linguistic features of individual languages (Herring 1990; Myhill 1992). Therefore, the observation of the urgency principle in Kaqchikel does not negate the general assumption of incrementality in sentence production, but rather suggests that Kaqchikel is a language in which both principles are involved. In sentence production, the final realized linguistic form is determined not only by a single factor but also by the competition among various factors (Bates and MacWhinney 1989; Yamashita and Chang 2001; Tanaka et al. 2011). Although the general tendency in Kaqchikel is to produce SVO word order under incremental processing, the urgency principle is believed to occasionally motivate the production of VOS word order in a way that is contrary to incremental processing. The results of this study show that these factors compete with one another in Kaqchikel when producing a linguistic form.

## 6 Summary and remaining issues

Previous studies of sentence production have drawn attention to the problem that although a universal sentence production mechanism has often been assumed, the languages studied have been typologically limited (Jaeger and Norcliffe 2009). Therefore, previous research may have only revealed specific aspects of human language. Against this background, our study employed experimental methods to demonstrate that the hitherto overlooked principle of urgency motivates the production of VOS word order in Kaqchikel. In the future, it will be necessary to determine languages in which the urgency principle plays a manifest role, and this is an issue that should be considered for various languages.

## References

Ajsivinac Sian, Juan Esteban, Lolmay Pedro Oscar García Mátzar, Martín Chacach Cutzal, & Ixkusamil Estela Alonzo Guaján. 2004. *Gramática descriptiva del idioma maya Kaqchikel: Rutzijoxik rucholik ri Kaqchikel ch'ab' äl. Academia de las Lenguas Mayas de Guatemala*. Chimaltenango, Guatemala: Comunidad Lingüística Kaqchikel.

Arnold, Jennifer E., Anthony Losongco, Thomas Wasow & Ryan Ginstrom. 2000. Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language* 76(1). 28–55.

Bates, Elizabeth & Brian MacWhinney. 1989. Functionalism and the competition model. In Brian MacWhinney & Elizabeth Bates (eds.). *The Crosslinguistic Study of Sentence Processing*, 3–73. Cambridge: Cambridge University Press.

Bock, J. Kathryn & David E. Irwin. 1980. Syntactic effects of information availability in sentence production. *Journal of Verbal Learning and Verbal Behavior* 19. 467–484.

Bock, J. Kathryn & Richard K. Warren. 1985. Conceptual accessibility and syntactic structure in sentence formulation. *Cognition* 21(1). 47–67.

Branigan, Holly P., Martin J. Pickering, & Mikihiro Tanaka. 2008. Contributions of animacy to grammatical function assignment and word order during production. *Lingua* 118(2). 172–189.

Broadwell, George Aaron. 2000. Word order and markedness in Kaqchikel. In *Proceedings of the LFG00 Conference*, 1–19. Stanford University: CSLI Publications.

Brown, R. McKenna, Judith M. Maxwell, & Walter E. Little. 2006. *La Ütz Awäch?: Introduction to Kaqchikel Maya language*. Austin, TX: University of Texas Press.

England, Nora. 1991. Changes in basic word order in Mayan languages. *International Journal of American Linguistics* 57(4). 446–486.

Feleki, Eleonora & Holly P. Branigan. 1999. Conceptual accessibility and serial order in Greek speech production. In *Proceedings of the 21st Annual Conference of the Cognitive Science Society*, 96–101. Mahaw, NJ: Lawrence Erlbaum Associates.

Ferreira, Fernanda & Paul E. Engelhardt. 2006. Syntax and production. In Matthew J. Traxler & Morton A. Gernsbacher (eds.). *Handbook of Psycholinguistics*, 61–91. New York: Academic Press.

Ferreira, Victor S. & Hiromi Yoshita. 2003. Given-new ordering effects on the production of scrambled sentences in Japanese. *Journal of Psycholinguistic Research* 32(6). 669–692.

Fox, Andrew. 1983. Topic continuity in biblical Hebrew narrative. In T. Givón (ed.), *Topic Continuity in Discpurse: A Quantitative Cross-Language Study*, 215–254. Amsterdam: John Benjamins.

Gibson, Edward, Steven T. Piantadosi, Kimberly Brink, Leon Bergen, Eunice Lim, & Rebecca Saxe. 2013. A noisy-channel account of crosslinguistic word-order variation. *Psychological Science* 24(7). 1079–1088.

Givón, Talmy. 1983a. Topic cocntinuity in discourse : an introduction. In T. Givón (ed.), *Topic Continuity in Discourse: A Quantitative Cross-Language Study*, 3–41. Amsterdam: John Benjamjns.

Givón, Talmy. 1983b. Topic continuity and word order pragmatics in Ute. In T. Givón (ed.), *Topic Continuity in Discourse: A Quantitative Cross-Language Study*, 141–214. Amsterdam: John Benjamins.

Givón, Talmy. 1983c. Topic continuity in spoken English. In T. Givón (ed.), *Topic Continuity in Discourse: A Quantitative Cross-Language Study*, 343–363. Amsterdam: John Benjamins.

Givón, Talmy. 1991. Isomorphism in the grammatical code: Cognitive and biological considerations. *Studies in Lang*uage 15(1). 85–114.

Gundel, Jeanette K. 1988. Universals of topic-comment structure. In Michael Hammond, Edith Moraucsik, & Jessica Wirth (eds.), *Studies in Syntactic Typology*, 209–239. Amsterdam: John Benjamins.

Herring, Susan C. 1990. Information structure as a consequence of word order type. In *Proceedings of the Sixteenth Annual Meeting of the Berkeley Linguistics Society* 16(1), 163–174.

Jaeger, T. Florian. 2008. Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59(4). 434–446.

Jaeger, T. Florian & Elisabeth J. Norcliffe. 2009. The cross-linguistic study of sentence production. *Language and Linguistics Compass* 3(4). 866–887.

Kempen, Gerard & Edward Hoenkamp. 1987. An incremental procedural grammar for sentence formulation. *Cognitive Science* 11. 201–258.

Kim, Sok-Ju. 2011. Word-order variables in Patzun Kaqchikel. *Kansas Working Papers in Linguistics* 32. 120–144.

Kubo, Takuya. 2014. Kakuchikerugo ni okeru gojyun no sentaku ni eikyo o ataeru yoin – danwateki takuritsuse o tegakari to shite – [Factors affecting the choice of word order in Kaqchikel: Evidence from discourse saliency]. *Bulletin of the Graduate School of Education, Hiroshima University*, Part 2, Vol. 63. 243–251.

Levelt, Willem J. 1989. *Speaking: From Intention to Articulation*. Cambridge: MIT press.

McDonald, Janet L., Kathryn Bock, & Michael H. Kelly. 1993. Word and world order: semantic, phonological, and metrical determinants of serial position. *Cognitive Psychology* 25. 188–230.

Myhill, John. 1992. *Typological Discourse Analysis: Quantitative Approaches to the Study of Linguistic Function*. Cambridge: Blackwell publishers.

Prat-Sala, Mercè & Holly P. Branigan. 2000. Discourse constraints on syntactic processing in language production: A cross-linguistic study in English and Spanish. *Journal of Memory and Language* 42(2). 168–182.

Prince, Ellen F. 1981. Toward a taxonomy of given-new information. In P. Cole (ed.), *Radical Pragmatics*, Vol. 3, 235–255. New York: Academic Press.
Rodríguez Guaján, José Obispo. 1994. *Rutz'ib'axik ri Kaqchikel: Manual de Redaccíon Kaqchikel*. Guatemala City: Editorial Cholsamaj.
Slevc, L. Robert. 2011. Saying what's on your mind: Working memory effects on sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37(6). 1503–1514.
Tanaka, Mikihiro N., Holly P. Branigan, & Janet F. McLean. 2011. Conceptual influences on word order and voice in sentence production: Evidence from Japanese. *Journal of Memory and Language* 65(3). 318–330.

Rei Emura

# Chapter 7
# Sentence comprehension in Central Alaskan Yup'ik: The effects of case marking, agreement, and word order

## 1 Introduction

In the study of sentence comprehension, the issue of how parsers form internal structures from surface forms has been explored for several decades. It has been suggested that grammatical relations are determined through parallel competition among several syntactic (e.g., word order and case marking), semantic (e.g., animacy and plausibility), and prosodic (e.g., stress and pitch) cues (Bates and MacWhinney 1982, 1989). However, previous studies have not clarified how multiple cues interact with each other in determining grammatical relations.

Focusing on Central Alaskan Yup'ik (hereinafter, Yup'ik), an ergative language with free word order, this chapter examines the effects of word order and its interaction with case marking and agreement cues on the judgment of grammatical relations. The acceptability judgment experiment presents evidence that word order serves as a cue, especially when the case marking is ambiguous. I also compare the results in Yup'ik with those in Japanese.

### 1.1 The effects of word order

Several languages permit basic constituents to be ordered in several ways. This chapter focuses on the sequence of subject (S), object (O), and verb (V) in a transitive sentence, which can be divided into two categories: (i) the relative order of S and O (i.e., SO vs. OS; *the argument order* hereinafter) and (ii) the position of V (verb-initial, verb-medial, verb-final).

Argument order has been shown to play an important role in sentence processing in various languages, including both SO- and OS-ordered languages with both accusative and ergative case marking systems (Bahlmann et al. 2007; Matzke et al.

**Acknowledgments:** This work was supported in part by JSPS KAKENHI Grant Number 19H05589. I would like to thank Masatoshi Koizumi (Tohoku University), Sachiko Kiyama (Tohoku University), and Al'aq (University of Alaska, Fairbanks) for their support and comments.


https://doi.org/10.1515/9783110778946-007

2002 for German; Chujo 1983; Mazuka, Itoh, and Kondo 2002; Tamaoka et al. 2005 for Japanese; Kaiser and Trueswell 2004 for Finnish; Sekerina 1997 for Russian; Erdocia et al. 2009 for Basque; Kiyama et al. 2013; Koizumi et al. 2014 for Kaqchikel Maya). These studies demonstrated that the canonical word order requires a lower processing load than the scrambled word order. Note that in this chapter, canonical word order refers to the syntactically simplest sequence of S, O, and V in a transitive sentence (Koizumi 2023, see also England 1991).

On the other hand the effect of verb position has not yet been well-studied. Basque showed a scrambled effect for SOV versus OSV, but no significant effect for SVO versus OVS (Laka and Erdocia 2012). This result suggests that the verb position may affect SO preference. From the perspective of cognitive tendency, it has been suggested that the most cognitively preferred word order of language universals is SOV rather than SVO because the actor-patient-action order, analogous to the SOV order, was used the most frequently to express transitive events by gesture (Goldin-Meadow et al. 2008). However, actor-action-patient and patient-actor-action were used more frequently in reversible conditions than in nonreversible ones. This may underlie the "role conflict model" that assumes that the patient-action order, or OV order, is avoided because patient-action leads to a conflict with actor-action (Hall, Mayberry, and Ferreira 2013; see also Koizumi 2023: 66–74 for review).

## 1.2 Interaction between word order, case marking, and agreement

The use of morpho-syntactic cues, such as word order, case marking, and agreement, to determine grammatical relations is posited by two models: the "competition model" (Bates and MacWhinney 1982, 1989) and the "diagnosis model" (Fodor and Inoue 1994, 2000). The competition model assumes that cue validity predicts cue strength. In other words, readers/listeners rely on information that is accessible to them and that is unambiguous in sentence comprehension. For instance, Italian, which allows all possible word orders in informal conversation and has a rich S-V agreement system, strongly relies on the agreement cue rather than word order (MacWhinney et al. 1984). However, the interaction between word order and ambiguity in case marking and agreement has not been well examined. In German, case marking ambiguity induced a higher processing cost only in non-canonically argument-ordered sentences, as evidenced by a P600 component from electroencephalogram (ERP) and the activation of the left supramarginal gyrus from functional magnetic resonance imaging (Bahlmann et al. 2007; Matzke et al. 2002). Erdocia et al. (2009) further showed the same result for Basque, an ergative language, in an ERP experiment. However, the sentences used in these studies also included the ambiguity of verb agreement with a subject (and an object in Basque) as well as the ambiguity of case marking, leading to confusion regarding the effects of case marking and agreement. Thus, we need to distinguish between the effects of case marking and agreement and clarify which cue ambiguities interact with word order effects.

The diagnosis model focuses on the difference between case and agreement features, assuming that the case information is a “positive symptom,” which builds sentence structure, whereas the agreement information is a “negative symptom,” which simply invalidates the incorrect structure but does not say what is the subject or the object. For instance, in the study for German, reanalysis from garden-path was more effective when the sentences were disambiguated due to the case features rather than the number features, one of the verb agreement cues. On the other hand, the sentences with incorrect number features were more successfully rejected than the sentences with incorrect case marking (Meng and Bader 2000). The model supposes a language-universal mechanism of use of case and agreement cues but does not assume how the mechanism relates to word order information.

## 1.3 Some relevant properties of Yup’ik

Yup’ik is an Eskimo language spoken in southeast Alaska. The number of its speakers is approximately 10,000 (Lewis, Gary, and Charles 2016). The majority of native Yup’ik speakers also speak English as a second language.

Yup’ik has a free word order, allowing all six types of word orders, SOV, SVO, OSV, OVS, VOS, and VSO (Miyaoka 1986). However, according to typological studies on Yup’ik, SOV is the most “neutral” (Miyaoka 1986; Jacobson and Jacobson 1995; Mather, Meade, and Miyaoka 2002). A corpus study also showed that SOV is the most frequently used (Fortescue 1993). According to Fortescue’s data, which include two narrative texts and one speech, SOV was the most frequently used, followed by SVO in both genres. However, the proportion of SOV was somewhat smaller and that of SVO was larger in the speech than in the narratives. There were also OVS and OSV orders to some extent. This chapter refers to SOV and SVO as the canonical word orders of Yup’ik based on their frequency.

Yup’ik has an ergative-absolutive case marking system in which subjects are marked with the ergative case and objects with the absolutive case in transitive sentences.[1] In addition, transitive verbs agree with both the subject and the object

**1** Since the inflection of the ergative and genitive cases is the same, many Yup’ik studies (Miyaoka 1986, 2012: 94; Jacobson and Jacobson 1995) treat the two cases together as relative cases. In this in number and person, which means that agreement does not show ergativity in Yup'ik, as seen in (1) and (2). This chapter focused particularly on sentences with indicative mood and third-person arguments.[2]

(1) Atsaq-ø tepsarq-uq.
berry-ABS.SG stink-IND.3SG
'The berry stinks.'

(2) Angte-m atsaq-ø ner-'aqa.
man-ERG.SG berry-ABS.SG eat-IND.3SG/3SG
'The man is eating the berry.'

The case markers and agreement are occasionally ambiguous in Yup'ik. First, as shown in Table 1, *-k* and *-t* are ambiguous in the sense that they can mark both ergative and absolutive. Therefore, the case markers cannot be used to judge grammatical relations when an argument is dual or plural. For instance, in (3), since both arguments are marked with ambiguous case markers (*-k* and *-t*), parsers cannot determine the grammatical relations from case marking information, but agreement information clarifies them.

**Table 1:** Third-person noun endings of the absolutive case and ergative case.

| | SG | DU | PL |
|---|---|---|---|
| **Absolutive Case** | ø | (e)k | (e)t |
| **Ergative Case** | (e)m | (e)k | (e)t |

(3) Arna-t angu-k tangerr-gket.
woman-ERG or ABS.PL man-ERG or ABS.DU see-IND.3PL/3DU
'The women are seeing the two men.'

On the other hand, agreement ambiguity is seen only when the two arguments share the same values of person and number, as in (4). Because both arguments in (4) are third-person singular, parsers determine the grammatical relations from case marking information, not agreement, contrary to (3).

study, I refer to the former as the ergative case and the latter as the genitive case in order to distinguish between its use as a subject of a transitive verb and its use as a possessor.

**2** The following abbreviations are used: 3 = third person. ACC = accusative. DU = dual. ERG = ergative. NOM = nominative. PL = plural. SG = singular.

(4) Arna-m angun-ø tangerr-aa.
woman-ERG.SG man-ABS.SG see-IND.3SG/3SG
'The woman is seeing the man.'

A sentence with ambiguities in both case markers and agreement results in a globally ambiguous sentence, as exemplified in (5). In this sentence, both nouns are marked with the ambiguous case marker *-t* and both arguments are third-person plurals, leading to a lack of morphological cues to judge grammatical relations.

(5) Arna-t angu-t tangerr-ait.
woman-ERG or ABS.PL man-ERG or ABS.PL see-IND.3PL/3PL
'The women are seeing the men.' or 'The men are seeing the women.'

Regarding the interaction between word order and ambiguity in Yup'ik, Miyaoka (2012: 180–181) states that ambiguous sentences tend to be parsed as SVO, SOV, and VSO, but this has not been quantitatively tested.

## 1.4 Our study

Since Yup'ik allows for a variety of word orders, morpho-syntactic factors such as case marking and agreement are necessary for determining the grammatical relations. I hypothesized that word order should also function as a cue to determine grammatical relations since the effects of word order are observed in a wide array of languages and Miyaoka's observation about its effects in Yup'ik. The hypothesis was verified using an acceptability judgment experiment. To examine the effect of word order, this study focused on four types of word order: SOV, SVO, OSV, and OVS. I categorized these word orders in terms of two factors: Argument Order and Verb Position. Argument Order includes the conditions SO and OS, and Verb Position includes the conditions Medial (VM) and Final (VF). Table 2 shows the conditions of the two factors of Argument Order and Verb Position.

**Table 2:** Correspondence between word order, argument order, and verb position.

| Argument Order | Verb Position | Word Order |
|---|---|---|
| SO | Medial | SVO |
| SO | Final | SOV |
| OS | Medial | OVS |
| OS | Final | OSV |

While previous studies have shown that word order is related to the ambiguity of other morpho-syntactic factors as mentioned in Section 1.2, this study clarified which ambiguity factors moderate the word order effect. Hence, focusing on the ambiguity of case marking and agreement, I set the conditions for Case Ambiguity as Unambiguous (CU) and Ambiguous (CA), and the conditions for Agreement Ambiguity as Unambiguous (AU) and Ambiguous (AA). Table 3 shows the combinations of the numbers of subject and object and their example sentences for the respective conditions of Case Ambiguity and Agreement Ambiguity.

**Table 3:** Summary of the conditions according to the two factors of Case Ambiguity and Agreement Ambiguity.

| Case | Agreement | SUB-OBJ | Example stimuli (SOV) |
|---|---|---|---|
| CU | AU | 3SG-3DU<br>3SG-3PL<br>3DU-3SG<br>3PL-3SG | ***Arna-m angute-t tangerr-ai.***<br>woman-ERG.SG man-ABS.PL see-IND.3SG/3PL<br>'The woman is seeing the men.' |
| CU | AA | 3SG-3SG | ***Arna-m angun-ø tangerr-aa.***<br>woman-ERG.SG man-ABS.SG see-IND.3SG/3SG<br>'The woman is seeing the man' |
| CA | AU | 3DU-3PL<br>3PL-3DU | ***Arna-t angu-k tangerr-gket.***<br>woman-ERG or ABS.PL man-ERG or ABS.DU see-IND.3PL/3DU<br>'The women are seeing the two men.' |
| CA | AA | 3DU-3DU<br>3PL-3PL | ***Arna-t angte-t tangerr-ait.***<br>woman-ERG or ABS.PL man-ERG or ABS.PL see-IND.3PL/3PL<br>'The women are seeing the men.'<br>or 'The men are seeing the women.' |

*Note.* SUB = Subject. OBJ = Object. AA = Agreement Ambiguous. AU = Agreement Unambiguous. CA = Case Ambiguous. CU = Case Unambiguous. SG = singular. DU = dual. PL = plural. ABS = absolutive. ERG = ergative. IND = indicative. 3 = third person.

Taken together, this study explored how the three morpho-syntactic cues of word order, case marking, and agreement affect the determination of grammatical relations in Yup'ik as an ergative and free word order language. Specifically, this study investigated two questions:

(6) Does word order (argument order and verb position) serve as a cue for the determination of grammatical relations?

(7) How do the effects of word order interact with the ambiguity of case marking and agreement?

# 2 Methods

## 2.1 Participants

The participants in this experiment included 26 native Yup'ik speakers. Two participants were excluded from the analysis because they did not seem to pay enough attention to the task (see details in Section 2.2). The age distribution of the 24 participants was as follows: six participants were 20–29 years of age, three participants were 30–39 years of age, four participants were 40–49 years of age, seven participants were 50–59 years of age, and four participants were 60–69 years of age. This research was approved by the Graduate School of Arts and Letters, Tohoku University. None of the participants were paid for their participation, and informed consent was obtained from all participants prior to the experiment.

All participants reported Yup'ik as the main language spoken during their childhood or/and the main language spoken in their elementary school. The primary language spoken in secondary and higher education institutions (equivalent to middle school, high school, and university) was Yup'ik for eight participants and English for 17 participants. With regard to language use at the time of this experiment, five participants mainly used Yup'ik both at home and at work, 13 at either home or work, and six did not mainly use it at home or work.

## 2.2 Materials

All stimuli, including target, filler, and practice sentences, were semantically reversible transitive sentences, including *arnaq* 'woman' and *angun* 'man' as arguments. The four factors: Argument Order (SO/OS), Verb Position (VM/VF), Case Ambiguity (CU/CA), and Agreement Ambiguity (AU/AA), were crossed in the target sentences, resulting in 16 conditions (see Appendix for the example stimuli of each condition). As shown in Table 3, Case Ambiguity and Agreement Ambiguity were manipulated by the number of subjects and objects.

The agent-patient combinations (woman–man or man-woman) were counterbalanced. Ten transitive verbs, namely *itegmig-* 'kick', *kaugtur-* 'hit', *nunur-* 'scold', *assike-* 'like', *cinge-* 'push', *nuteg-* 'shoot', *ceńirte-* 'visit', *qenrute-* 'get angry', *tanger-* 'see', *tuqute-* 'kill' were used to create reversible sentences.[3]

**3** Some people use *cing'e-* for 'push' instead of *cinge-*.

A total of 640 target sentences were distributed across 20 lists, following a Latin square design, and each list included 32 target stimuli. Fifteen filler sentences were added to each list to balance the acceptable and unacceptable sentences.

The present experiment used four types of filler sentences: sentences in which a subject and an object were opposite to the picture; sentences in which the numbers of subject and object were different from the picture; sentences with wrong verbs; and ungrammatical sentences with incorrect verb agreement. The verbs and nouns in the filler sentences were the same as those in the target sentences. Those who responded to filler sentences with high acceptability (6–8) for more than 8 out of 15 sentences were considered as not having paid enough attention to the task and were removed from the analysis.

For the presented pictures, the number of people depicted was set to three when an argument was plural. The characters depicted in the pictures did not vary by verb or condition. Figure 1 shows examples of the presented pictures.

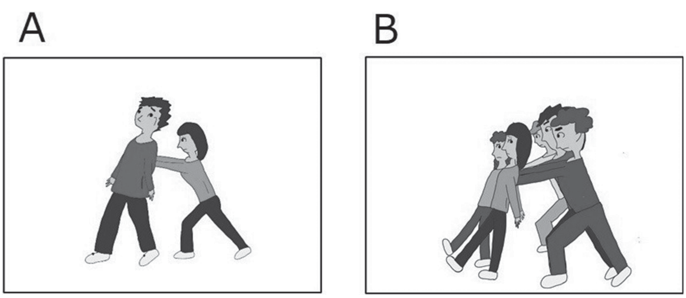



**Figure 1:** Examples of pictures (changed to grayscale from color). (A) *Arnam angun cingaa.* 'The woman is pushing the man.' (B) *Angutet arnak cingagket.* 'The men are pushing the two women'.

## 2.3 Procedure

The acceptability judgment experiment was conducted using Google Forms (https://www.google.com/forms/about/). The experiment was conducted in the following sequence: confirming the correspondence between verbs and pictures, and acceptability judgment.

In the acceptability judgment task, the participants were asked to imagine the following situations in which the given sentences were spoken: (i) the participants were looking at the given picture, (ii) the participants were describing the picture to a friend, and (iii) the picture was not visible to the friend. They were asked to judge how acceptable the given sentences were to describe the given pictures in the situation. The acceptability judgment task was scored on an eight-point Likert scale, with

1 at the low end and 8 at the high end of acceptability. The Likert scale usually refers to an odd-numbered scale. However, because the acceptability of ambiguous sentences was important in this experiment, I made it impossible to select the median value to avoid the centralizing tendency, that is, the tendency to select the middle value when one is unsure of a decision. Six sentences (two were unambiguous and canonical word ordered sentences, two were ambiguous and non-canonical word ordered sentences, and two were ungrammatical sentences) were presented as practice trials to allow the participants to arrive at a fixed standard of acceptability. The participants were not told that these sentences were practice trials.

Before the acceptability judgment task, participants were asked to confirm the correspondence between the verbs and the pictures presented in the task. These verbs were presented in the direct 3SG-3SG conjugation, which is the basic form found in the dictionary. The corresponding pictures were used with one man as the actor and one woman as the participant for all verbs.

### 2.4 Analysis

First, the raw acceptability ratings (1–8) were transformed into *z*-scores for each participant. Statistical analyses were conducted using a linear mixed-effects (LME) model (Baayen 2008) fitted with the lme4 package (Bates et al. 2015) in R studio (version 1.4.1717). The model included Argument Order (SO/OS), Verb Position (VM/VF), Case Ambiguity (CU/CA), and Agreement Ambiguity (AU/AA) as fixed factors, as well as Number of Subject (SG/DU/PL), Number of Object (SG/DU/PL), and sex of agents and patients (Woman-Man/Man-Woman) for non-interest. Participants and items were treated as random factors. I selected the final model based on Akaike's information criterion and calculated the factors of Number of Subject and Number of Object based on SG. *P*-values were calculated by submitting the final models to the lmer function of the lmerTest package (Kuznetsova, Brockhoff, and Christensen 2017).

## 3 Results

The mean acceptability of each condition is presented in Figure 2. Table 4 shows the results of the LME analysis. The main effect of Argument Order was significant ($t = -3.36$, $p < 0.001$), indicating that SO order was more acceptable than OS order. The main effect of Verb Position was not significant ($t = -0.05$, $p = 0.963$), but the interaction between Argument Order and Verb Position was marginally significant ($t = 1.75$, $p = 0.081$). In the post hoc analysis, Argument Order and Verb Position were integrated into the factor "Word Order" (SVO/SOV/OSV/OVS), then all the pairwise differences within Word Order were calculated using Tukey's method (Figure 3; Table 5). SVO and SOV were the most acceptable, followed by OSV and OVS (SVO = SOV >> OSV >> OVS).

The main effect of Case Ambiguity was not significant ($t$ = −0.89, $p$ = 0.373). However, the interaction between Argument Order and Case Ambiguity was marginally significant ($t$ = −1.90, $p$ = 0.058), and the interaction between Argument Order, Verb Position, and Case Ambiguity was significant ($t$ = 3.09, $p$ = 0.002). Because the interaction between the three factors was significant, in the post hoc analysis, the pairwise differences between Word Order in each Case Ambiguity condition were calculated using Tukey's method (Figure 4; Table 6). In the condition of CU, SVO and SOV were more acceptable than OSV and OVS (SVO = SOV >> OSV = OVS). In the condition of CA, SVO was marginally more acceptable than SOV, SOV was as acceptable as OSV, and OVS was the worst acceptable (SVO > SOV = OSV >> OVS).

Finally, Agreement Ambiguity had no effects, including the main effect and the interaction between Agreement Ambiguity and any other factors.

**Table 4:** Linear Mixed-Effect Model for the results.

| | Estimate | SE | *df* | *t* | *p* | |
|---|---|---|---|---|---|---|
| (Intercept) | 0.62 | 0.69 | 766.96 | 0.89 | 0.372 | |
| Arg. Order | −0.48 | 0.14 | 761.41 | −3.36 | < 0.001 | *** |
| Verb Position | −0.01 | 0.14 | 763.53 | −0.05 | 0.963 | |
| Agreement A. | 0.20 | 0.70 | 764.95 | 0.29 | 0.772 | |
| Case A. | −0.65 | 0.72 | 765.43 | −0.89 | 0.373 | |
| Number of Subject -DU | 0.12 | 0.70 | 765.52 | 0.17 | 0.864 | |
| Number of Subject -PL | 0.33 | 0.70 | 765.18 | 0.47 | 0.637 | |
| Number of Object -DU | 0.18 | 0.70 | 765.70 | 0.25 | 0.800 | |
| Number of Object -PL | 0.18 | 0.70 | 765.70 | 0.25 | 0.800 | |
| Sex of Agents and Patients | 0.17 | 0.70 | 765.67 | 0.24 | 0.810 | |
| Arg. Order × Verb Position | 0.09 | 0.05 | 762.90 | 1.75 | 0.081 | † |
| Arg. Order × Agreement A. | 0.07 | 0.20 | 761.95 | 0.35 | 0.724 | |
| Verb Position × Agreement A. | 0.13 | 0.20 | 762.68 | 0.66 | 0.510 | |
| Arg. Order × Case A. | −0.38 | 0.20 | 761.90 | −1.90 | 0.058 | † |
| Verb Position × Case A. | −0.22 | 0.20 | 766.05 | −1.11 | 0.267 | |
| Agreement A. × Case A. | 0.16 | 0.71 | 764.86 | 0.22 | 0.827 | |
| Arg. Order × Verb Position × Agreement A. | 0.05 | 0.28 | 763.79 | 0.19 | 0.849 | |
| Arg. Order × Verb Position × Case A. | 0.88 | 0.28 | 764.82 | 3.09 | 0.002 | ** |
| Arg. Order × Agreement A. × Case A. | −0.24 | 0.28 | 762.79 | −0.84 | 0.404 | |
| Verb Position × Agreement A. × Case A. | −0.04 | 0.28 | 766.81 | −0.14 | 0.886 | |
| Arg. Order × Verb Position × Agreement A. × Case A. | −0.47 | 0.40 | 766.64 | −1.17 | 0.244 | |

*Note.* †$p$ < 0.1. *$p$ < 0.05. **$p$ < 0.01. ***$p$ < 0.001

A. = Ambiguity. Arg. = Argument. SE = Standard Error. *df* = degrees of freedom.

The effects of Number of Subject and Number of Object were calculated based on SG.

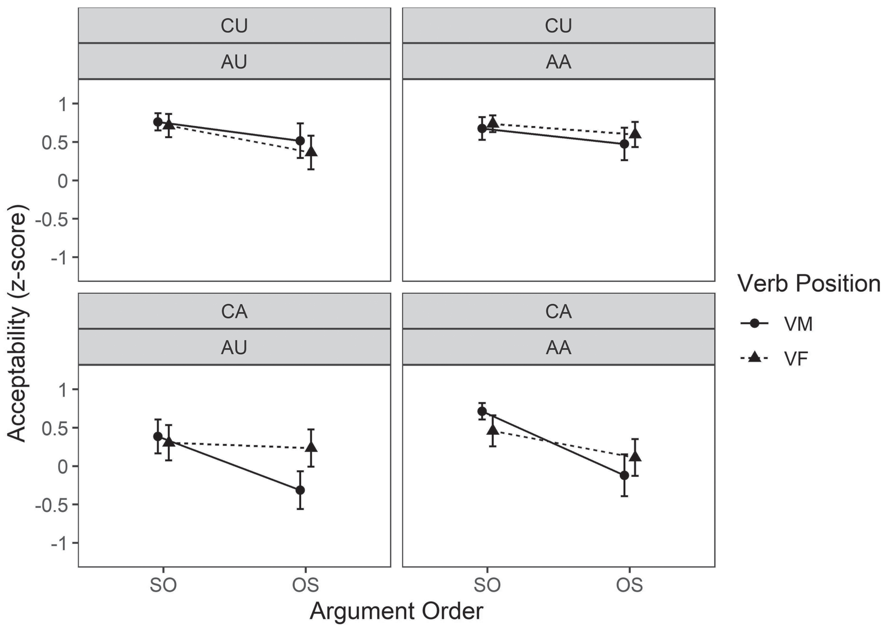



**Figure 2:** Mean acceptability (*z*-score) of each condition.
*N* = 24. Error bars represent 95% intervals.

**Table 5:** Multiple comparisons within word order.

| | Estimate | SE | df | *t* | *p* | |
|---|---|---|---|---|---|---|
| SOV-SVO | −0.11 | 0.07 | 762 | −1.60 | 0.380 | |
| SOV-OSV | −0.25 | 0.07 | 761 | 3.49 | 0.003 | ** |
| SOV-OVS | −0.56 | 0.07 | 763 | 7.81 | < 0.001 | *** |
| SVO-OSV | −0.36 | 0.07 | 763 | 5.07 | < 0.001 | *** |
| SVO-OVS | −0.67 | 0.07 | 760 | 9.38 | < 0.001 | *** |
| OSV-OVS | 0.31 | 0.07 | 762 | 4.29 | < 0.001 | *** |

*Note*. †$p < 0.1$. *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$
SE = Standard Error. df = degrees of freedom.

# 4 Discussion

The experiment showed two important findings: (i) preference for word order regardless of ambiguity and (ii) the interaction between word order preference and case marking ambiguity.

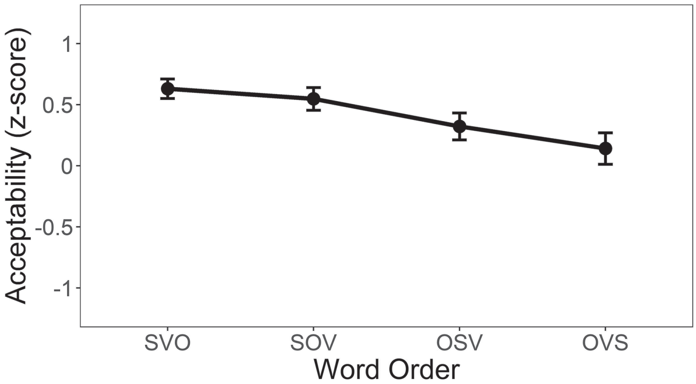



**Figure 3:** Mean acceptability (*z*-score) of each word order.
*N* = 24. Error bars represent 95% intervals.

**Table 6:** Multiple comparison within word order in each condition of Case Ambiguity.

| **Case Ambiguity = CU** | | | | | | |
|---|---|---|---|---|---|---|
| | **Estimate** | **SE** | **df** | ***t*** | ***p*** | |
| SOV-SVO | 0.01 | 0.10 | 772 | 0.08 | 1.000 | |
| SOV-OSV | −0.32 | 0.10 | 774 | 3.22 | 0.007 | ** |
| SOV-OVS | −0.43 | 0.10 | 771 | 4.23 | < 0.001 | *** |
| SVO-OSV | −0.31 | 0.10 | 766 | 3.13 | 0.010 | ** |
| SVO-OVS | −0.42 | 0.10 | 771 | 4.14 | < 0.001 | *** |
| OSV-OVS | 0.11 | 0.10 | 767 | 1.04 | 0.727 | |
| **Case Ambiguity = CA** | | | | | | |
| | **Estimate** | **SE** | **df** | ***t*** | ***p*** | |
| SOV-SVO | −0.23 | 0.10 | 771 | −2.32 | 0.096 | † |
| SOV-OSV | −0.18 | 0.10 | 775 | 1.72 | 0.314 | |
| SOV-OVS | −0.69 | 0.10 | 763 | 6.81 | < 0.001 | *** |
| SVO-OSV | −0.41 | 0.10 | 773 | 4.01 | < 0.001 | *** |
| SVO-OVS | −0.92 | 0.10 | 771 | 9.13 | < 0.001 | *** |
| OSV-OVS | 0.51 | 0.10 | 778 | 5.01 | < 0.001 | *** |

*Note*. †$p < 0.1$. *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$

SE = Standard Error. df = degrees of freedom.

## 4.1 The effects of word order

The present offline experiment showed that argument order strongly affected the acceptability of semantically reversible transitive sentences in Yup'ik, which shows morphological ergativity with free word order (Figure 3). In particular, SO order

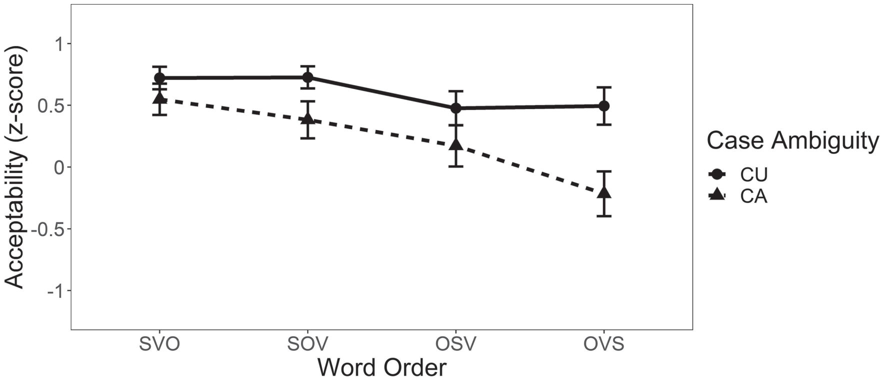



**Figure 4:** Mean acceptability (*z*-score) of each word order.
*N* = 24. Error bars represent 95% intervals.

was more acceptable than OS order regardless of the ambiguity of case marking and agreement. As any word order is grammatical in Yup'ik, this difference derives from the magnitude of the processing load (Bader and Häussler 2010; Fanselow and Frisch 2006). This result is consistent with previous studies in other ergative languages in that canonical word order requires less processing load in reversible sentences (Erdocia et al. 2009 for Basque; Kiyama et al. 2013 and Koizumi et al. 2014 for Kaqchikel Maya). While these previous studies used online methods to measure the processing load, this study used the offline method. This study showed that the SO preference is visible not only in the online data (e.g., reaction time and ERP) but also in the acceptability rate.

The results also showed the interaction between Argument Order and Verb Position. This interaction depends on the effect of Case Ambiguity, thus I discuss it below.

## 4.2 Interaction between word order and case marking

The effect of word order preference partly depended on the effect of Case Ambiguity (Figure 4). In the case-unambiguous condition, SVO and SOV were more acceptable than OSV and OVS, meaning that only SO preference was visible. On the other hand, in the case-ambiguous condition, the effect of Verb Position appeared. Specifically, SVO was the most acceptable, SOV and OSV were in the middle, and OVS was the worst acceptable. It suggests that the sequence in which objects are immediately followed by verbs is less preferred in the case-ambiguous sentences. This is compatible with the role conflict model, under which the patient-action (OV) sequence is avoided in reversible sentences. That is, SVO is SO ordered and does not include the OV order, which means it should be the most acceptable. SOV is SO ordered but includes the OV order, which means it should be a little less acceptable than SVO. OSV is OS ordered but does not include the OV order, which means it should be a little less acceptable than SVO. Finally, OVS is OS ordered and also includes the OV order, which means it should be the worst acceptable. To summarize the interaction between word order and case marking in Yup'ik, there was a tendency to prefer the SO order and avoid the OV order, but the latter was visible only in the case-ambiguous condition.

However, why does the OV avoidance strategy appear only in the case-ambiguous condition? In the gesture production experiment for English speakers, SVO was produced more than SOV only in the reversible condition (Hall et al. 2013). They also mentioned that some participants used gestures that functioned like case markers, but "the case marking gestures were especially uncommon in SVO sequences, just as case marking is rare in SVO languages" (p.13). Similarly, it could be that the Yup'ik participants in the current study used the OV avoidance strategy when the case marker cue was unavailable due to case ambiguity, but did not have to use the strategy when the case marker cue was available (case marking was unambiguous).

Finally, it should be noted that the current study used only sentences in which the subjects were always actors and the objects were always patients. The role conflict model is directly related to the thematic role, not grammatical relations. Therefore, if Yup'ik speakers truly use the role conflict strategy in case-ambiguous sentences, the results would be different when subjects and objects have other thematic roles. This possibility can be the subject of future studies.

## 4.3 Interaction between word order, case marking, and agreement

Interestingly, there were no effects related to verb agreement, although there were some effects of word order and case marking as mentioned above. This suggests that Yup'ik speakers depend on both word order (SO preference and OV avoidance) and case marking to judge grammatical relations, but the agreement cue does not play a predominant role in the sentence comprehension of Yup'ik.[4]

**4** The results are possibly due to the paradigm of the experiment in which both arguments were spelled out. Future research needs to experiment with sentences that omit arguments so that the effects of agreement could be further clarified.

Here, I applied the results in Yup'ik to the competition model and the diagnosis model. The competition model predicts that the judgment of grammatical relations should depend on case marking and agreement rather than word order in Yup'ik because case marking and agreement can always be fixed, leading to high validity, but the word order is free, meaning low validity. Contrary to the model, there was a strong word order effect, but an agreement effect was not observed in the experiment. Therefore, I argue that the strength of morpho-syntactic cues in determining grammatical relations does not necessarily connect with whether they are fixed or flexible.

On the other hand, from the perspective of the asymmetry of case marking and verb agreement, the results support the diagnosis model, under which the case marking cue is used to analyze sentences while the agreement cue is used to invalidate the incorrect analysis. The current study showed how morpho-syntactic cues were used when a sentence was analyzed, which means that the case marking, but not the agreement cue, is predicted to use under the model. Here, I applied the current results to the diagnosis model. While determining the grammatical relations, Yup'ik speakers used word order and case marking cues, but not agreement cues because the sentences in the experiment did not need to be nullified. The original diagnosis model only discusses the asymmetry of case marking and agreement, but the current study substantiated the interaction between case marking and word order. The interaction between agreement and word order when the analysis should be invalidated would be an interesting study in the future.

## 4.4 Comparison with Japanese

Finally, I compared the results in Yup'ik with the study in Japanese. First, Japanese showed results consistent with this study in terms of the effects of argument order. Although Japanese has an accusative case marking system in general, potential sentences show ergativity, in which a nominative case marker *ga* is attached to an object, as in (8).

| | | | |
|---|---|---|---|
| (8) | *Hanako-ni* | *eigo-ga* | *hanaseru-darooka.* |
| | Hanako-ERG | English-NOM | can speak-wonder |

'I wonder if Hanako can speak English.'
(Tamaoka et al. 2005)

Tamaoka et al. (2005) showed that the SO order was processed faster than the OS order in this type of sentence, as well as other sentence types with nominative alignment. According to the results for both Yup'ik and Japanese, subject-first order is preferred even when a subject takes an ergative case.

On the other hand, unlike Yup'ik, Japanese does not show an interaction between case marking and word order (Chujo 1983). Chujo made case-ambiguous sentences by omitting case markers, such as (9).

(9) a. **Unambiguous**
*Osamu-ga nimotsu-o oi-ta.*
Osamu-NOM luggage-ACC put down-PAST

b. **Ambiguous**
*Osamu nimotsu oi-ta.*
Osamu luggage put down-PAST
'Osamu put down the luggage.'
(Chujo 1983)

In Japanese, even though word order had the main effect, that is, canonical word order was processed faster than non-canonical orders, the effect of word order was not influenced by the ambiguity of case marking. There are two possible reasons for this asymmetry between the Japanese and Yup'ik. First, Chujo's study used irreversible sentences (i.e., subjects were animate and objects were inanimate) to disambiguate globally ambiguous sentences, whereas this study used reversible sentences (i.e., both subjects and objects were animate) because the pictures were presented to disambiguate globally ambiguous sentences. Thus, the asymmetry between the Yup'ik and Japanese results might be due to whether the animacy cue is ambiguous, not to a difference between the languages. The second possible reason is that Yup'ik allows the OVS word order but Japanese does not. According to Figure 4, in Yup'ik, the acceptability of the OVS order was greatly affected by case marking ambiguity, but the OSV order was not. Therefore, the effect of case marking on word order may be invisible in Japanese, which allows only the OSV, not the OVS order.[5]

5 Another possible but weaker reason is that case markers in Japanese can be omitted while they cannot be omitted in Yup'ik. Thus, Japanese speakers face a lack of case marking information more frequently than Yup'ik speakers, and the frequency may affect the interaction between case marking and word order.

# 5 Conclusion

This study investigated the effects of word order, case marking, and agreement on sentence comprehension of transitive and reversible sentences in Yup'ik, an ergative and free word order language, using an acceptability judgment experiment. Regardless of the ambiguities of case marking and agreement, canonical word orders (SOV and SVO) were more acceptable than non-canonical ones (OSV and OVS), which is consistent with their frequency. Furthermore, when case marking was ambiguous, the OV order was less preferred, probably because patient-action order leads to role conflict. Finally, unlike case marking and word order, there were no agreement effects. This suggests that Yup'ik speakers use the case marking cue and the word order cue, but not the agreement cue to determine grammatical relations.

# Appendix

Example stimuli of each condition.

SO Order (SO/OS), Verb Position (VM/VF), Case Ambiguity (CU/CA), Agreement Ambiguity (AU/AA)

**SO, VM, CU, AU**

| | | |
|---|---|---|
| Arna-m | cing-ak | angute-k. |
| woman-ERG.SG | push-3SG/3DU | man-ABS.DU |

'The woman is pushing the two men.'

**SO, VM, CU, AA**

| | | |
|---|---|---|
| *Arna-m* | *cing-aa* | *angun-ø.* |
| woman- ERG.SG | push-3SG/3SG | man-ABS.SG |

'The woman is pushing the man.'

**SO, VM, CA, AU**

| | | |
|---|---|---|
| *Arna-k* | *cinga-kek* | *angute-t.* |
| woman-ERG or ABS.DU | push-3DU/3PL | man-ERG or ABS.PL |

'The two women are pushing the men.'

**SO, VM, CA, AA**

| | | |
|---|---|---|
| *Arna-t* | *cing-ait* | *angute-t.* |
| woman-ERG or ABS.PL | push-3PL/3PL | man-ERG or ABS.PL |

'The women are pushing the men.' (or 'The men are pushing the women.')

**SO, VF, CU, AU**

| | | |
|---|---|---|
| *Arna-m* | *angute-k* | *cing-ak.* |
| woman-ERG.SG | man-ABS.DU | push-3SG/3DU |

'The woman is pushing the two men.'

**SO, VF, CU, AA**

| | | |
|---|---|---|
| *Arna-m* | *angun-ø* | *cing-aa.* |
| woman-ERG.SG | man-ABS.SG | push-3SG/3SG |

'The woman is pushing the man.'

**SO, VF, CA, AU**

| | | |
|---|---|---|
| *Arna-k* | *angute-t* | *cinga-kek.* |
| woman-ERG or ABS.DU | man-ERG or ABS. PL | push-3DU/3PL |

'The two women are pushing the men.'

**SO, VF, CA, AA**

| | | |
|---|---|---|
| *Arna-t* | *angute-t* | *cing-ait.* |
| woman-ERG or ABS.PL | man-ERG or ABS.PL | push-3PL/3PL |

'The women are pushing the men.' (or 'The men are pushing the women.')

**OS, VM, CU, AU**

| | | |
|---|---|---|
| *Angute-k* | *cing-ak* | *arna-m.* |
| man-ABS.DU | push-3SG/3DU | woman-ERG.SG |

'The woman is pushing the two men.'

**OS, VM, CU, AA**

| | | |
|---|---|---|
| *Angun-ø* | *cing-aa* | *arna-m.* |
| man-ABS.SG | push-3SG/3SG | woman-ERG.SG |

'The woman is pushing the man.'

**OS, VM, CA, AU**

| | | |
|---|---|---|
| *Angute-t* | *cinga-kek* | *arna-k.* |
| man-ERG or ABS.PL | push-3DU/PL | woman-ERG or ABS. DU |

'The two women are pushing the men.'

**OS, VM, CA, AA**

| | | |
|---|---|---|
| *Angute-t* | *cing-ait* | *arna-t.* |
| man-ERG or ABS.PL | push-3PL/3PL | woman-ERG or ABS.PL |

'The women are pushing the men.' (or 'The men are pushing the women.')

**OS, VF, CU, AU**

| | | |
|---|---|---|
| *Angute-k* | *arna-m* | *cing-ak.* |
| man-ABS.DU | woman-ERG.SG | push-3SG/DU |

'The woman is pushing the two men.'

**OS, VF, CU, AA**

| | | |
|---|---|---|
| *Angun-ø* | *arna-m* | *cing-aa.* |
| man-ABS.SG | woman-ERG.SG | push-3SG/3SG |

'The woman is pushing the man.'

**OS, VF, CA, AU**

| | | |
|---|---|---|
| *Angute-t* | *arna-k* | *cinga-kek.* |
| man-ERG or ABS.PL | woman-ERG or ABS.DU | push-3DU/3PL |

'The two women are pushing the men.'

**OS, VF, CA, AA**

| | | |
|---|---|---|
| *Angute-t* | *arna-t* | *cing-ait.* |
| man-ERG or ABS.PL | woman-ERG or ABS.PL | push-3PL/3PL |

'The women are pushing the men.' (or 'The men are pushing the women.')

## References


Baayen, R. Harald. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge/New York: Cambridge University Press.

Bader, Markus & Jana Häussler. 2010. Toward a model of grammaticality judgments. *Journal of Linguistics* 46(2). 273–330.

Bahlmann, Jörg, Antoni Rodriguez-Fornells, Michael Rotte, & Thomas F. Münte. 2007. An fMRI study of canonical and noncanonical word order in German. *Human Brain Mapping* 28(10). 940–949.

Bates, Douglas, Martin Mächler, Ben Bolker, & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48.

Bates, Elizabeth & Brian MacWhinney. 1982. Functionalist approaches to grammar. In Eric Wanner & Lila R. Gleitman (eds.), *Language Acquisition: The State of the Art*, 173–218. Cambridge/New York: Cambridge University Press.

Bates, Elizabeth & Brian MacWhinney. 1989. Functionalism and the competition model. In Brian MacWhinney & Elizabeth Bates (eds.), *The Crosslinguistic Study of Sentence Processing*, 3–73. Cambridge/New York: Cambridge University Press.

Chujo, Kazumitsu. 1983. Nihongo tanbun-no rikai katei -Bunrikai sutoratejii no sougo kankei [The interrelationships among strategies for sentence comprehension]. *Japanese Journal of Psychology* 54(4). 250–256.

England, Nora C. 1991. Changes in basic word order in Mayan languages. *International Journal of American Linguistics* 57(4). 446–486.

Erdocia, Kepa, Itziar Laka, Anna Mestres-Missé, & Antoni Rodriguez-Fornells. 2009. Syntactic complexity and ambiguity resolution in a free word order language: Behavioral and electrophysiological evidences from Basque. *Brain and Language* 109(1). 1–17.

Fanselow, Gisbert & Frisch, Stefan. 2006. Effects of processing difficulty on judgments of acceptability. In Gisbert Fenselow, Caroline Féry, Ralf Vogel, & Matthias Schlesewsky (eds.), *Gradience in Grammar: Generative Perspectives*, 291–316. New York: Oxford University Press.

Fodor, J. Dean & Atsu Inoue. 1994. The diagnosis and cure of garden paths. *Journal of Psycholinguistic Research* 23(5). 407–434.

Fodor, J. Dean & Atsu Inoue. 2000. Syntactic features in reanalysis: Positive and negative symptoms. *Journal of Psycholinguistic Research* 29(1). 25–36.

Fortescue, Michael. 1993. Eskimo word order variation and its contact-induced perturbation. *Journal of Linguistics* 29(2). 267–289.

Goldin-Meadow, Susan, Wing Chee So, Asli Özyürek, & Carolyn Mylander. 2008. The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences of the United States of America* 105(27). 9163–9168.

Hall, L. Matthew, Rachel I. Mayberry, & Victor S. Ferreira. 2013. Cognitive constraints on constituent order: Evidence from elicited pantomime. *Cognition* 129(1). 1–17.

Jacobson, Steven A. & Anna W. Jacobson. 1995. *A Practical Grammar of the Central Alaskan Yup'ik Eskimo language*. Fairbanks: Alaska Native Language Center and Program, University of Alaska.

Kaiser, Elsi & John C. Trueswell. 2004. The role of discourse context in the processing of a flexible word-order language. *Cognition* 94(2). 113–147.

Kiyama, Sachiko, Katsuo Tamaoka, Jungho Kim, & Masatoshi Koizumi. 2013. Effect of animacy on word order processing in Kaqchikel Maya. *Open Journal of Modern Linguistics* 3(3). 203–207.

Koizumi, Masatoshi. 2023. *Constituent Order in Language and Thought: A Case Study in Field-Based Psycholinguistics*. Cambridge/New York: Cambridge University Press.

Koizumi, Masatoshi, Yoshiho Yasugi, Katsuo Tamaoka, Sachiko Kiyama, Jungho Kim, Juan Esteban Ajsivinac Sian, & Lolmary Pedro Odcar Garcia Mátzar. 2014. On the (non) universality of the preference for subject-object word order in sentence comprehension: A sentence-processing study in Kaqchikel Maya. *Language* 90(3). 722–736.

Kuznetsova, Alexandra, Per B. Brockhoff, & Rune H. B. Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13). 1–26.

Laka, Itziar & Kepa Erdocia. 2012. Linearization preference given 'Free Word Order'; subject preferences given ergativity: A look at Basque. In Esther Torrego (ed.), *Of Grammar, Words, and Verses*, 115–142. Amsterdam/Philadelphia: Benjamins.

Lewis, M. Paul, Gary F. Simons, & Charles D. Fennig (eds.). 2016. *Ethnologue: Languages of the World*, 19th edn. Dallas/Texas: SIL International.

MacWhinney, Brian, Elizabeth Bates, & Reinhold Kliegl. 1984. Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior* 23(2). 127–150.

Mather, Elsie, Marie Meade, & Osahito Miyaoka. 2002. *Survey of Yup'ik grammar revised*. Osaka: Endangered Languages of the Pacific Rim, Faculty of Informatics, Osaka Gakuin University.

Matzke, Mike, Heinke Mai, Wido Nager, Jascha Rüsseler, & Thomas Münte. 2002. The costs of freedom: An ERP –study of non-canonical sentences. *Clinical Neurophysiology* 113(6). 844–852.

Mazuka, Reiko, Kenji Itoh, & Tadahisa Kondo. 2002. Costs of scrambling in Japanese sentence processing. In Nakayama Mineharu (ed.), *Sentence Processing in East Asian Languages*, 167–188. Stanford, California: Stanford University Center for the Study.

Meng, Michael & Markus Bader. 2000. Mode of disambiguation and garden-path strength: An investigation of subject-object ambiguities in German. *Language and Speech* 43(1). 43–74.

Miyaoka, Osahito. 1986. Eskimogo no nokakusei [Ergativity in Alaskan Yupik Eskimo]. *Gengo kenkyu (Journal of the Linguistic Society of Japan)* 90. 97–118.

Miyaoka, Osahito. 2012. *A grammar of Central Alaskan Yupik (CAY)*. Berlin: De Gruyter Mouton.
Sekerina, Irina A. 1997. *The syntax and processing of Russian scrambled constructions*. New York: City University of New York dissertation.
Tamaoka, Katsuo, Hiromu Sakai, Jun-ichiro Kawahara, Yayoi Miyaoka, Hyunjung Lim, & Masatoshi Koizumi. 2005. Priority information used for the processing of Japanese sentences: Thematic roles, case particles or grammatical functions? *Journal of Psycholinguistic Research* 34(3). 281–332.

Mari Kugemoto and Shota Momma

# Chapter 8 Producing long-distance dependencies in English and Japanese

## 1 Introduction

In sentence production, it is widely assumed that speakers can start speaking sentences without extensive *look-ahead*; later-coming words and structures in a sentence are not necessarily planned before its articulation onset (e.g., Griffin 2001; Levelt 1989; De Smedt 1990; Allum and Wheeldon 2007, 2009; Schriefers, Teruel, and Meinshausen 1998, Brown-schmidt et al. 2006; Brown-Schmidt and Konopka 2008; among others). For instance, Griffin (2001) suggested that when uttering sentences like *The A and the B are above the C*, speakers began to speak “The A. . .” before planning “B” and “C”. However, previous studies mostly examined relatively simple sentences where sentence-initial constituents do not depend on later coming words, and those studies tend to focus on whether the later-coming *words* are planned before the initiation of an utterance and thus little is known about how the later-coming *structures* are planned before the initiation of an utterance (cf. Wheeldon et al. 2013). Consequently, how speakers plan structural representations of complex sentences is largely unknown. For example, it remains unclear how speakers plan sentences that contain *filler-gap dependencies,* as in *what do you think the dog ate?* In this sentence, the sentence-initial constituent *(who)* is the “*filler*” that fills the “*gap*,” the missing object position, after the verb *ate.* Filler-gap dependencies are intensively studied in analytical linguistics (Chomsky 1957, 1965, 1995; Frank 2004; Kroch and Joshi 1985; Pollard and Sag 1994; Ross 1967; among many other) and in sentence comprehension research (Aoshima, Phillips, and Weinberg 2004; Frazier and Clifton 1989; Fodor 1978; Frazier, Clifton, and Randall 1983; Frazier and d’Arcais 1989; Garnsey, Tanenhaus, and Chapman 1989; Omaki et al. 2015; Wanner and Maratsos 1978; among many other). In comparison, limited attention has been paid to filler-gap dependency production. Studying filler-gap dependency production is important in constructing theories of production that are not limited in scope and connecting sentence production research with analytical linguistics. It is also likely to be useful in understanding the relationship between production and working memory mechanisms. Against this background, the current chapter aims to study the nature of the production mechanisms involved in planning sentences involving filler-gap dependencies, specifically focusing on the production of wh-dependencies, a type of filler-gap dependencies found in wh-questions.


https://doi.org/10.1515/9783110778946-008

## 1.1 Two strategies for producing wh-dependencies

The current study investigates the time-course of wh-dependency formation in English and Japanese sentence production. We compare two possible hypotheses about how speakers form wh-dependencies in speaking: the *late commitment hypothesis* and the *early commitment hypothesis* (Momma 2021). The late commitment hypothesis claims that the grammatical status of the gap is not specified when speaking the filler. For example, in sentences like *What do you think the dog ate?*, the grammatical status of *what* is not specified when *what* is uttered; in the extreme case, it may not be determined up until the materials immediately preceding the gap need to be uttered. The late commitment hypothesis allows sentence production to proceed flexibly because speakers can keep various options open throughout their production. This flexibility may be beneficial because speakers can avoid having to say a word that they are not ready to say (Ferreira 1996). For instance, when an agent noun is difficult to retrieve, speakers may want to use the passive voice to postpone it (e.g., when speakers have difficulty retrieving the word *professor*, speakers may want to say *Who was introduced by the professor?* instead of saying *Who was the professor introducing?*). If speakers commit to the object status of the filler before saying *what*, this strategy would be unavailable. At the same time, the late commitment strategy may be disadvantageous because speakers could "talk themselves into the corner." For example, when the gap happens to correspond to the participant of the event described by a relative clause, wh-dependencies would fail to be established due to the constraint that the gap cannot be posited inside a relative clause (i.e., due to the relative clause island; Ross 1967 among others). If speakers do not decide the structural position of the gap until late in the utterance, they might start speaking the filler and later realize that the filler cannot be associated with the appropriate grammatical position due to various constraints on filler-gap dependencies (Ross 1967).

In contrast to the late commitment hypothesis, the early commitment hypothesis claims that the grammatical status of the filler is already determined before the filler is uttered. For example, in *What do you think the dog ate?*, before *what* is spoken, speakers already represent *what* as the object of the verb *ate*. This strategy is beneficial because speakers can avoid positing an illicit gap. But one disadvantage is that speakers lose flexibility in their production. For example, the passivization strategy discussed above would not be available if speakers have already decided the grammatical position of the filler before starting to speak it. Considering both the late and early commitment hypotheses have advantages and disadvantages, either of those hypotheses is plausible. The present study aims to test those hypotheses in both English (Experiment 1) and Japanese (Experiment 2).

Of course, speakers of different languages may use different strategies for filler-gap dependency production, depending on the properties of the languages they speak. For instance, English and Japanese differ in the usual position of wh-phrases. In English, wh-phrases are moved to the left edge of a clause in most cases, while in Japanese they often stay in their canonical positions. When wh-phrases are moved in Japanese, the movement may be driven by a different cause than in English. Because wh-phrases in English and Japanese show different distributional properties and their movement may be driven by distinct causes, English and Japanese speakers may plan filler-gap dependencies differently. For instance, in English, speakers may develop the strategy to plan the grammatical status of the filler early to avoid violating the various constraints on long-distance dependencies, according to the early commitment hypothesis. In contrast, Japanese speakers may not adopt the early commitment strategy because wh-phrases often do not appear at the sentence-initial position, and perhaps because the constraints on long-distance dependencies may be generally more relaxed in Japanese (Kuno 1973; Omaki et al. 2020). Thus, English and Japanese speakers may reasonably differ in how they establish filler-gap dependencies in production. But they may also use fundamentally similar mechanisms for filler-gap dependency production. The present study thus aims to compare the time-course of wh-dependency formation in English and Japanese, to explore how typological differences may affect wh-dependency planning mechanisms.

## 1.2 A method for investigating the time-course of wh-dependency production

To investigate wh-dependency planning processes in English and Japanese, we used a close relative of the method that Momma (2021) used to investigate the time-course of filler-gap dependency planning. Before we elaborate on the current method, the basic logic of the method used in Momma (2021) should be explained. The method relied on two previously well-established phenomena: the structural priming effect (Bock 1986; see Pickering and Ferreira 2008 and Mahowald et al. 2016 for a recent overview) and the *that*-trace constraint (Perlmutter 1971; see Pesetsky 2017 for a recent overview).

The structural priming effect refers to a phenomenon that speakers tend to re-use the same structures they recently encountered (Bock 1986). For instance, after encountering a prepositional dative sentence like *I showed my drawing to her*, speakers are more likely to produce the prepositional dative structure *The boy gave the ball to the dog* than its double object counterpart, *The boy gave the dog the ball.* Structural priming can occur without any overlap in words between prime and target sentences (Bock 1986). Usually, the structural priming effect is measured as the increase in the production rate of a particular structure, but structural priming has also been shown to speed up the production of the primed structure (Wheeldon and Smith 2003; Seagert, Wheeldon, and Hagoort 2016). Most relevantly for current purposes, the complementizer *that* can be structurally primed. Ferreira (2003) reported that sentences with the complementizer *that* increased the likelihood of speakers using *that* in the subsequent production. For example, after encountering prime sentences like *The director announced that Hollywood's hottest actor would be playing the part*, speakers were more likely to produce *that* in target sentences like *The jury believed that the young witness told the truth* than after encountering minimally different prime sentences like *The director announced Hollywood's hottest actor would be playing the part*. This complementizer priming is not reducible to the priming of the phonological form of *that*. This is because Ferreira (2003) showed that the demonstrative *that* as in *that dog* did not prime the complementizer *that*, and because the null complementizer also primed the null complementizer. Thus, the complementizer priming is best characterized as priming at the structural level, not the phonological level.

Momma (2021) also used the constraint known as the *that*-trace effect (Perlmutter 1971; see Ritchart et al. 2016 for laboratory-based experimental evidence for this effect). The *that*-trace constraint bans the structures where the complementizer *that* is followed by the gaps in the following:

(1) *Which girl do you think that ate the cake?

This effect is not observed in sentences where the gap corresponds to the embedded object position, as in the following:

(2) Which cake do you think that the girl ate?

Importantly, the *that*-priming effect and *that*-trace constraint conflict with each other. The *that*-priming effect encourages speakers to say *that* while the *that*-trace constraint prohibits them to say *that.* Momma (2021) showed that this conflict between the *that*-priming effect and the *that*-trace constraint slowed down the planning process. That is, speakers are slower to speak sentences like *Who do you think met the girl?* given prime sentences with *that* like *The boy thinks that the dog liked them* than given minimally different prime sentences without *that*, presumably due to the conflict between the *that*-priming effect and the *that*-trace constraint. Critically, in a series of picture description experiments, it was observed that this slow-down effect appeared before the sentence onset of utterances, that is, before starting to say the filler. This suggests that speakers already plan the grammatical function of the filler, as well as the complementizer structure of the gap-containing clause, in accordance with the early commitment hypothesis.

## 1.3 Current experiments

Having explained the logic used in Momma (2021), we are now ready to describe the current experiments. There are two experiments in the current study. Previous studies on sentence planning often used picture description tasks (Allum and Wheeldon 2007, 2009; Schriefers, Teruel, and Meinshausen 1998; Smith and Wheeldon 1999; Konopka and Meyer 2014; among others). However, because it is difficult to elicit complicated target sentences of interest in English and Japanese using a picture description task, the current study alternatively used a variant of sentence recall task. The working assumption is that sentence recall involves the regeneration of memorized sentences from their conceptual representations (Potter and Lombardi 1998). In both Experiments 1 and 2, participants memorized one target sentence and one prime sentence in this order and recited the target sentence. In this task, because the prime sentence is the last sentence they encounter before uttering the target sentence, the structure of the prime sentence would be primed in the target production.

In Experiment 1, we examined if English speakers plan the grammatical status of the gap before starting to speak the filler, as in Momma (2021), but using the sentence recall task. We aim to evaluate if the results from Momma (2021) can be conceptually replicated and if they can generalize to different task contexts. In Experiment 1, prime and target sentences were like the following:

(3) Prime sentences
   a. Do you think <u>that</u> the student solved the question? (*that* prime)
   b. Do you think the student solved the question? (null prime)

(4) Target sentences
   a. Which trainer do you think loved the lion? (subject extraction)
   b. Which trainer do you think the lion loved? (object extraction)

Given prime sentences with the complementizer *that* like (3a), speakers should be more inclined to say *that* in target sentences than given prime sentences like (3b). However, when the target sentence is an embedded subject wh-question like (4a), the complementizer *that* cannot be used because the *that*-trace constraint prohibits the complementizer *that* followed by the subject gap. Thus, the *that* priming and the *that*-trace constraint creates a conflict in production of sentences like (4a) given a prime sentence with the complementizer *that* like (3a).

Experiment 2 aimed to test whether Japanese speakers plan wh-dependencies before speaking the wh-phrase scrambled to the sentence-initial position. However, because, as far as we know, Japanese does not have a structure that can potentially violate the *that*-trace constraint, we used a different effect to make inferences about the timing of filler-gap dependency planning. Namely, we used simple structural priming on two types of wh-questions with two different scope relations. In Japanese, wh-phrases are associated with the question particle, -*ka*. When a sentence is bi-clausal and a wh-phrase is extracted from the embedded clause, the position of the Q-particle determines the scope of the wh-phrase.

(5) a. どの ライオンが 逃げた と 言いました か? (matrix)
Which lion-NOM ran-away that said-POLITE Q
'Which lion did you say ran away?'
b. どの ライオンが 逃げた か 言いました か? (embedded)
Which lion-NOM ran-away Q said-POLITE Q
'Did you say which lion ran away?'

In both sentences, the wh-phrase *which lion* occurs in the initial position of the sentence. But in (5a), it is associated with the sentence-final Q-particle and has the matrix scope. In contrast, when the wh-phrase is associated with the Q-particle in the embedded clause as in (5b), it is usually interpreted to have the embedded scope, although it could have the matrix scope when prosodically licensed. Based on the finding by Wheeldon and Smith (2003) and Segaert et al. (2017) that speakers are faster to speak the primed structures, we predicted that speakers should be faster to plan target sentences when prime sentences have the same scope relation as target sentences. If this potential speed-up effect is observed in the onset latency of target utterances where wh-fillers are fronted, it can be inferred that speakers plan (a) whether the wh-filler is associated with the embedded or matrix complementizer and (b) the type of complementizer used for the embedded and matrix clause, before starting to speak sentence-initial wh-fillers. If this prediction is met, it can be argued that Japanese speakers plan the structural representations of wh-dependencies early, before starting to speak the scrambled wh-filler, just like English speakers. More specifically, it can be argued that both English and Japanese speakers minimally plan the complementizer structure of the clause that the relevant wh-phrase is taking scope over, before starting to speak the sentence-initial wh-fillers.

## 2 Experiment 1

Like in Momma (2021), Experiment 1 examined the timing of wh-dependency formation in English using the conflict between the *that*-priming effect and the *that*-trace constraint in subject extracted wh-questions. The early commitment hypothesis predicts that this conflict would cause a slow-down effect at the onset of subject-extracted wh-questions, but not object-extracted wh-questions.

## 2.1 Method

### 2.1.1 Participants

Forty-eight monolingual English speakers were recruited via Prolific Academic. Informed consent was obtained from each participant. Each participant was paid five US dollars as compensation for the 20–30 minutes experiment. We replaced eleven participants who did not follow instructions or whose recordings were not intelligible and two additional participants who had less than half error-free trials.

### 2.1.2 Materials

For the target sentences, forty-eight pairs of subject-extracted wh-questions and object-extracted wh-questions like (4a) and (4b) were constructed (see Table 1). All sentences began with *Which NP do you think. . .* The prime sentences like (3a) and (3b) were forty-eight yes-no questions either with or without the complementizer *that*. They began with either *Do you. . .* or *Do they. . .* The prime sentences were paired with the target sentences so that they did not share the content words aside from the embedding verb *think*. They also did not have any obvious semantic relationship.

**Table 1:** The four conditions in Experiment 1.

| condition | target sentence | prime sentence |
|---|---|---|
| subject-extraction / *that* prime | Which trainer do you think loved the lion? | Do you think that the student solved the question? |
| subject-extraction / null prime | | Do you think the student solved the question? |
| object-extraction / *that* prime | Which trainer do you think the lion loved? | Do you think that the student solved the question? |
| object-extraction / null prime | | Do you think the student solved the question? |

### 2.1.3 Procedure

The experiment was conducted online using PCIbex (Zehr and Schwarz 2018). At the beginning of the experiment, there were three practice trials, which had the same task structure as the experimental trials. The experimental trials were structured as follows. First, a target sentence was presented for 5000 ms. Participants were instructed to read it aloud and memorize it. Subsequently, a prime sentence was presented for 5000 ms, which participants also read aloud and memorized. After a blank screen presented for 2000 ms, either '1' or '2' in the red font was presented as the prompt for recall. Participants were instructed to recite the first sentence when '1' was presented. They were instructed to recite the second sentence when '2' was presented. In critical trials, '1' was always presented, as the target sentences were always presented as the first sentence. In filler trials, which were indistinguishable from critical trials from the participants' perspectives, participants were presented with '2.' Thus, speakers could not reliably predict which sentence they needed to recall. There were 48 critical trials and 24 filler trials.

### 2.1.4 Scoring and analysis

All audio files were first transcribed and coded for errors. Errors were defined as any deviations from target sentences. Incomplete utterances, trials where participants were still uttering the previous sentence after the recall prompt, and trials where participants uttered overt hesitation (uh, am, etc.) before finishing the sentence, were also coded as erroneous. The erroneous trials were excluded from the subsequent analysis. Trials where participants said the complementizer *that* in the object extracted wh-question (e.g., *Which trainer do you think that the lion loved?*) and trials where participants replaced *you* with *they* (e.g., *Which trainer do they think the lion loved?* for the target *Which trainer do you think the lion loved?*) were included in the analysis. The onset latency of the error-free trials was manually measured using Praat, by the authors and a research assistant who were all blind to the prime type condition.

Using R (R Core Team 2020) and *lmer* package (Bates et al. 2015), a linear mixed-effects model was fit for the onset latency of target sentences. The model was initially maximal in the sense of Barr et al. (2013), but due to the convergence issue, the random slopes were removed from the model. When simplifying the model, the random slope that accounted for the least amount of variance was removed successively, until the model converged. The final model had PrimeType (that vs. null), ExtractionType (subject vs. object) and their interaction as fixed effects, and by-subject and by-item random intercepts.

## 2.2 Result

In Experiment 1, 30.8 % of the trials (606 out of 1968 trials) were excluded from the subsequent analyses as erroneous trials. The error rates in each condition are shown in Table 2. The trials where the onset latency is longer than 2500 ms (16 out of 1968 trials; 0.8%) were excluded as well.

**Table 2:** Error rates in each condition in Experiment 1.

| condition | error rate |
|---|---|
| subject-extraction / *that* prime | 28.7 % |
| subject-extraction / null prime | 32.5 % |
| object-extraction / *that* prime | 30.5 % |
| object-extraction / null prime | 28.2 % |

As shown in Figure 1, in the subject extraction condition, speakers were 46 ms slower in the *that* prime condition than in the null prime condition, but in the object extraction condition, they were 8 ms slower in the *that* prime condition than in the null prime condition. Supporting this pattern, the statistical model showed that the interaction between ExtractionType and PrimeType was significant ($\beta = 0.05$, $SE = 0.02$, $|t| = 2.13$, $p = 0.03$). In addition, the planned comparison based on the nested models showed that the simple effect of PrimeType was significant in the subject extraction condition ($\beta = 0.05$, $SE = 0.02$, $|t| = 2.73$, $p = 0.006$), but not in the object extraction condition ($p = 0.78$). The main effect of Prime Type was marginally significant ($p = 0.08$), but this is not interpretable given the interaction involving this term. The main effect of Extraction Type was not significant ($p = 0.77$).

## 2.3 Discussion

The results of Experiment 1 showed that there was a slow-down effect in onset latency selectively in the subject extraction condition, but not in the object extraction condition. This pattern replicates Momma (2021) but in a different task environment. This suggests that speakers know that, as early as at the sentence onset, the subject-extracted question is not compatible with the *that* complementizer. That is, speakers plan the structural properties of wh-dependency, specifically the grammatical function of the extracted wh-phrase and the complementizer type of the gap-containing clause before uttering it. This supports the early com-

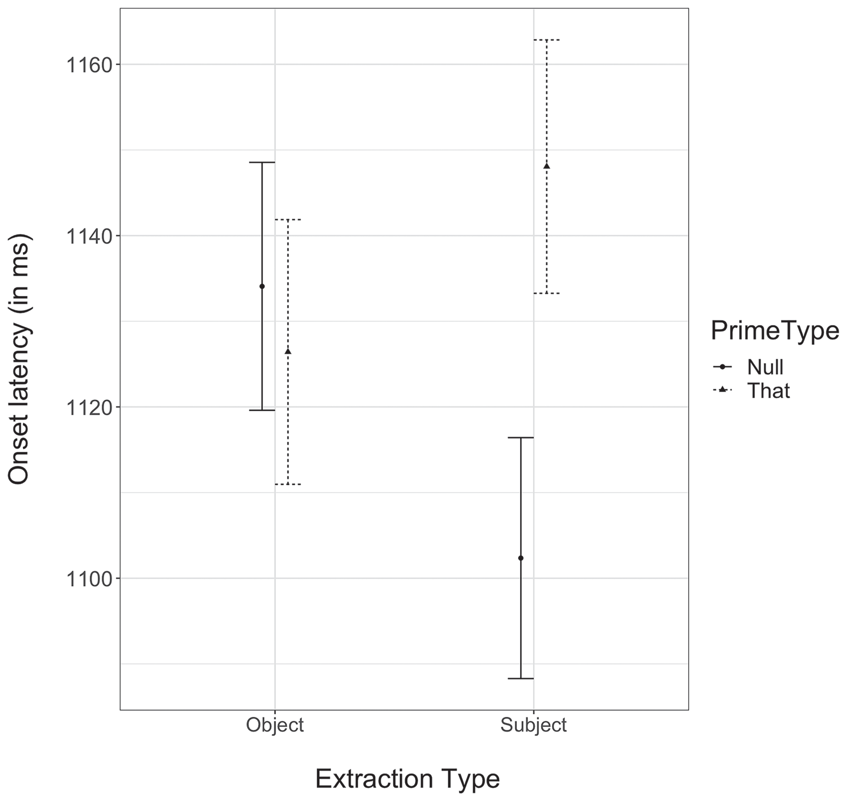



**Figure 1:** By-subject mean onset latency across four conditions in Experiment 1. Error bars represent the standard error of the means.

mitment hypothesis, which claims that speakers plan the grammatical details of wh-dependencies before uttering the sentence-initial filler.

## 3 Experiment 2

Experiment 1 showed that speakers plan the grammatical details of wh-dependency in English. However, this early commitment strategy may be language-specific. For example, because some constraints on filler-gap dependencies may be relaxed (or even absent) in Japanese (Kuno 1973; Omaki et al. 2020), Japanese speakers may have weaker motivations for planning the grammatical status of the filler/gap before the filler production. Experiment 2 investigated if Japanese speakers nevertheless use the early commitment strategy for planning wh-dependency despite relevant typological differences. As discussed in the introduction, we used the potential speed-up effect in onset latency due to structural priming. Specifically, we hypothesized that the scope of wh-phrases can be primed, and this priming effect would lead to faster onset latency when target sentences share the same wh-scope with prime sentences. If this potential speed-up effect is observed before the onset of sentence-initial wh-phrases, it can be inferred that speakers plan at least the scope relation of wh-phrases and by extension the complementizer type of the embedded clause, before starting to speak the wh-filler.

## 3.1 Method

### 3.1.1 Participants

Thirty-five native Japanese speakers participated in Experiment 2 online. For those who live outside of Japan, it was confirmed that they acquired Japanese in their infancy and use Japanese daily via a questionnaire. No demographic information was collected other than language backgrounds. Each participant was paid ten US dollars or 1000 yen per an hour as compensation for the 30–45 minute experiment. We replaced two participants who did not follow instructions or whose recordings were not intelligible and nine additional participants who had less than half error-free trials.

### 3.1.2 Materials

The stimuli were questions like (5a) and (5b). Table 3 shows the four conditions of the prime and target sentence combinations. All sentences had the same matrix verb and ending 言いましたか ('said-POLITE-Q') to make the sentences easier to memorize. In addition, to make the sentences as simple as possible, wh-phrases were always the subject and all verbs were intransitive verbs or verbs whose objects can be omitted naturally without contextual support. Because Japanese is a pro-drop language that allows pronouns to be omitted, the sentences like (5a) and (5b) are in principle ambiguous between the parse where the matrix subject is dropped and the parse where the embedded subject is dropped. However, to force participants to interpret the subject as extracted from the embedded clause, all subject noun phrases were headed by non-human nouns except for 'the baby'. This would prevent the parse where the embedded subject is dropped because the parse where non-human noun phrases function the subject of the matrix verb *say* yields implausible interpretation (e.g., *Which lion said you ran away?*).

**Table 3:** The four conditions in Experiment 2. The sentences are translated from Japanese.

| condition | target sentence | prime sentence |
|---|---|---|
| matrix scope / matching scope prime | Which lion did you say ran away? | Which train did you say stopped? |
| matrix scope / mismatching scope prime | | Did you say which train stopped? |
| embedded scope / matching scope prime | Did you say which lion ran away? | Which train did you say stopped? |
| embedded scope / mismatching scope prime | | Did you say which train stopped? |

#### 3.1.3 Procedure

The same procedure as in Experiment 1 was used.

#### 3.1.4 Scoring and analysis

All audio files were transcribed and coded for errors using the same criteria as in Experiment 1. Onset latencies were measured with the same procedure as in Experiment 1. The onset latency of target sentences was analyzed using linear mixed-effects modeling. The model was initially maximal but was simplified in the same way as in Experiment 1 due to the convergence issue. The final model had PrimeType (match vs. mismatch), Scope (matrix vs. embedded) of the target sentence, and their interaction as fixed effects, and by-subject and by-item random intercepts.

### 3.2 Result

In Experiment 2, 31 % of the trials (521 out of 1680 trials) were excluded from the subsequent analyses as erroneous trials. The error rates in each condition are shown in Table 4. Onset latencies longer than 2500 ms (0.5 %, 8 out of 1680 trials) were also excluded.

As can be seen in Figure 2, in the embedded scope condition, speakers were 61.1 ms slower in the mismatch condition than in the match condition, but in the matrix scope condition, they were 30.6 ms faster in the mismatch condition than in the match condition. Supporting this pattern, the statistical model showed that the interaction between Scope and PrimeType was significant ($\beta = -0.08$, $SE = 0.02$, $|t| = 3.16$, $p = 0.002$).

**Table 4:** Error rates in each condition in Experiment 1.

| condition | error rate |
|---|---|
| matrix scope / matching scope prime | 21.7 % |
| matrix scope / mismatching scope prime | 46.4 % |
| embedded scope / matching scope prime | 15.5 % |
| embedded scope / mismatching scope prime | 40.5 % |

In addition, the planned comparison based on the nested models showed that the simple effect of PrimeType was significant in the embedded scope condition ($\beta = 0.04$, $SE = 0.02$, $|t| = 2.6$, $p = 0.01$), and it was marginally significant in the matrix scope condition ($\beta = -0.03$, $SE = 0.02$, $|t| = 1.87$, $p = 0.06$). Neither the main effect of Prime Type ($p = 0.65$) nor the main effect of Scope ($p = 0.44$) was significant.

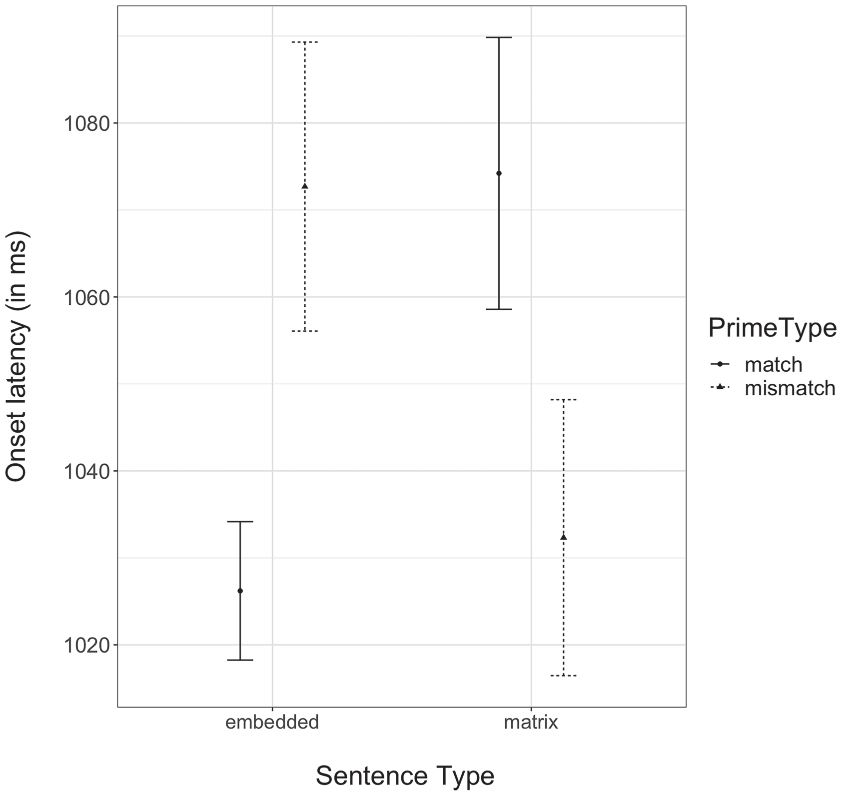



**Figure 2:** By-subject mean onset latency across four conditions in Experiment 2. Error bars represent the standard error of the means.

## 3.3 Discussion

The results showed that speakers were faster to start speaking sentences with the embedded scope given prime sentences with the embedded scope. In contrast, speakers were marginally slower to start speaking sentences with the matrix scope given prime sentences with the matrix scope. There was an interaction between Prime Type and Scope. We suggest that this interaction can be explained by assuming a facilitatory effect of scope priming (cf. Wheeldon and Smith 2003) and an inhibitory effect of similarity-based interference (Lewis 1996), which to some extent cancel each other out. First, the similarity-based interference slows production planning when the prime and the target sentences are similar in scope, perhaps because two sentences are less discriminable from each other when they have the same scope properties (in the match condition) than when they have distinct scope properties (in the mismatch condition). This effect of similarity-based interference is masked by the facilitatory effect of scope-related structural priming in the embedded scope condition. However, the similarity-based interference effect in the matrix scope condition remains observable because the scope-related structural priming effect is less strong in the matrix scope condition. The reason that the structural priming effect is less strong in the matrix scope condition may be due to the effect known as the *inverse preference* effect (Jaeger and Snider 2008; Reitter, Keller, and Moore 2011; Bernolet and Hartsuiker 2010; Ferreira 2003; among others). In the structural priming literature, it is widely observed that less frequent structures are more easily primed than more frequent structures. It is reasonable to assume that the matrix scope is less primable than the embedded scope because the matrix scope wh-questions occur even in sentences without any embedded clauses (i.e., in mono-clausal sentences). If the matrix wh-scope in bi-(or multi-) clausal sentences and mono-clausal sentences are treated as the same type of dependency configuration, the matrix wh-scope would be more frequent than the embedded scope interpretation. Given the inverse preference effect, it may be harder to prime the matrix scope structures than to prime the embedded scope structures. If this is the case, the similarity-based interference effect should mask the small structural priming effect in the matrix scope condition, but the relatively large structural priming effect should mask the similarity-based interference effect in the embedded scope condition. Thus, the combination of similarity-based interference and structural priming may explain the pattern we observed in the current data, although this explanation remains speculative and the assumptions we made here should be independently verified with further studies.

## 4 General discussion

Both Experiment 1 and 2 show that the structural properties of wh-dependency are planned before the wh-phrase is spoken, as the early commitment hypothesis predicts. In Experiment 1 in English, the slow-down effect caused by the conflict between the *that*-priming and the *that*-trace constraint was observed in onset latency, replicating Momma (2021). This suggests that speakers already plan the grammatical function of the filler and the complementizer of the gap-containing clause before starting to speak the wh-filler, across different task environments. In Experiment 2, we use the structural priming of wh-scopes in Japanese to make inferences about the time-course of wh-dependency planning. The results show a complicated pattern, but under our interpretation, they minimally suggest that speakers plan the scope of wh-phrases early and, assuming that the complementizer structures are critically involved in determining the scope relation, the complementizer of the clause that wh-phrase is taking scope over. For the target sentences with the embedded wh-scope, speakers were faster to start speaking when the prime sentences also had the embedded wh-scope. In contrast, for the target sentences with the matrix wh-scope, speakers were marginally slower when the prime sentences also had the matrix wh-scope. Although this pattern was not entirely predicted and deserves further investigation, we speculate that this pattern was caused by the interplay between the facilitatory effect of scope priming and the inhibitory effect of similarity-based interference. Taken together, Experiment 1 and Experiment 2 both suggest that speakers plan the complementizer structure of the clause containing the gap before starting to speak the filler. This in turn suggests some abstract similarity between how English and Japanese speakers plan wh-dependencies, despite surface differences in how such dependencies are realized.

We argue that the inhibitory effect found in the matrix scope condition was due to the similarity-based interference. Previous research suggested that the similarity-based interference arises in the process of retrieving words from memory during comprehension (see Van Dyke and McElree 2006 for an overview). The similarity-based interference in word retrieval also occurs in production. For instance, in Smith and Wheeldon (2004), the latency of sentences containing two semantically related nouns such as *the saw and the axe move down* is longer than when the two nouns are not related as in *the saw and the cat move down*, suggesting that the later-coming nouns interfered with the retrieval of the initial noun (at least when they are planned together). Thus, the similarity-based interference arises both in comprehension and production when a word similar to the retrieval target is co-present in memory. Given that the current study uses a memory-based task, it is conceivable that the retrieval of a sentence with the embedded or matrix scope can be more difficult in the presence of another sentence with the same scope property in memory. The relevant notion of similarity here can be about the complementizer type (question particle vs. declarative complementizer), the scope relation (embedded vs. matrix), or the sentence type (wh- vs. yes-no question). Experiment 2 does not provide evidence to determine which of those properties are relevant to the similarity-based interference effect we postulated here. Nevertheless, the slow-down effect we found in the matrix scope condition may reflect the interference based on the similarity of the properties related to wh-scope.

We also speculate that the similarity-based interference arises in both the embedded and matrix scope conditions, but it is canceled out by the facilitatory effect of scope-related structural priming in the embedded scope condition. We attribute the lack of the facilitatory priming effect in the matrix scope condition to the inverse preference effect, based on the assumption that the mono-clausal wh-scope is also counted as the matrix wh-scope. That is, the matrix scope is difficult to prime because it is frequent. Although that assumption about frequency counting needs to be tested independently, the matrix wh-dependency structure in multi-clausal sentences is the same as that in the mono-clausal questions in the sense that they both involve the dependency between the wh-phrase and the question particle in the matrix clause.

Under our interpretation, the current results suggest that both English and Japanese speakers plan the grammatical details of wh-dependencies before starting to speak the wh-filler. This way of planning sentences involving wh-dependencies is generally congruent with the broad class of production theories that allow the generation of structural representations before selecting lexical items (e.g., Garrett 1975; see Bock and Ferreira 2014). In the current studies, we provide evidence that structural representations encoding wh-dependencies are at least to some extent planned, but we cannot tell from current results if words and structures intervening the filler and the gap are planned or not planned before the speech onset. However, Momma (2021) showed that the words intervening between the filler and the gap are likely not planned before the filler production. Momma (2021) argued that the formalism known as Tree-Adjoining Grammar (Joshi, Levy, and Takahashi 1975; Frank 2004) naturally captures this idea that the planning of sentences involving filler-gap dependencies starts with first building the non-contiguous parts of sentences (the filler and the gap). Under this view, words and structures intervening between the filler and the gap are planned later. In other words, filler-gap dependency production can still be incremental, in the sense that planning and articulation are still frequently interleaved in the production of a single sentence. Although the current experiments do not provide direct evidence for or against this view, given that speakers in current experiments took only slightly more than 1 second to start speaking, we deem it implausible that speakers in the current experiments planned all words and structures intervening the filler and the gap in details before starting to speak. Thus, the current results are naturally compatible with the view that the structural representations of non-contiguous parts (the filler and the gap) are planned as a unit, and the words and structures intervening the filler and the gap are inserted later. This hypothesis about filler-gap dependency production can be subsumed under the view that speakers can build structural representations prior to lexical selection (e.g., Garrett 1975; see Bock and Ferreira 2014 for a recent overview). Given that the current results show some high-level parallelism between English and Japanese, this view might be applied to both English and Japanese sentence production.

Lastly, we acknowledge that the current study has the limitation that sentence recall tasks may differ from naturalistic language production processes in relevant respects. Sentence recall tasks are not widely used as a method to investigate the time-course of language production, and speakers in recall experiments may deploy planning procedures that are fundamentally different from those in naturalistic production. However, it is worth noting that the accessibility effect on word order, the effect that is usually assumed to arise from the temporal dynamics of sentence planning, can be observed in recall-based experiments (e.g., Bock and Irwin 1980; McDonald, Bock, and Kelly 1993; Tanaka et al. 2011), as in naturalistic production (e.g., Kempen and Harbusch 2011). Also, in our lab, several lines of study show that the time-course of verb planning is similar between recall-based experiments and picture-description experiments (e.g., Momma and Yoshida 2021). Thus, we assume that the time-course of sentence-recall mirrors the time-course of naturalistic sentence production as a reasonable starting point, although of course this assumption should be evaluated further. Finally, we also acknowledge that the results of Experiment 2 have an alternative interpretation. For example, it may be that speakers were simply slower to start speaking after reading and memorizing a matrix scope prime sentence (that is, after reading a match prime sentence in the matrix scope condition and after reading a mismatch prime sentence in the embedded scope condition), perhaps because matrix scope sentences are more complex than embedded scope sentences. This possibility cannot be ruled out in the current study, but future studies should examine the relationship between the complexity of prime sentences and the production latency of the target sentence production in the current task. If this interpretation is correct, more complex prime sentences should increase the onset latency of subsequent target production.

## 5 Conclusion

The current study shows the grammatical details of wh-dependencies are predominantly planned before the utterance of the sentence-initial wh-phrases both in English and Japanese, in accordance with the early commitment hypothesis. Of course, the current study does not show that the planning processes involved in wh-dependency formation in production are identical between English and Japanese. However, English and Japanese sentence production may plausibly involve similar planning mechanisms for formulating wh-dependencies, despite the surface differences in how wh-dependencies are realized in the two languages. Specifically, wh-dependency formation in English and Japanese may both involve planning the complementizer structure before producing the filler.

## References


Aoshima, Sachiko, Colin Phillips, & Amy Weinberg. 2004. Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language* 51. 23–54.

Allum, Paul H. & Linda R. Wheeldon. 2007. Planning scope in spoken sentence production: the role of grammatical units. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33(4). 791.

Allum, Paul H. & Linda Wheeldon. 2009. Scope of lexical access in spoken sentence production: Implications for the conceptual–syntactic interface. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 35(5). 1240.

Bates, Douglas, Martin Mächler, Ben Bolker, & Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1). 1–48.

Barr, Dale J., Roger Levy, Christoph Scheepers, & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278.

Bernolet, Sarah & Robert Hartsuiker. 2010. Does verb bias modulate syntactic priming? *Cognition* 114(3). 455–461.

Bock, Kathryn J. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18(3). 355–387.

Bock, Kathryn & Victor S. Ferreira. 2014. Syntactically speaking. In Matthew Goldrick, Victor Ferreira, & Michele Miozzo (eds.), *The Oxford Handbook of Language Production*, 21–46. Oxford: Oxford University Press.

Bock, Kathryn J. & David E. Irwin. 1980. Syntactic effects of information availability in sentence production. *Journal of Verbal Learning and Verbal Behavior* 19(4). 467–484.

Brown-Schmidt, Sarah & Agnieszka E. Konopka. 2008. Little houses and casas pequeñas: Message formulation and syntactic form in unscripted speech with speakers of English and Spanish. *Cognition* 109(2). 274–280.

Brown-Schmidt, Sarah & Michael K. Tanenhaus. 2006. Watching the eyes when talking about size: An investigation of message formulation and utterance planning. *Journal of Memory and Language* 54(4). 592–609.

Chomsky, Noam. 1957. *Syntactic Structure*. The Hague: Morton.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge: MIT Press.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge: MIT Press.

De Smedt, Koenraad. 1990. IPF: An incremental parallel formulator. In Robert Dale, Chris Mellish, & Michael Zock (eds.), *Current Research in Natural Language Generation,* 167–192. Academic Press.

Ferreira, Victor S. 1996. Is it better to give than to donate? syntactic flexibility in language production. *Journal of Memory and Language* 35(5). 724–755.
Ferreira, Victor S. 2003. The persistence of optional complementizer production: Why saying "that" is not saying "that" at all. *Journal of Memory and Language* 48. 379–398.
Fodor Janet D. 1978. Parsing strategies and constraints on transformations. *Linguistic Inquiry* 9. 427–473.
Frank, Robert. 2004. *Phrase Structure Composition and Syntactic Dependencies*. Cambridge, MA: MIT Press.
Frazier, Lyn, Charles Clifton, & Janet Randall. 1983. Filling gaps: Decision principles and structure in sentence comprehension. *Cognition* 13(2). 187–222.
Frazier, Lyn & Charles Clifton. 1989. Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes* 4(2). 93–126.
Frazier, Lyn & Giovanni B. Flores d'Arcais. 1989. Filler driven parsing: a study of gap filling in Dutch. *Journal of Memory and Language* 28. 331–344.
Garrett, Merrill F. 1975. The analysis of sentence production. In G. H. Bower (ed.), *Psychology of Learning and Motivation*, Vol. 9, 133–177. New York: Academic Press.
Garnsey Susan M., Michael K. Tanenhaus, & Robert M. Chapman. 1989. Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistic Research* 18. 51–60. 10.1007/BF01069046
Griffin, Zenzi M. 2001. Gaze durations during speech reflect word selection and phonological encoding. *Cognition* 82(1). B1–B14.
Jaeger, T. Florian & Neal E. Snider. 2008. Implicit learning and syntactic persistence: Surprisal and cumulativity. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, Vol. 827812. Austin: Cognitive Science Society.
Joshi, Aravind K., Leon S. Levy, & Masako Takahashi. 1975. Tree adjunct grammars. *Journal of Computer and System Sciences* 10(1). 136–163.
Kempen, Gerard & Karin Harbusch. 2011. Word order scrambling as a consequence of incremental sentence production. In Holden Härtl & Heike Tappe (eds.), *Mediating Between Concepts and Grammar*, 141–164. Berlin: De Gruyter Mouton.
Konopka, Agnieszka E. & Antje S. Meyer. 2014. Priming sentence planning. *Cognitive Psychology* 73. 1–40.
Kroch, Anthony S. & Aravind K. Joshi. 1985. The linguistic relevance of tree adjoining grammar. *Technical Reports MS-CIS-85-18*.
Kuno, Susumu. 1973. Constraints on internal clauses and sentential subjects. *Linguistic Inquiry* 4(3). 363–385.
Lewis, Richard L. 1996. Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research* 25. 93–115.
Levelt, Willem J. 1989. *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.
Mahowald, Kyle, Ariel James, Richard Futrell, & Edward Gibson. 2016. A meta-analysis of syntactic priming in language production. *Journal of Memory and Language* 91. 5–27.
McDonald, Janet L., Kathryn Bock, & Michael H. Kelly. 1993. Word and world order: Semantic, phonological, and metrical determinants of serial position. *Cognitive Psychology* 25(2). 188–230.
Momma, Shota. 2021. Filling the gap is gap filling: Long-distance dependency formation in sentence production. *Cognitive Psychology* 129. 104411.
Momma, Shota & Masaya Yoshida. 2021. Syntax guides sentence planning: evidence from multiple dependency constructions. Short talk at the 34th annual CUNY conference on human sentence processing. Philadelphia, PA, United States, 5–7 March.

Omaki, A, Shin Fukuda, Chizuru Nakao, & Maria Polinsky. 2020. Subextraction in Japanese and subject-object symmetry. *Natural Language and Linguistic Theory* 38. 627–669.

Omaki, Akira, Ellen F. Lau, Imogen Davidson White, Myles L. Dakan, Aaron Apple, & Colin Phillips. 2015. Hyper-active gap filling. *Frontiers in Psychology* 6. 384.

Perlmutter, David M. 1971. *Deep and Surface Structure Constraints in Syntax.* New York: Holt, Reinhart, and Winston.

Pesetsky, David. 2017. Complementizer-trace effects. In M. Everaert & H. van Riemsdijk (eds.), *The Blackwell Companion to Syntax*. 2nd edn. Blackwell.

Pickering, Martin J. & Victor S. Ferreira. 2008. Structural priming: A critical review. *Psychological Bulletin* 134(3). 427.

Pollard, Carl & Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Potter, Mary C. & Linda Lombardi. 1998. Syntactic priming in immediate recall of sentences. *Journal of Memory and Language* 38(3). 265–282.

R Core Team 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. Accessed 22 June 2020.

Reitter, David, Frank Keller, & Johanna D. Moore. 2011. A Computational Cognitive Model of Syntactic Priming. *Cognitive Science* 35(4). 587–637.

Ritchart, Amanda, Grant Goodall, & Marc Garellek. 2016. Prosody and the that-trace effect: An experimental study. In K. Kim, P. Umbal, T. Block, Q. Chan, T. Cheng, K. Finney, M. Katz, S. Nickel-Thompson, & L. Shorten (Eds.), *33rd West Coast Conference on Formal Linguistics*, 320–328. Somerville: Cascadilla Proceedings Project.

Ross, John Robert. 1967. *Constraints on Variables in Syntax*. Cambridge: Massachusetts Institute of Technology dissertation.

Schriefers, Herbert, Encarna Teruel, & Raik-Michael Meinshausen. 1998. Producing simple sentences: Results from picture–word interference experiments. *Journal of Memory and Language* 39(4). 609.

Segaert, Katrien, Linda Wheeldon, & Peter Hagoort. 2016. Unifying structural priming effects on syntactic choices and timing of sentence generation. *Journal of Memory and Language* 91. 59–80.

Smith, Mark & Linda Wheeldon. 1999. High level processing scope in spoken sentence production. *Cognition* 73(3). 205–246.

Smith, Mark & Linda Wheeldon. 2004. Horizontal information flow in spoken sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30(3). 675–686.

Tanaka, Mikihiro N., Holly P. Branigan, Janet F. McLean, & Martin J. Pickering. 2011. Conceptual influences on word order and voice in sentence production: Evidence from Japanese. *Journal of Memory and Language* 65(3). 318–330.

Van Dyke, Julie A. & Brian McElree. 2006. Retrieval interference in sentence comprehension. *Journal of Memory and Language* 55(2). 157–166.

Wanner, Eric & Michael P. Maratsos. 1978. An ATN approach to comprehension. In Morris Hale, Joan Bresnan, & George A. Miller (eds.), *Linguistic Theory and Psychological Reality*. Cambridge: MIT Press.

Wheeldon, Linda & Mark Smith. 2003. Phrase structure priming: A short-lived effect. *Language and Cognitive Processes* 18(4). 431–442.

Wheeldon, Linda, Natalie Ohlson, Aimee Ashby, & Sophie Gator. 2013. Lexical availability and grammatical encoding scope during spoken sentence production. *Quarterly journal of Experimental Psychology* 66(8). 1653–1673.

Zehr, Jeremy & Florian Schwarz. 2018. PennController for Internet Based Experiments (IBEX).

Koichi Otaki, Manami Sato, Hajime Ono, Koji Sugisaki,
Noriaki Yusa, Yuko Otsuka, and Masatoshi Koizumi

# Chapter 9
# Case and word order in children's comprehension of *wh*-questions: A cross-linguistic study

## 1 Introduction

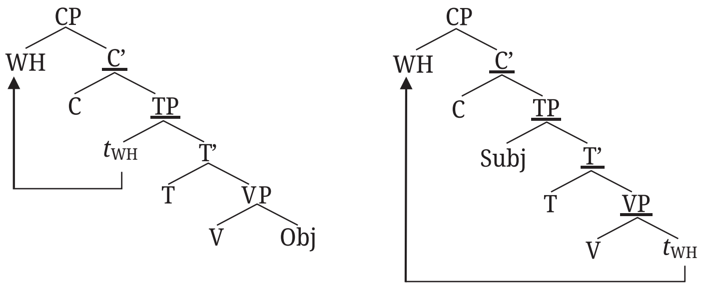
In this chapter, building on our own experimental data of Japanese and other typologically distinct languages, we attempt to elucidate the source of the subject preference widely observed in children's comprehension of *wh*-questions. Subject preference refers to the observation that it is easier for children to comprehend subject *wh*-questions compared to object *wh*-questions. Most of the previous studies relevant to this observation have focused on typologically similar languages with a nominative/accusative case system and S(ubject)-before-O(bject) word order (e.g., English: Tyack and Ingram 1977; Dutch: van der Meer et al. 2001; German: Roesch and Chondrogianni 2014; see also Lau and Tanaka 2021 for a comprehensive review of the subject preference in relative clauses). This limitation, however, makes it difficult to consider the role of case and word order in children's S-over-O preference. In this study, we test several hypotheses concerning children's subject preference against experimental data obtained from monolingual children acquiring Japanese, Tongan, and Kaqchikel. Results from three languages with typologically distinct case and word order characteristics argue in favor of the proposal made by O'Grady (1997) and Hawkins (2004) that the structural distance between a moved *wh*-phrase and its gap is the key factor to explain children's subject preference.

The Structural Distance Hypothesis by O'Grady (1997) and Hawkins (2004) posits that structural distance between an operator (a *wh*-phrase, a relative operator, etc.) that undergoes syntactic movement to a higher functional projection and its gap is reflected in processing costs. Let us see the structures of a subject *wh*-question and an object *wh*-question. In the structure of the subject *wh*-question in (1a), the *wh*-phrase is generated in [Spec, TP] and moves to [Spec, CP], resulting in skipping two intervening nodes (C' and TP). In the structure of the object *wh*-question in (1b), on the other hand, object *wh*-movement skips four intervening nodes (C', TP, T', and VP). The Structural Distance Hypothesis argues that structural distance is determined by the number of intervening nodes between a moved operator and its gap, and that crossing more intervening nodes leads to heavier processing loads.


https://doi.org/10.1515/9783110778946-009

Therefore, the Structural Distance Hypothesis naturally explains the subject preference observed in children's comprehension of *wh*-questions – subject *wh*-questions always result in fewer intervening nodes between a moved *wh*-phrase and its gap than object *wh*-questions.

(1) a. Subject *wh*-question b. Object *wh*-question




In this chapter, we compare the Structural Distance Hypothesis with three other hypotheses regarding children's subject preference: the Conceptual Accessibility Hypothesis, the Case Accessibility Hypothesis, and the Linear Distance Hypothesis. As far as we focus on languages such as English, which has overt *wh*-movement, a nominative/accusative case system, and S-before-O word order, all of these four hypotheses equally predict the subject preference in children's comprehension of *wh*-questions. However, by looking at the acquisition of typologically distinct languages, we can tease apart these hypotheses and pin down the source of children's subject preference. To this end, in the remainder of this chapter, we will report the results from the comprehension experiments that we conducted with children speaking Japanese (Section 2), Tongan (Section 3), and Kaqchikel (Section 4). The results from our cross-linguistic acquisition studies suggest that structural prominence strongly affects children's comprehension of *wh*-questions, thus lending support to the Structural Distance Hypothesis among the four hypotheses.

## 2 Experiment 1: Japanese

In this section, we discuss the Conceptual Accessibility Hypothesis, which is potentially relevant to children's reported subject preference, and analyze whether it can explain children's comprehension of *wh*-questions in languages without overt *wh*-movement, such as Japanese. Many sentence-production studies have shown that agentive, animate, concrete, and salient referents are conceptually more accessible in an event, and tend to be placed in the sentence-initial subject position (cf.

Bock and Warren 1985; Tanaka et al. 2011).[1] In other words, conceptually more accessible subjects are easier to process than other conceptually less accessible referents (such as direct objects and obliques). If conceptual accessibility has an effect on children's comprehension of an event, then children should be able to comprehend conceptually more accessible agent subjects more easily than conceptually less accessible patient objects.

As far as languages having overt *wh*-movement are concerned, the Conceptual Accessibility Hypothesis and the Structural Distance Hypothesis make the same prediction for the acquisition of *wh*-questions: the subject preference. However, different predictions could emerge if we examine languages without overt *wh*-movement. Given that *wh*-phrases in languages without overt *wh*-movement are assumed to undergo covert (LF) *wh*-movement to [Spec, CP] (Huang 1982, Lasnik and Saito 1984, Nishigauchi 1990) and covert *wh*-movement also incurs processing difficulties (cf. Xiang et al. 2014) as overt *wh*-movement does, then both the Conceptual Accessibility Hypothesis and the Structural Distance Hypothesis predict the subject preference. However, if it is only visible filler-gap dependencies that induce processing difficulties (cf. Liu, Hyams, and Mateu 2020), then the Conceptual Accessibility Hypothesis and the Structural Distance Hypothesis make a different prediction: the former predicts the subject preference, while the latter does not (no difficulty should be observed either in subject or object *wh*-questions). To test these predictions, we conducted a comprehension experiment with children speaking Japanese – a language that allows in-situ *wh*-questions (see Yoshinaga 1996 for children's production of *wh*-questions in Japanese).

## 2.1 Participants

Participants were 20 Japanese-speaking children aged 4 to 5 years (mean age, 4;10). They were recruited from a kindergarten in Tsu City, Mie Prefecture.

## 2.2 Materials and procedure

We employed a question-after-story method. The participants were shown two pictures (Figure 1) placed side by side on a laptop computer screen. An experimenter

**1** For example, Bock and Warren (1985) define "conceptual accessibility" as follows: "Conceptual accessibility is the ease with which the mental representation of some potential referent can be activated in or retrieved from memory."

provided each child with a brief explanation for each picture, and then a puppet (manipulated by the other experimenter) asked the child either a subject *wh*-question or an object *wh*-question as in (2) (X indicates the puppet's name). The child's task was to answer these questions; for example, saying "The father" to (2a) and "The mother" to (2b).[2]

(2) Sample context and *wh*-questions

| | | | | | |
|---|---|---|---|---|---|
| Kocchi-no | e-de-wa | onnanoko-ga | okaasan-o | oshita | yo. |
| this-GEN | picture-in-TOP | girl-NOM | mother-ACC | pushed | SFP |

'In this picture, the girl pushed the mother.'

| | | | | | |
|---|---|---|---|---|---|
| Kocchi-no | e-de-wa | otoosan-ga | onnanoko-o | oshita | yo. |
| this-GEN | picture-in-TOP | father-NOM | girl-ACC | pushed | SFP |

'In this picture, the father pushed the girl.'

| | | | |
|---|---|---|---|
| Jaa, | X-ga | shitsumonsuru | yo. |
| now | X-NOM | ask.question | SFP |

'Now, X is going to ask a question.'

a. Subject *wh*-question

| | | | |
|---|---|---|---|
| Dare-ga | onnanoko-o | oshita | kana? |
| who-NOM | girl-ACC | pushed | Q |

'Who pushed the girl?'

b. Object *wh*-question

| | | | |
|---|---|---|---|
| Onnanoko-ga | dare-o | oshita | kana? |
| girl-NOM | who-ACC | pushed | Q |

Lit. 'The girl pushed who?'

Importantly, as Japanese has SOV word order and allows *wh*-phrases to remain in-situ, the object *wh*-phrase *dare-o* in (2b) appears in the base-generated object position. The children were given four *wh*-questions in each of the subject and object conditions, as well as four intransitive subject *wh*-questions as filler items, resulting in a total of 12 test sentences per child.

**2** The following abbreviations are used: 3: third-person, ABS: absolutive, ACC: accusative, AF: agent focus, CL: classifier, CP: completive, DEF: definite, DIM: diminutive, ERG: ergative, GEN: genitive, IC: incompletive, NOM: nominative, pl: plural, PRED: predicate, PRES: present, PROG: progressive, Q: question, RP: resumptive pronoun, SFP: sentence final particle, sg: singular, TOP: topic

**Figure 1:** Sample pictures used with a *wh*-question.

## 2.3 Results

The results of Experiment 1 are summarized in Table 1.

**Table 1:** Summary of Experiment 1 (Japanese).

| | Correct Responses | % of Correct Responses |
|---|---|---|
| Subject *wh*-questions | 75/80 | 93.8% |
| Object *wh*-questions | 78/80 | 97.5% |
| Filler *wh*-questions | 80/80 | 100% |

Table 1 clarifies that Japanese-speaking children had no difficulty in comprehending both subject and object *wh*-questions; they answered both types of *wh*-questions correctly over 90% of the time. Let us recall that the Conceptual Accessibility Hypothesis predicts that it is easier for children to comprehend subject *wh*-questions than object *wh*-questions, irrespective of processing difficulties that covert *wh*-movement may (or may not) incur. The results from child Japanese were not consistent with this prediction; in the acquisition of Japanese, which is a *wh*-in-situ language, children did not show any difficulty comprehending either the subject *wh*-questions or the object *wh*-questions, supporting the view that covert *wh*-movement incurs no processing difficulties. These findings suggest that conceptual accessibility is not a major factor to explain children's subject preference in languages such as English, and that children's difficulty in comprehending object *wh*-questions reported in previous studies is caused by overt syntactic movement (or filler-gap dependency) (but see Section 5 for a discussion of the potential effects of conceptual accessibility).[3]

## 3 Experiment 2: Tongan

It is well known that there exists a hierarchy that underlies the availability of NP movement in *wh*-questions and relative clauses cross-linguistically. For instance, Keenan and Comrie (1977, 1979) propose the Noun Phrase Accessibility Hierarchy in (3).[4]

(3) Noun Phrase Accessibility Hierarchy (Keenan and Comrie 1977, 1979)
Subject >> Direct Object >> Indirect Object >> Oblique >> Genitive
('A >> B' means A is syntactically more prominent than B.)

This hierarchy states that if a language allows extraction of NPs of a particular type (say oblique, for example), then NPs that are higher in the hierarchy (subject, direct object, and indirect object) are also allowed to undergo extraction. This is because the higher an NP ranks in the hierarchy, the more prominent it is syntactically. Given that the Noun Phrase Accessibility Hierarchy has a direct bearing on sentence processing (Hawkins 1999), the subject preference in the acquisition of *wh*-questions receives a natural explanation: as the subject is ranked higher than the direct object in the hierarchy, the former is easier to process/comprehend than the latter.

**3** Consistent with the results of our Japanese experiment, Liu et al. (2020) report that children speaking Mandarin Chinese – another *wh*-in-situ language – have no difficulty in comprehending in-situ embedded (subject and object) *wh*-questions. Interestingly, they also report that the subject preference emerges among children at age 5–6 when they comprehend corresponding sluicing-like constructions. They argue that this is because sluicing-like constructions in Mandarin Chinese optionally require focus movement (followed by ellipsis), and it is only after a certain age that children can use such a grammatical option. In particular, focus movement triggers intervention effects in object sluicing-like constructions, and the observed subject preference results.

Liu et al.'s (2020) study suggests that the subject preference could be observed in Japanese if we look at more complex cases involving A'-dependences such as sluicing, clefts, and relative clauses. (We thank a reviewer for bringing this point to us.) In fact, Suzuki (2011) reports that, after controlling confounding factors relevant to case, Japanese-speaking children can comprehend both subject and object relative clauses very well (i.e., no subject preference). Given the logic of Liu et al. (2020), this supports the view that relative clauses in Japanese do not involve A'-movement (e.g., Murasugi 1991; Saito 1985, among others).

**4** Part of this section was reported in Otaki et al. (2020).

The Noun Phrase Accessibility Hierarchy in (3), however, faces a problem when we consider languages with ergative/absolutive case alignment (Keenan and Comrie 1977, 1979; Aldridge 2008). In a certain number of ergative/absolutive languages, NPs with absolutive case (i.e., intransitive subjects and transitive objects) undergo *wh*-movement and relativization leaving a gap behind, whereas NPs with ergative case (i.e., transitive subjects) do not, resorting to other syntactic or morphological operations (such as anti-passivization and resumptive pronouns). In other words, the cross-linguistic availability of NP extraction cannot be explained in terms of grammatical relations such as subject and object.

Building on the insights of Marantz (1991) and Chomsky (1993), Otsuka (2006) proposes that the Accessibility Hierarchy should be restated in terms of case, as illustrated in (4).

(4) Case Accessibility Hierarchy (Otsuka 2006)
Unmarked Case (NOM/ABS) >> Marked Case (ACC/ERG) >> Oblique

The Case Accessibility Hierarchy in (4) states that NPs with unmarked case (nominative and absolutive, which syntactically appear as the sole argument of an intransitive verb and phonologically tend to receive a zero exponent) are syntactically more prominent than NPs with marked case (accusative and ergative) and oblique case. This explains the cross-linguistic observation that extraction of NPs with unmarked case is more widely available than extraction of NPs with other types of case. Note that both the Noun Phrase Accessibility Hierarchy in (3) and the Case Accessibility Hierarchy in (4) are markedness hierarchies, which predict that unmarked structures tend to be acquired prior to marked structures, and the only difference between the two hierarches lies in whether they are defined by grammatical relations or case.

Most of the previous studies regarding the acquisition of *wh*-questions have targeted languages with nominative/accusative case alignment, for which the Case Accessibility Hierarchy in (4) uniformly predicts the subject/nominative preference, and only a few studies investigated the acquisition of *wh*-questions in ergative/absolutive languages (but see Gutierrez-Mangado 2011; Muāgututi'a 2017, for notable exceptions). Interestingly, the Case Accessibility Hierarchy makes a different prediction for the acquisition of *wh*-questions in ergative/absolutive languages, namely an absolutive object preference. This is because NPs with absolutive case are considered syntactically more prominent compared to NPs with ergative case. In this section, we report the design and results of our experimental study that investigated the acquisition of *wh*-questions in Tongan, an Austronesian language with ergative/absolutive case alignment.

## 3.1 Tongan grammar

Before going into the details of the experiment, let us take a quick detour to basic Tongan grammar. First, the basic word order is VSO in Tongan, as shown in (5).[5]

(5) VSO sentence in Tongan

| | | | | | | |
|---|---|---|---|---|---|---|
| Na'e | 'ofa'i | ['e | Sione] | ['a | e | fefine]. |
| PAST | love | ERG | John | ABS | DEF | woman |

'John loved the woman.'

In (5), the combination of the past tense marker *na'e* and the verb *'ofa'i* (love) comes first, followed by the ergative subject *'e Sione* and the absolutive object *'a e fefine*. This VSO word order is considered to be the most basic word order in Tongan, although other orders such as VOS, SVO, and OVS are also possible (Otsuka 2000; Custis 2004).

Second, Tongan exhibits an ergative/absolutive case alignment. As we can see in (6), the subject of the intransitive sentence *e fefine* (the woman) bears the absolutive case marker *'a*, which is identical to the one used with the transitive object in (5).

(6) Intransitive sentence in Tongan

| | | | | | | |
|---|---|---|---|---|---|---|
| Na'e | 'alu | ['a | e | fefine] | ki | Tonga. |
| PAST | go | ABS | DEF | woman | to | Tonga |

'The woman went to Tonga.'

Third, *wh*-phrases in Tongan can either stay in the original position or move to the sentence-initial position, as exemplified in (7) and (8), respectively.[6]

(7) In-situ absolutive object *wh*-question

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 'Oku | tuli | 'e | he | sipi | 'a | e | manu | fē? |
| PRES | chase | ERG | DEF | sheep | ABS | DEF | animal | which |

Lit. 'The sheep is chasing which animal?'

5 Strictly speaking, the article *e* (allomorph *he*) indicates specificity and not definiteness. The latter is expressed in Tongan phonologically as *definitive accent*, stress on the final vowel of the final word of the relevant noun phrase, orthographically indicated as an acute accent, as in *fefiné* vs. *fefine*. In this article, however, we gloss *e/he* as definite and dispense with orthographic representation of definitive accent in Tongan examples for the sake of simple exposition.

6 In what follows, Δ indicates the position of the gap resulting from syntactic operator movement.

(8) Absolutive object question with a *wh*-phrase in sentence-initial position

Ko e manu fē 'oku tuli 'e he sipi Δ ?
PRED DEF animal which PRES chase ERG DEF sheep
'Which animal is the sheep chasing?'

The sentence in (7) is an in-situ *wh*-question, where the *wh*-phrase *'a e manu fē* (which animal) remains in the original object position. In (8), however, the *wh*-phrase moves to the sentence-initial position, leaving behind a gap. Note that the moved *wh*-phrase is accompanied by the particle *ko*, which marks predicate nominals. This is because *wh*-questions like (8) involve a pseudo-cleft construction, the structure of which is illustrated in (9).

(9) Structure of (8)

Ko [e manu fē] [$_{CP}$ $OP_i$ 'oku tuli 'e he sipi $\Delta_i$ ]?
PRED DEF animal which PRES chase ERG DEF sheep

In this pseudo-cleft construction, it is not the *wh*-phrase *per se* but the null operator *OP* that undergoes movement.[7] Nevertheless, *wh*-questions such as (9) differ from *wh*-in-situ questions in that the former involves the A-bar extraction.

Lastly, ergativity in Tongan is not limited to the domain of morphology (e.g., morphological forms of case markers), but it also appears in the domain of syntax. Let us consider the example of an ergative subject *wh*-question in (10).

(10) Ergative subject *wh*-question

Ko e manu fē 'oku ne tuli Δ 'a e sipi?
PRED DEF animal which PRES RP chase ABS DEF sheep
'Which animal is chasing the sheep?'

Much like the absolutive object *wh*-question in (8), the *wh*-phrase appears in the sentence-initial position with the predicate particle *ko*. What is different from the absolutive object *wh*-question in (8) is that the clitic pronoun *ne* appears between

7 One piece of evidence showing that the pseudo-cleft construction in (9) involves A'-movement comes from the fact that it shows a weak crossover effect, as shown in (i) (Otsuka 2005: 251).

(i) *Ko $hai_i$ [$_{CP}$ $OP_i$ na'e fili 'e $he'ene_i$ tamai $\Delta_i$ ]?
PRED who PAST choose ERG his father
'$Who_i$ did $his_i$ father choose?'

the tense marker and the verb. This kind of resumptive pronoun is obligatory when an ergative subject *wh*-phrase appears in the sentence-initial position.

The ergative nature of Tongan grammar provides us with a good testing ground for investigating the extent to which the Case Accessibility Hierarchy is responsible for children's subject preference. As absolutive case – being unmarked – is ranked higher than ergative case, it is predicted that Tongan-speaking children will show an absolutive object preference.

## 3.2 Participants

To examine whether Tongan-speaking children show the absolutive-object preference, we conducted a comprehension experiment with 27 Tongan-speaking children aged 4 to 5 years (mean age, 4;10). Additionally, 60 Tongan-speaking adults (mean age, 24;02) also participated as a control group. Children were recruited from a kindergarten in Nuku'alofa, Tonga. Adult participants were students and staff members recruited from the University of South Pacific, Tonga Campus.

## 3.3 Materials and procedure

In this experiment, we used the four types of Tongan *wh*-questions in (11) as test items.

(11) a. Absolutive-subject question (ABS SUBJ)

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Ko | **e** | **manu** | **fē** | 'oku | hiva | Δ | mo | e | pusi? |
| PRED | DEF | animal | which | PRES | sing | | with | DEF | cat |

'Which animal is singing with the cat?'

b. Ergative-subject question (ERG SUBJ)

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Ko | **e** | **manu** | **fē** | 'oku | ne | tuli | Δ | 'a | e | pusi? |
| PRED | DEF | animal | which | PRES | RP | chase | | ABS | DEF | cat |

'Which animal is chasing the cat?'

c. Absolutive-object question (ABS OBJ)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Ko | **hai** | 'oku | teke | 'e | he | fa'ē | Δ ? |
| PRED | who | PRES | push | ERG | DEF | mother | |

'Who is the mother pushing?'

d. In-situ absolutive-object question (IN-SITU)

| | | | | | | |
|---|---|---|---|---|---|---|
| 'Oku | teke | 'e | he | ta'ahine | 'a | **hai**? |
| PRES | push | ERG | DEF | girl | ABS | who |

Lit. 'The girl is pushing who?'

The example in (11a) is an absolutive-subject *wh*-question containing the intransitive verb *hiva* (sing). The second example in (11b) is an ergative-subject *wh*-question with the resumptive pronoun *ne* (3sg.). The third example in (11c) is an absolutive-object *wh*-question. The last example in (11d), which is an absolutive-object *wh*-question, differs from the other three types in that the *wh*-phrase *hai* (who) does not undergo overt *wh*-movement and remains in the original position. Each sentence type had three tokens (one with human characters and *hai* (who), and the others with animal characters and *manu fē* (which animal)), yielding a total of 12 test sentences. The order of the test sentences was semi-randomized so that participants did not hear the same sentence types consecutively.

An experimenter, who is a native speaker of Tongan, asked *wh*-questions to child participants with pictures shown on a computer display during the experiment. Prior to the main experiment, we had a practice session in which participants were asked some simple *wh*-questions, as illustrated in Figure 2.

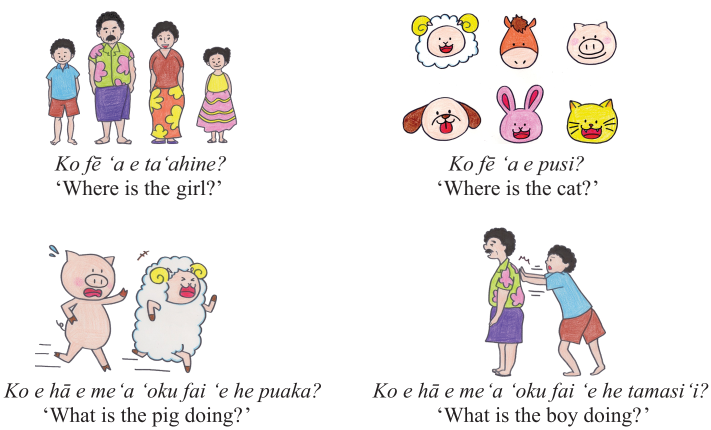



**Figure 2:** Sample pictures and questions used in the practice session.

If participants either hesitated to answer or gave an incorrect response, the experimenter provided them with the correct answer. This ensured that the participants knew the words referring to the characters and actions used in the main experiment.

In the main experiment involving intransitive actions, participants were presented with intransitive *wh*-questions such as (11a) using pictures in which two of the characters perform the same intransitive action (e.g., singing), while the other performs a different intransitive action (e.g., listening), as shown in Figure 3.[8, 9]

*Ko e manu fē ʻoku hiva mo e pusi?*
ʻWhich animal is singing with the cat?ʼ

**Figure 3:** Sample picture and test sentence in the intransitive condition.

If participants correctly understand the absolutive-subject *wh*-question above, they are expected to choose "dog" as their answer.

Test sentences involving transitive actions such as (11b), (11c), and (11d) were presented with pictures as in Figure 4, where one character acts on another, who in turn acts on yet another character in a transitive manner (cf. Longenbaugh and Polinsky 2016). The correct answer for the ergative subject *wh*-question above is "rabbit," but there is a good chance that the children may choose "dog" as the answer in case they misunderstand it as the absolutive-object *wh*-question. Test trials were split into two lists, half presented in an inverted order and with mirror-image pictures to avoid potential order and directionality effects.

## 3.4 Predictions

If the Case Accessibility Hierarchy, discussed at the beginning of this section, affects children's processing/comprehension of *wh*-questions, it makes a different prediction from the Structural Distance Hypothesis for the comprehension of the test

**8** Immediately prior to presenting the test sentences, the experimenter again asked the children the names of the characters, using the expressions like "Ko e hā ē? (What's this?)" and "Ko hai ē? (Who is this?)." Just like our Japanese and Kaqchikel experiments in which we provided children with simple contexts before presenting the test sentences, the simple questions asking character names helped the children access relevant lexical items that they were expected to use in the experiment.

**9** Zuckerman et al. (2016) indicate that presenting multiple events at the same time in the picture selection task sometimes interferes with children's comprehension of a particular type of sentences (such as passives). To avoid this problem, unlike our Japanese and Kaqchikel experiments where we had used two pictures involving different events, we used a single picture in which three characters perform an intransitive or transitive action in the Tongan experiment.

*Ko e manu fē 'oku ne tuli 'a e pusi?*
'Which animal is chasing the cat?'

**Figure 4:** Sample picture and test sentence in the transitive conditions.

sentences given in (11b) and (11c), namely the ERG SUBJ condition and the ABS OBJ condition. The Structural Distance Hypothesis predicts that it will be easier for Tongan-speaking children to comprehend ERG SUBJ *wh*-questions in (11b) compared to ABS OBJ *wh*-questions in (11c) because subjects are generated in a structurally higher position than objects. The Case Accessibility Hypothesis, on the other hand, predicts that ERG SUBJ *wh*-questions will be more difficult for Tongan-speaking children to comprehend than ABS OBJ *wh*-questions as ergative case – being marked – is lower in the Case Accessibility Hierarchy in (4) than absolutive case.

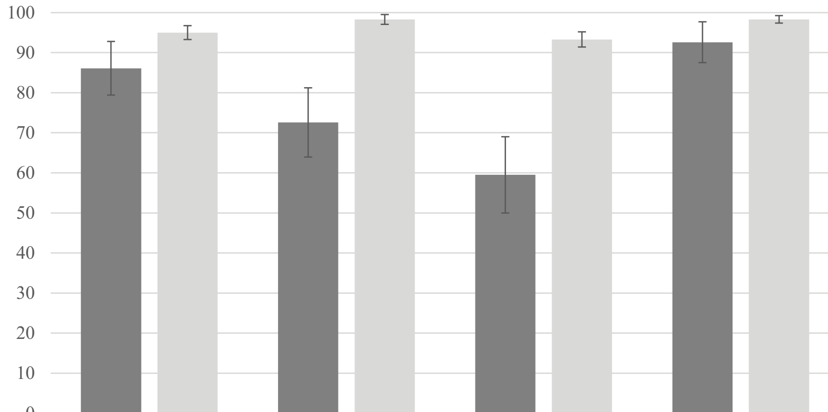
The (intransitive) ABS SUBJ condition in (11a) constitutes a baseline and children should have no difficulty comprehending this type of *wh*-questions as no competing argument NPs exist within the sentence. Likewise, the IN-SITU (absolutive-object) condition in (11d) is also predicted to be easy for children, because there is no movement involved in the sentence and children do well with in-situ *wh*-questions, as we have already seen in the Japanese experiment.

## 3.5 Results

The results of Experiment 2 are summarized in Figure 5 (the error bars indicate standard errors).

Adult controls behaved as expected, with over 90% correct responses across the conditions. Child participants also performed well with the IN-SITU condition (92.6%) and the ABS SUBJ condition (86.1%). Their performances on the ERG SUBJ condition and the ABS OBJ were less reliable; the correct response rates for each condition were 72.6% and 59.5%, respectively.[10]

**10** The following three types of verbs are used in the transitive conditions: *teke* (push), *tuli* (chase), and *'emo* (lick). We checked the correct response rates by verb types and found no differences: *teke* (78.1%), *tuli* (77.9%), and *'emo* (78.0%). This ensures it is not the case that the participants faced difficulty with a particular sentence type but not with the others.



| | ABS SUBJ | ERG SUBJ | ABS OBJ | IN-SITU |
|---|---|---|---|---|
| Child | 86.1 | 72.6 | 59.5 | 92.6 |
| Adult | 95 | 98.3 | 93.3 | 98.3 |

**Figure 5:** Summary of Experiment 2 (Tongan).

The child data were analyzed using logistic mixed-effects models (Baayen, Davidson, and Bates 2008) fitted with the *glmer* function of the *lme4* package in R (Bates et al. 2015). The models included the independent variables *sentence type* (i.e., ABS SUBJ, ERG SUBJ, ABS OBJ, and IN-SITU) as fixed factors. They were dummy-coded and ERG SUBJ was set as the reference level in the analysis so that we could see the contrast between this condition and the other conditions (the contrast between the ERG SUB condition and the ABS OBJ condition, in particular). We also included both the participants and the items as random factors. The dependent variable was whether the response was correct (coded as 1) or not (coded as 0). Model selection was performed using the backward stepwise method, comparing models using the *anova* function of the *lme4* package.

The analysis revealed significant differences between the ERG SUBJ condition and the ABS SUBJ condition ($\beta = 0.94$, $SE = 0.43$, $z = 2.17$, $p = 0.03$), and between the ERG SUBJ condition and the IN-SITU condition ($\beta = 1.68$, $SE = 0.51$, $z = 3.28$, $p < 0.01$). Additionally, the difference between the ERG SUBJ condition and the ABS OBJ condition was marginally significant ($\beta = -0.63$, $SE = 0.37$, $z = -1.72$, $p = 0.08$).

These results suggest the following three points. First, the difficulty in comprehending *wh*-questions occurs only when a *wh*-phrase undergoes movement.[11] This

11 This does not mean that overt *wh*-movement always triggers processing difficulties. In fact, the Tongan-speaking children did well in the ABS SUBJ condition, suggesting that the presence of an intervening argument is also relevant for children's processing difficulties. For example, the Relativized Minimality Account (e.g., Friedmann, Belletti, and Rizzi 2009) claims that a phrase that

**Table 2:** Summary of the fixed effects in the logistic mixed-effects model.

| | Estimate | *SE* | *z*-value | *p* | |
|---|---|---|---|---|---|
| Intercept | 1.09 | 0.32 | 3.40 | < 0.01 | |
| ABS SUBJ | 0.94 | 0.43 | 2.17 | 0.03 | * |
| ABS OBJ | −0.63 | 0.37 | −1.72 | 0.08 | + |
| IN-SITU | 1.68 | 0.51 | 3.28 | < 0.01 | ** |

is consistent with the results of our Japanese experiment (Experiment 1), which also observed that children had no difficulty in comprehending in-situ *wh*-questions. Second, the fact that comprehending ergative subject *wh*-questions was more difficult for children than comprehending absolutive subject *wh*-questions suggests that ergativity hinders children's comprehension of *wh*-questions to some extent. This is consistent with the prediction of the Case Accessibility Hypothesis and also with previous findings that ergativity is acquired late (e.g., Muāgututi'a 2017). Lastly, the effect of structural distance on children's comprehension of *wh*-questions is still robust, as we can see in the difference (though marginal) between the ERG SUBJ condition and ABS OBJ conditions. In other words, the difficulty in comprehending absolutive-object *wh*-questions suggests that the effect of structural distance on comprehending *wh*-questions arises independently of the Case Accessibility Hierarchy.

We believe that the findings reported in this section contribute to further understanding of the source of children's subject preference, because only by looking at the acquisition of *wh*-questions in ergative/absolutive languages can we tease apart the effects of structural distance and case accessibility. However, there remains another possibility that linear distance could have a potential effect on children's subject preference. As Tongan is a language with VSO word order, subject *wh*-movement always results in shorter linear distance than object *wh*-movement (and the same is true for other languages that have S-before-O word order). To discuss the effect of linear distance on children's comprehension of *wh*-questions, we will report in the next section the results of the experiment that we conducted with children speaking a language called Kaqchikel.

structurally intervenes between a moved *wh*-phrase and its gap counts as an intervener only when it has identical morphological features as the moved *wh*-phrase. As the purpose of our experiments is to first investigate whether the subject preference holds in typologically different languages using the most basic paradigm employed in previous studies (with sentences having two animate arguments), we leave for future research to test predictions of the Relativized Minimality Account.

## 4 Experiment 3: Kaqchikel

So far we have considered languages such as English, Japanese, and Tongan, all of which have S-before-O word order. One might suspect that the subject preference observed in children's comprehension of *wh*-questions simply reflects shorter linear distance involved in subject *wh*-questions, as schematically represented in (12) and (13).

(12) English (SVO)

a. Subject *wh*-question

Who Δ is chasing the boy?

b. Object *wh*-question

Who is the boy chasing Δ ?

(13) Tongan (VSO)

a. Subject *wh*-question

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Ko | **hai** | ‘oku | ne | tuli | Δ | ‘a | e | tamasi‘i? |
| PRED | who | PRES | RP | chase | | ABS | DEF | boy |

b. Object *wh*-question

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Ko | **hai** | ‘oku | tuli | ‘e | he | tamasi‘i | Δ ? |
| PRED | who | PRES | chase | ERG | DEF | boy | |

In S-before-O word order languages, object *wh*-questions necessarily have linearly longer *wh*-movement, because the object *wh*-phrase must move to the sentence-initial position by skipping over the subject. If linear distance involved in *wh*-movement has some impact on the processing/comprehension of *wh*-questions (and children are more susceptive to such effects than adults), then children's subject preference receives a natural explanation.

In fact, among the six logically possible word orders (i.e., SOV, SVO, VSO, VOS, OVS, and OSV), a vast majority of languages in the world have S-before-O word order (SOV, SVO, and VSO), and languages that show basic O-before-S word order (VOS, OVS, and OSV) are quite rare (3.5%, according to Dryer 2005). To the best of our knowledge, all of the language acquisition studies that investigated the comprehension of *wh*-questions have dealt with S-before-O word order languages, and the question, to what extent linear distance involved in *wh*-movement affects children's processing/comprehension of *wh*-questions, still remains to be answered.

In this section, we report on an experiment that we conducted with children speaking Kaqchikel, a Mayan language with basic VOS word order. Before going into the details of our experiment, we will briefly review the fundamental grammatical properties of Kaqchikel in the following sub-section.

## 4.1 Kaqchikel grammar

First, let us consider a simple transitive sentence in Kaqchikel.

(14) Kaqchikel VOS sentence

| | | | | |
|---|---|---|---|---|
| X-ø-u-b'a' | ri | tz'i' | ri | me's. |
| CP-3sg.ABS-3sg.ERG-bite | the | dog | the | cat |

'The cat bit the dog.'

In (14) there are two noun phrases following the verb: the first is *ri tz'i'* (the dog) and the second is *ri me's* (the cat), and this whole sentence means "the cat bit the dog," suggesting that the sentence has VOS word order. Although Kaqchikel also allows other word orders such as SVO and VSO, VOS is argued to be the syntactically canonical, basic word order from both theoretical and psycholinguistic viewpoints (García Matzar and Rodríguez Guaján 1997; Broadwell and Smith 2001; Koizumi et al. 2004; Yasunaga et al. 2015, among others).

Second, Kaqchikel is a head-marking and morphologically ergative language. As we can see in (14), the subject and the object are not overtly case-marked. Instead, they are obligatorily cross-referenced with the agreement morphemes attached to the verb: the subject *ri me's* (the cat) is cross-referenced with the third-person singular ergative marker *-u*, and the object *ri tz'i'* (the dog) is cross-referenced with the third-person singular absolutive morpheme, which has a zero exponent in Kaqchikel.

Third, Kaqchikel has obligatory overt *wh*-movement, as shown in (15).

(15) a. Subject *wh*-question

| | | | | | |
|---|---|---|---|---|---|
| Achike | x-ø-nim-o | la | ti | xtän | Δ ? |
| who | CP-3sg.ABS-push-AF | the | DIM | girl | |

'Who pushed the girl?'

b. Object *wh*-question

| | | | | | |
|---|---|---|---|---|---|
| Achike | x-ø-u-nim | Δ | la | ti | xtän? |
| who | CP-3sg.ABS-3sg.ERG-push | | the | DIM | girl |

'Who did the girl push?'

After the *wh*-phrase *achike* (who) undergoes movement, the subject *wh*-question in (15a) and the object *wh*-question in (15b) result in the same word order, that is, “WH V NP.” The only difference between these two types of *wh*-questions lies in the form of the verb: in the subject *wh*-question, a special morphology called Agent Focus (AF) is used and the verb is intransitivized, as we can see that the verb in (15a) exhibits only an absolutive agreement, whereas in the object *wh*-question, the AF morphology is not necessary and the verb continues to be transitive (with both ergative and absolutive agreement).

What is important for the purpose of this study is the fact that, unlike *wh*-questions found in S-before-O word order languages, object *wh*-questions have linearly shorter *wh*-movement than subject *wh*-questions, as illustrated in (16).

(16) a. Subject *wh*-question

| | | | | | |
|---|---|---|---|---|---|
| Achike | x-ø-nim-o | la | ti | xtän | Δ ? |
| who | CP-3sg.ABS-push-AF | the | DIM | girl | |

b. Object *wh*-question

| | | | | | |
|---|---|---|---|---|---|
| Achike | x-ø-u-nim | Δ | la | ti | xtän? |
| who | CP-3sg.ABS-3sg.ERG-push | | the | DIM | girl |

Lastly, Otaki et al. (2019) report that subjects are structurally higher than objects even in sentences with VOS word order in Kaqchikel, based on the facts in (17) (see also Henderson 2012 for similar data). In (17a) the subject is the conjoined NP *a Lolmay chunqa’ a Xwan* (Lolmay and Juan), which triggers ergative third-person plural agreement, *ki*, and the object is the anaphor *k-i’* (each other), which is cross-referenced with absolutive third-person singular agreement, ø, a phonetically null morpheme. In (17b), the order of the subject and the object is reversed, and the verb bears an ergative third-person singular agreement (cross-referencing the subject anaphor) and absolutive third-person plural agreement (cross-referencing the conjoined NP object). The fact that (17a) is acceptable but (17b) is not indicates that subjects c-command objects in Kaqchikel VOS sentences (Condition A of Binding Theory, Chomsky 1981).

(17) Reciprocal binding in Kaqchikel (Otaki et al. 2019)

a.

| | | | | |
|---|---|---|---|---|
| X-ø-ki-tz’ët | (jub’ey chik) | k-i’ | a | Lolmay |
| CP-3sg.ABS-3pl.ERG-see | (once again) | each.other | CL | Lolmay |

| | | |
|---|---|---|
| chunqa’ | a | Xwan. |
| and | CL | Juan |

‘Lolmay and Juan saw each other (again).’

b.

| | | | |
|---|---|---|---|
| *X-e'-ru-tz'ët | (jub'ey chik) | a | Lolmay |
| CP-3pl.ABS-3sg.ERG-see | (once again) | CL | Lolmay |

| | | | |
|---|---|---|---|
| chunqa' | a | Xwan | k-i' |
| and | CL | Juan | each.other |

Given the grammatical properties above, Kaqchikel *wh*-questions provide a good testing ground for the role of linear distance in children's comprehension of *wh*-questions. More specifically, the Linear Distance Hypothesis in (18) predicts that object *wh*-questions should be easier for Kaqchikel-speaking children to process/comprehend than subject *wh*-questions, because object *wh*-movement has shorter linear distance compared to subject *wh*-movement.

(18) The Linear Distance Hypothesis (cf. Gibson 1998, 2000)
Longer linear distance between a filler and a gap incurs more processing difficulty.

In contrast, the Structural Distance Hypothesis predicts the opposite behavior. As discussed above, subjects are generated in a structurally higher position than objects even in VOS sentences. Therefore, the Structural Distance Hypothesis predicts that subject *wh*-questions should be easier for Kaqchikel-speaking children to process/comprehend than object *wh*-questions.

## 4.2 Participants

Eighteen Kaqchikel-speaking children aged from 5;03 to 6;03 years (mean age, 5;08) participated in our comprehension experiment conducted in Guatemala. They spoke Kaqchikel as their primary language at home, although some of the participants also used Spanish as their second language.

## 4.3 Materials and procedure

The method used in the Kaqchikel experiment was similar to that employed in the Japanese experiment. In each trial, a child was presented with two pictures depicting different events (Figure 6) and an experimenter – a native speaker of Kaqchikel – asked a *wh*-question after narrating a short story that explained the situations depicted in each of the pictures, as in (19).

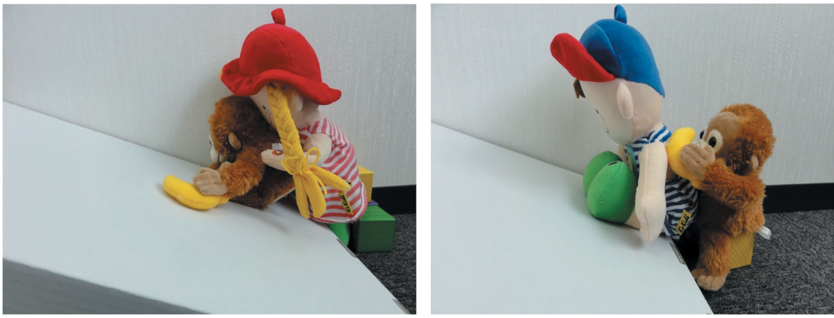


**Figure 6:** Sample pictures used with target *wh*-questions.

(19) Sample story and test sentences
In this picture, the girl is pushing the monkey.
In this picture, the monkey is pushing the boy.
Now, I'm going to ask you a question:

a. Subject *wh*-question [x 3]

| | | | | | | |
|---|---|---|---|---|---|---|
| Achike | tajin | ni-ø-chakmay-in | la | ti | k'oy | Δ ? |
| who | PROG | IC-3sg.ABS-push-AF | the | DIM | monkey | |

'Who is pushing the monkey?'

b. Object *wh*-question [x 3]

| | | | | | | |
|---|---|---|---|---|---|---|
| Achike | tajin | n-ø-u-chakmayij | Δ | la | ti | k'oy? |
| who | PROG | IC-3sg.ABS-3sg.ERG-push | | the | DIM | monkey |

'Who is the monkey pushing?'

In addition to the two types of *wh*-questions in (19), participants were given filler *wh*-questions such as (20), in which inanimate objects, instead of animate characters, were used (see also Figure 7 for sample pictures used with filler *wh*-questions).

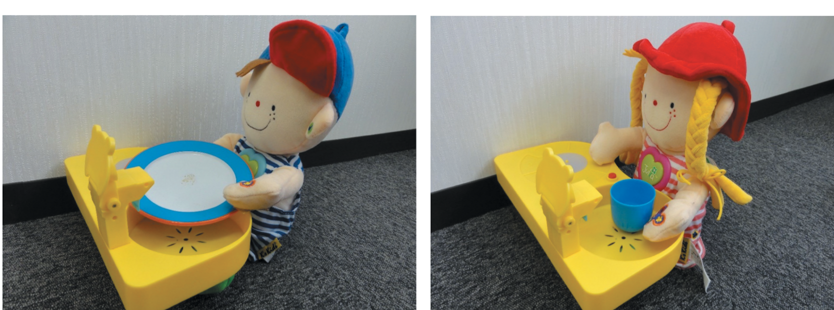


**Figure 7:** Sample pictures used with filler *wh*-questions.

(20) Sample filler *wh*-questions

a. Subject *wh*-question

| | | | | | |
|---|---|---|---|---|---|
| Achike | tajin | ni-ø-ch'aj-o | la | taza | Δ ? |
| who | PROG | IC-3sg.ABS-wash-AF | the | cup | |

'Who is washing the cup?'

b. Object *wh*-question

| | | | | | | |
|---|---|---|---|---|---|---|
| Achike | tajin | n-ø-u-ch'äj | Δ | la | ti | ala'? |
| what | PROG | IC-3sg.ABS-3sg.ERG-wash | | the | DIM | boy |

'What is the boy washing?'

Some previous studies report that children's difficulty in comprehending object *wh*-questions arises only when the moved *wh*-phrase shares features with the intervening subject (Friedmann et al. 2009). Therefore, participants are expected to do well with these filler *wh*-questions because the subject and the object used in each filler sentence do not share features in animacy. Each child participant was given 10 test sentences in total: three target subject *wh*-questions, three target object *wh*-questions, two filler subject *wh*-questions, and two filler object *wh*-questions.

## 4.4 Results

The results are summarized in Table 3 below. The Kaqchikel-speaking children who participated in our experiment did very well with the filler *wh*-questions (97.2%), which ensured that they understood what they were expected to do in the experiment and that processing difficulty in comprehending object *wh*-questions did not emerge when the moved *wh*-phrase had different features than the intervening subject. As for the target *wh*-questions, they understood subject *wh*-questions better compared to object *wh*-questions (77.8% vs. 57.4%). The results of the target *wh*-questions were analyzed using logistic mixed-effects models (Baayen et al. 2008) fitted with the *glmer* function of the *lme4* package in R (Bates et al. 2015). The models included the independent variables *sentence type* (i.e., SUBJ-WH and OBJ-WH) as fixed factors, and the participants and the items as random factors. The dependent variable was whether the response was correct (coded as 1) or not (coded as 0). Model selection was performed using the backward stepwise method, comparing models using the *anova* function of the *lme4* package. The analysis revealed significant difference between the SUBJ-WH condition and the OBJ-WH condition ($\beta = -0.96$, $SE = 0.43$, $z = -2.22$, $p = 0.03$).[12, 13, 14]

**Table 3:** Summary of Experiment 3 (Kaqchikel).

| | Correct Responses | % of Correct Responses |
|---|---|---|
| Subject *wh*-questions | 42/54 | 77.8% |
| Object *wh*-questions | 31/54 | 57.4% |
| Filler *wh*-questions | 70/72 | 97.2% |

These results are incompatible with the Linear Distance Hypothesis, which incorrectly predicts the opposite behavior, namely the object preference. The Structural Distance Hypothesis, on the other hand, is consistent with the results obtained in the experiment: subject *wh*-questions are easier to process/comprehend compared to object *wh*-questions as subject *wh*-movement is structurally shorter than object *wh*-movement.

**12** The two incorrect responses were made with the filler object *wh*-question "Achike tajin n-ø-u-tïj la ti ixtän? (What is the girl eating?)".

**13** Additionally, the two-tailed binomial tests show that their responses on object *wh*-questions did not differ from chance level ($p = 0.34$), while they answered correctly on subject *wh*-questions significantly more often than would be expected by chance ($p = 0.001$). This indicates that the children were incapable of comprehending object *wh*-questions and resorted to guessing when giving their answers.

**14** As one of the reviewers points out, the correct response rates in both subject and object *wh*-questions look relatively low (77.8% and 57.4%), considering participants' age (5–6 years old). There seem to be at least two potential reasons for their general difficulty in comprehending Kaqchikel *wh*-questions. First, as shown in (15), the only cue for distinguishing between subject and object *wh*-questions lies in a slight difference in verbal agreement morphology, and it might be the case that some Kaqchikel-speaking children have problems with detecting the difference in verbal agreement even after age 5. (Unfortunately, we do not have data concerning how children deal with verbal agreement in simple declarative sentences, but it is interesting to see if children are sensitive to differences in meaning arising from different verbal morphology.) Second, there remains a possibility that the contexts given right before the test sentences were not felicitous. More specifically, the contexts we provided in the experiment corresponded to the answers of the target *wh*-questions. It might be strange to ask a question whose answer was already given in the context right before the question, and this might have affected children's performance. To make the *wh*-questions felicitous, it is better to use contexts such as: "This is mother, a girl, and a father. They are doing some pushing event, aren't they?" (We thank the reviewer for suggesting this solution to us.)

# 5 Discussion and conclusion

The findings from the series of our cross-linguistic experiments are summarized in Table 4, along with the comparative information on English as a representative nominative/accusative S-before-O language. The most important finding is that the effect of structural distance arises even after the other relevant potential factors such as conceptual accessibility, case accessibility, and linear distance are controlled. This suggests that children refer to the hierarchical structure when processing/comprehending *wh*-questions but sometimes fail to establish a filler-gap dependency because of their limited processing capacity in integrating structurally distant elements.

**Table 4:** Summary of the cross-linguistic experiments.

| | English | Japanese (4;10) | Tongan (4;10) | Kaqchikel (5;08) |
|---|---|---|---|---|
| **Conceptual Accessibility** | ✓ | × | × (in situ) | ✓ |
| **Case Accessibility** | ✓ | ✓ | Δ | – (no case) |
| **Linear Distance** | ✓ | – (no wh-mvt) | ✓ | × |
| **Structural Distance** | ✓ | ✓ | ✓ | ✓ |

The experimental data from Kaqchikel-speaking children show that children do not rely on the linear distance, which is readily detectable from actual speech and hence would be easy for children to make use of. Instead, they rely on the structural distance, which neither appears on surface word order nor is directly detectable from what they actually hear in conversation. This is consistent with the view that the application of rules in human language refers to structures, not to linear orderings, and even young children adhere to structural information when they produce/comprehend sentences (e.g., Crain and Nakayama 1987; Gualmini and Crain 2005; Sugisaki 2016).

The statistically significant differences between the ABS SUBJ and ERG SUBJ conditions (86.4% vs. 71.6%), and between the ERG SUBJ and ABS OBJ conditions (71.6% vs. 60.5%) found in our Tongan experiment indicate that both structural distance and case accessibility affect children's comprehension of *wh*-questions, and that the former has greater influence than the latter. The claim that structural distance and case accessibility are independent factors affecting the processing of filler-gap dependencies is consistent with the findings of Longenbaugh and Polinsky (2016), who report that speakers of Niuean – an Austronesian language closely related to Tongan – took longer to process ergative-subject and absolutive-object relative clauses compared to absolutive-subject relative clauses.

Finally, our observations that Japanese- and Tongan-speaking children had no problem in comprehending in-situ object *wh*-questions suggest that merely being an object (thus conceptually less accessible than a subject) is insufficient to explain children's poor performance on object *wh*-questions, and that it is displacement of a *wh*-phrase from its original thematic position that induces their difficulty in comprehending *wh*-questions. We take this as a piece of evidence against the Conceptual Accessibility Hypothesis, but there still remains a possibility that conceptual accessibility affects children's comprehension of *wh*-questions in a different way. More specifically, our crucial test items used in the experiments had the word order '*wh*-phrase . . . NP' due to overt *wh*-movement, and if children adopted a strategy like 'take the first NP in a sentence to be agent' (the agent-first strategy), then their inability to comprehend object *wh*-questions follows (an object *wh*-phrase undergoing movement to the sentence-initial position was interpreted wrongly as agent). Although the agent-first strategy is reported to be prevalent in children's comprehension of sentences involving non-canonical word order (passives, scrambling, clefts, etc.), it is not always the case that children rely on the agent-first strategy. For example, Aravind, Hackl, and Wexler (2018) report that English-speaking children aged 4 to 7 years (mean age, 5;11) comprehended object cleft sentences such as in (21b) (as well as the subject cleft sentences as in (21a)) very well (Subject cleft: 84% and Object cleft: 83%), when felicitous contexts were given.[15]

(21) a. Subject cleft
It's a dog that is chasing the cat.
b. Object cleft
It's a cat that the dog is chasing.

Furthermore, Shimada et al. (2020) report that Japanese-speaking children had no difficulty in comprehending subject right dislocation sentences as in (22), which have non-canonical patient-before-agent word order.

**15** However, as indicated by Ohba, Sano, and Yamakoshi (2019), there remains a possibility that the children in Aravind, Hackl, and Wexler's (2018) experiment could give a correct answer without hearing a whole cleft sentence because there were only two animal characters involved in the context. In fact, Ohba et al. (2019) report that the agent-first strategy was observed in children's comprehension of the Japanese cleft construction if three animal characters were given in the context.

(22) Subject right dislocation

| | | | |
|---|---|---|---|
| Koarasan-o | oikake-teiru | yo, | oumasan-ga. |
| koala-ACC | chase-PROG | SFP | horse-NOM |

'The horse is chasing the koala.'

The child participants in Shimada et al.'s (2020) study correctly comprehended sentences with subject right dislocation 86.1% of the time (and they also correctly comprehended sentences with object right dislocation 100% of the time). Therefore, to disentangle the effects of the agent-first strategy from the effects of structural distance, we need to identify the factors that make it possible for children to not resort to the agent first strategy (felicity conditions, types of movement, etc.), and determine whether the effects of structural distance remain after such factors are controlled. We leave this problem for future research.

To conclude, our cross-linguistic investigations pertaining to children's comprehension of *wh*-questions made it possible to tease apart the potential sources of their subject preference and narrow down the factors that trigger their difficulty in comprehending *wh*-questions. Languages such as Tongan and Kaqchikel have important characteristics that cannot be found in English and other typologically similar languages (exhibiting nominative/accusative case alignment and S-before-O word order). Including these languages in our investigations, this study demonstrated that testing children's comprehension in typologically distant languages has the potential of making a unique contribution to language acquisition research.

**Acknowledgments:** First, we would like to thank the participants of our experiments in Japan, Tonga, and Guatemala, and their kindergarten teachers who made their participation possible. We are also grateful to Soana Kaitapu, 'Ana Heti Veikune, Peseti Vea, and Raelyn Esau, without whose assistance we could not have conducted the experiment in Tonga. As for the Kaqchikel experiment, we are indebted to Juan Esteban Ajsivinac Sián, Lolmay Pedro Oscar García Mátzar, Filiberto Patal Majzul, and Yoshiho Yasugi for helping us create and analyze relevant Kaqchikel sentences. Lastly, we would like to thank two anonymous reviewers and the audience at Crosslinguistic Perspectives on Processing and Learning (Zürich, Switzerland), Boston University Conference on Language Development 44 (Boston, USA), and International Symposium on Issues in Japanese Psycholinguistics from Comparative Perspectives (Online) for their invaluable comments and suggestions. All remaining errors are our own. Part of this research was supported by JSPS KAKENHI Grant Numbers #15H02603 (PI: Masatoshi Koizumi), #19H05589 (PI: Masatoshi Koizumi), #19K21634 (PI: Yuko Otsuka).

# References

Aldridge, Edith. 2008. Generative approaches to syntactic ergativity. *Language and Linguistics Compass* 2(5). 966–995.
Aravind, Athulya, Martin Hackel, & Kenneth Wexler. 2018. Syntactic and pragmatic factors in children's comprehension of cleft constructions. *Language Acquisition* 25(3). 284–314.
Baayen, Harald, Douglas J. Davidson, & Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4). 390–412.
Bates, Douglas M., Martin Mächler, Ben Bolker, & Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1). 1–48.
Bock, J. Kathryn & Richard K. Warren. 1985. Conceptual accessibility and syntactic structure in sentence formulation. *Cognition* 21(1). 47–67.
Broadwell, George A. & Timothy J. Smith. 2001. Form and function in Kaqchikel word order. *Proceedings of the 37th Meeting of the Chicago Linguistic Society* [Volume 2]. 1–15.
Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
Chomsky, Noam. 1993. A minimalist program for linguistic theory. In Ken Hale & Samuel Jay Keyser (eds.), *The View from Building* 20*. Essays in Linguistics in Honor of Sylvain Bromberger*, 1–52. Cambridge, MA: MIT Press.
Crain, Stephen & Mineharu Nakayama. 1987. Structure dependence in grammar formation. *Language* 63(3). 522–543.
Custis, Tonya. 2004. *Word order variation in Tongan: A syntactic analysis.* Minneapolis, MN: University of Minnesota dissertation.
Dryer, Matthew S. 2005. Order of subject, object, and verb. In Martin Haspelmath, Matthew S. Dryer, David Gil, & Bernard Comrie (eds.), *The World Atlas of Language Structures*, 330–333. Oxford: Oxford University Press.
Friedmann, Naama, Adriana Belletti, & Luigi Rizzi. 2009. Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua* 119(1). 67–88.
García Matzar, Lolmay Pedro & Pakal B'alam José Obispo Rodrígues Guaján. 1997. *Rukemik ri Kaqchikel chi': Gramática kaqchikel*. Guatemala City: Cholsamaj.
Gualmini, Andrea & Stephen Crain. 2005. The structure of children's linguistic knowledge. *Linguistic Inquiry* 36(3). 463–474.
Gutierrez-Mangado, Juncal. 2011. Children's comprehension of relative clauses in an ergative language: The case of Basque. *Language Acquisition* 18(3). 176–201.
Hawkins, John A. 1999. Processing complexity and filler-gap dependencies across grammars. *Language* 75(2). 244–285.
Hawkins, John A. 2004. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.
Henderson, Robert. 2012. A tour of Kaqchikel ergativity. Handout, McGill Ergativity Lab.
Huang, C.-T. James. 1982. Logical relations in Chinese and the theory of grammar. Cambride, MA: MIT doctoral dissertation.
Keenan, Edward & Bernard Comrie. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8(1). 63–99.
Keenan, Edward & Bernard Comrie. 1979. Data on the noun phrase accessibility hierarchy. *Language* 55(2). 333–351.
Koizumi, Masatoshi, Yoshiho Yasugi, Katsuo Tamaoka, Sachiko Kiyama, Jungho Kim, Juan Esteban Ajsivinac Sian, & Lolmay Pedro Oscar García Matzar. 2014. On the (non-)universality of the

preference for subject-object word order in sentence comprehension: A sentence processing study in Kaqchikel Maya. *Language* 90(3). 722–736.

Lau, Elaine & Nozomi Tanaka. 2021. The subject advantage in relative clauses: A review. *Glossa: A Journal of General Linguistics* 6(1). 1–34.

Lasnik, Howard & Mamoru Saito. 1984. On the nature of Proper Government. *Linguistic Inquiry* 15(2). 235–289.

Liu, Minqi, Nina Hyams, & Victoria Mateu. 2020. Late intervention effects in the acquisition of Mandarin sluice-like constructions. *Proceedings of Boston University Conference on Language Development* 44. 322–335.

Longenbaugh, Nicholas & Maria Polinsky. 2016. The processing of long-distance dependencies in Niuean. *Proceedings of Austronesian Formal Linguistics Association* 22. 98–120.

Marantz, Alec. 1991. Case and licensing. *Proceedings of Eastern States Conference on Linguistics '91*. 234–253.

van der Meer, Matthijs, Wouter van Atteveldt, Peter Coopmans, & William Philip. 2001. Subject-object asymmetry in Dutch children's comprehension of wie-questions. *Linguistics in the Netherlands* 2001. 167–176.

Muāgututi‘a, Grant. 2017. The acquisition of ergativity in Samoan. *University of Hawai‘i at Mānoa: Working Papers in Linguistics* 48. 1–14.

Murasugi, Keiko. 1991. Noun phrases in Japanese and English: A study in syntax, learnability and acquisition. Storrs, CT: University of Connecticut doctoral dissertation.

Nishigauchi, Taisuke. 1990. *Quantification in the Theory of Grammar*. Dordrechit: Kluwer.

O'Grady, William. 1997. *Syntactic Development*. Chicago, IL: University of Chicago Press.

Ohba, Akari, Tetsuya Sano, & Kyoko Yamakoshi. 2019. Children's acquisition of clefts revisited: New evidence from Japanese. *Proceedings of Boston University Conference on Language Development* 43. 483–495.

Otaki, Koichi, Manami Sato, Hajime Ono, Koji Sugisaki, Noriaki Yusa, Soana Kaitapu, 'Ana Heti Veikune, Peseti Vea, Yuko Otsuka, & Masatoshi Koizumi. 2020. The ergative subject preference in the acquisition of wh-questions in Tongan. *Proceedings of Boston University Conference on Language Development* 44. 465–478.

Otaki, Koichi, Koji Sugisaki, Noriaki Yusa, & Masatoshi Koizumi. 2019. Two routes to the Mayan VOS: From the view of Kaqchikel. *Gengo Kenkyu* 156. 25–45.

Otsuka, Yuko. 2000. Ergativity in Tongan. Oxford: University of Oxford doctoral dissertation.

Otsuka, Yuko. 2005. Scrambling and information focus: VSO-VOS alternation in Tongan. In Joachim Sabel & Mamoru Saito (eds.), *The Free Word Order Phenomenon: Its Syntactic Sources and Diversity*, 243–279. Berlin: Mouton de Gruyter.

Otsuka, Yuko. 2006. Syntactic ergativity in Tongan: Resumptive pronouns revisited. In Alana Johns, Diane Massam, & Juvenal Ndayiragije (eds.), *Ergativity: Emerging Issues*, 79–107. Berlin: Springer.

Roesch, Anne-Dorothée & Vasiliki Chondrogianni. 2014. Morphological cues in the comprehension of wh-questions in German-speaking simultaneous and early sequential bilingual children. *Proceedings of Boston University Conference on Language Development* 38 [Volume 2]. 380–392.

Saito, Mamoru. 1985. Some asymmetries in Japanese and their theoretical implications. Cambridge, MA: MIT doctoral dissertation.

Shimada, Hiroyuki, Yuko Masaki, Rika Okada, Akari Ohba, Kanako Ikeda, & Kyoko Yamakoshi. 2020. The agent-first strategy and word order: Children's comprehension of right dislocations and clefts in Japanese. *Proceedings of Boston University Conference on Language Development* 44. 586–595.

Sugisaki, Koji. 2016. Quantifier float and structure dependence in child Japanese. *Language Acquisition* 23(1). 75–88.
Suzuki, Takaaki. 2011. A case-marking cue for filler–gap dependencies in children's relative clauses in Japanese. *Journal of Child Language* 38(5). 1084–1095.
Tanaka, Mikihiro, Holly P. Branigan, Janet F. McLean, & Martin J. Pickering. 2011. Conceptual influences on word order and voice in sentence production: Evidence from Japanese. *Journal of Memory and Language* 65(3). 318–330.
Tyack, Dorothy & David Ingram. 1977. Children's production and comprehension of questions. *Journal of Child Language* 4(2). 211–224.
Xiang, Ming, Brian Dillon, Matt Wagers, Fengqin Liu, & Taomei Guo. 2014. Processing covert dependencies: An SAT study on Mandarin wh-in-situ questions. *Journal of East Asian Linguistics* 23(2). 207–232.
Yasunaga, Daichi, Masataka Yano, Yoshiho Yasugi, & Masatoshi Koizumi. 2015. Is the subject-before-object preference universal? An ERP study in the Kaqchikel Mayan language. *Language, Cognition and Neuroscience* 30(9). 1209–1229.
Yoshinaga, Naoko. 1996. WH-questions: A comparative study of their form and acquisition in English and Japanese. Manoa, HI: University of Hawai'i at Manoa doctoral dissertation.
Zuckerman, Shalom, Manuela Pinto, Elly Koutamanis, & Yoïn van Spijk. 2016. A new method for testing language comprehension reveals better performance on passive and principle B constructions. *Proceedings of Boston University Conference on Language Development* 40. 443–456.

Kazuko Yatsushiro

# Chapter 10 Crosslinguistic investigation of the acquisition of disjunction

## 1 Introduction

Logically, disjunction of the form *X or Y* is true when *X* is true but not *Y*, *Y* is true but not *X*, and crucially, both of *X* and *Y* are true, as shown in the truth table in Table 1. Let us call this *inclusive* interpretation of disjunction.

**Table 1:** The truth table for disjunction.

| X | Y | X or Y | X and Y |
|---|---|---|---|
| 0 | 0 | false | false |
| 1 | 0 | true | false |
| 0 | 1 | true | false |
| 1 | 1 | true | true |

When an adult speaker uses a disjunction, however, their interpretation depends on the sentential environment. When a disjunction appears in an upward entailing environment, such as (1), adult speakers accept the use of disjunction when (1a) or (1b) are true, but may not when (1c) is true. Let us call this *exclusive* interpretation. It specifically excludes the possibility of both disjuncts being true.

(1) Mika ate peaches or ice cream.
  a. Mika ate peaches but she did not eat ice cream.
  b. Mika ate ice cream, but she did not eat peaches.
  c. Mika ate peaches and she ate ice cream.

When the disjunction appears in downward entailing environments, such as under negation as in (2), on the other hand, adult speakers allow for the inclusive inter-

**Acknowledgments:** I would like to thank audiences at the International Symposium on Issues in Japanese Psycholinguistics from Comparative Perspectives and Uli Sauerland for discussion on earlier version of this chapter, and two anonymous reviewers for their very helpful comments. This work is supported by ERC (agreement No. 787929).


https://doi.org/10.1515/9783110778946-010

pretation of disjunction: They accept the use of disjunction when both disjuncts are true, as in (2c), in addition to environments such as (2a) and (2b).

(2) Mika didn't eat peaches or ice cream.
  a. Mika didn't eat peaches, but she ate ice cream.
  b. Mika didn't eat ice cream, but she ate peaches.
  c. Mika didn't eat peaches and she didn't eat ice cream.

There are currently two main approaches to explain how the exclusive interpretation arises. One is a Gricean reasoning (Grice 1975: Chapter 3), and the other approach is a grammatical approach (Chierchia et al. 2001; Fox 2007; Chierchia, Fox, and Spector 2012; and others). According to the Gricean reasoning view, scalar implicature (the exclusive interpretation in the present context) is derived by following pragmatic reasonings. Let us see how.[1]

Pragmatic maxim of Quantity says that a speaker should make their contribution as informative as is required. Let us assume that *or* and *and* form a scale (Horn 1972) and ordered by informativity. As shown in the truth table in Table 1, a sentence [x and y] with conjunction is true in a subset of environments where a sentence [x or y] with disjunction is true. According to the Gricean quantity maxim, a speaker would use the expression that is maximally informative. Sentences with both the disjunction and the conjunction are compatible with the situation in (1c). Because the sentence with a conjunction is not true in any other environment, however, conjunction should be preferred over disjunction whenever both x and y are true, because it is more informative than using the counterpart with the disjunction.

Since Chierchia et al. (2001), the acquisition of disjunction by children have received much attention. Chierchia et al. and subsequent works have shown that, in downward entailing environment, children interpret disjunction inclusively, similarly to adult speakers (Su and Crain 2013, Tieu et al. 2016). This has been shown to be true in languages like Japanese as well. In Japanese, adult speakers interpret disjunction only *"exclusively"*: adult Japanese speakers have been observed to accept the sentence (2) when (2a) and (2b) are true, but they reject them when (2c) is true (Goro 2007). Child Japanese speakers, on the other hand, accept (2c) (Goro and Akiba 2004), resembling the responses by English speaking adults and children more closely than adult Japanese speakers.

1 We will come back to the grammatical approach in the next section.

In the upward entailing environment, children have been found to interpret disjunction sometimes *conjunctively* (Singh et al. 2016; Tieu et al. 2017): they accept the sentence in the environment like (1c) but reject when (1a) or (1b) is true. In this chapter, we compare acquisition of disjunction in two languages, Japanese and German, focusing on the disjunction in upward entailing environment.

# 2 Semantic theory to acquisition

Why do children interpret disjunction conjunctively? According to Singh et al. (2016) and Tieu et al. (2017), the difference between children's and adults' interpretation is that children do not have the conjunction as the alternative to disjunction.

## 2.1 Grammatical approach to scalar implicature

The conjunctive interpretation is unavailable in upward entailing environment for adult speakers in English, except in special environment such as free-choice disjunction (Meyer 2015).[2]

Let us see how the exclusive interpretation arises first with the grammatical approach. Let us assume that a set of alternatives of a sentence with a disjunction *A or B* are {A or B, A and not B, B and not A, A and B} (Sauerland 2004 and others). The example in (1) are repeated below with a slight modification for illustration.

(3) a. Mika ate peaches or Mika ate ice cream.
 b. Mika ate peaches and she didn't eat ice cream.
 c. Mika ate ice cream and she didn't eat peaches.
 d. Mika ate peaches and she ate ice cream.

Let us further assume that there is a silent operator (*exh*), which is skin to *only*. For concreteness, let us follow further the analysis in Fox (2007). Informally, the meaning of *only* is that the sentence that contains it presupposes that only sentences that are true are the ones that the sentence with that *only* combines with (the prejacent) entails. Fox proposes that *exh* is similar, but asserts (rather than presupposes) that

**2** In other languages, *manu* in Walpiri (Bowler 2014) and *ya* in Japanese (Sudo 2014) have been observed to exhibit conjunctive interpretation in upward entailing environment, although their interpretation in downward entailing environment suggest that they are disjunctions.

only sentences that are true are those entailed by the prejacent. As a result, when *exh* applies to a sentence, it has an effect of negating all the alternatives that are not entailed by the sentence. All the alternatives of (3a) ((3b), (3c), (3d)) are not entailed by (3a).

(4) a. It's not the case that Mika ate peaches and she didn't eat ice cream
-> Mika didn't eat peaches but she ate ice cream.
b. It's not the case that Mika didn't peaches but she ate ice cream
-> Mika ate peaches but she didn't eat ice cream.
c. It's not the case that Mika ate both peaches and ice cream.

Because (4a) and (4b) together lead to a contradiction, they are excluded from part of the meaning, leading us to the exclusive *or* interpretation: *'Mika ate peaches or ice cream',* and *Mika didn't eat both peaches and ice cream'* (Chierchia et al. 2012; Fox 2007).

Let us now ask how to derive the conjunctive interpretation. As mentioned above, the conjunctive interpretation is observed in adult speakers's free-choice interpretation. Fox argues that the conjunctive interpretation of free choice arises by the mulitple application of *exh*. According to Fox (2007), the operator *exh* applies when the application of it results in a meaning change. First application of *exh* results in the exclusive interpretation. If we apply *exh* another time, the meaning ends up being conjunctive, because after the first application of *exh*, the conjunction is not included as one of the alternatives, and hence, is not negated.

## 2.2 Previous acquisition studies

One of the earliest studies on disjunction is Chierchia et al. (2001). Chierchia et al. investigated children's interpretation of disjunction in English in downward and upward entailing environments. In the downward entailing environment, implicatures are cancelled, and as a result, adult speakers interpret disjunction inclusively. Chierchia et al. found that, in the downward entailing environments, such as within the restrictor of a universal quantifier ((5)), 3- to 6-year old children interpreted disjunction inclusively, accepting its use when both disjuncts were true.

(5) Every dwarf who chose a banana or a strawberry received a jewel.

In upward entailing environments, such as within the scope of universal quantifier ((6)), on the other hand, adult speakers rejected the use of disjunction when both disjuncts were true (exclusive interpretation), whereas children accepted its use 50% of the time.

(6) Every boy chose a skateboard or a bike.

More recently, Singh et al. (2016) investigated children's interpretation of disjunction in upward entailing environments. Singh et al. observes that 3- to 6-year-old English speaking children interpreted disjunction as if it is conjunction, rejecting the use of disjunction when only one of the disjuncts is true, while accepting its use when both disjuncts are true.

(7) Every boy is holding an apple or a banana. (Singh et al. 2016)
   a. α: Every boy is holding an apple and a banana.
   b. *β : Every boy is holding an apple or a banana, and not both.

Similarly, Tieu et al. (2017) observe that French and Japanese children seem to interpret disjunction as if it is conjunction.

According to Singh et al. (2016) and Tieu et al. (2017), the difference between children's and adults' interpretation is that children do not have the conjunction as the alternative to disjunction. Let us see what happens when conjunction is not an alternative. If the set of alternatives are {A or B, A and not B, B and not A} for children, the application of *exh* operator creates a set with the following members: {A or B, not A and B, not B and A}. Crucially, the meaning does not include *not (A and B)*, and through strengthening, the resulting interpretation is that of conjunction.

What Tieu et al.'s study differs from the previous studies mentioned so far is that it tested children's interpretation of complex disjunctions (e.g., *ka . . . ka* 'or. . . or' in Japanese) which must be interpreted exclusively (Nicolae and Sauerland 2016; Spector 2014) in addition to simple disjunctions (e.g., *ka* 'or' in Japanese) which does not require exclusive interpretation. Spector proposes that the difference between simple and complex disjunction is that, for the latter, the application of *exh* operator is obligatory.

This theory makes a prediction: if it is the case that children arrive at the conjunctive interpretation because they apply exhaustification without a conjunction as an alternative, we may observe children to reach conjunctive interpretation more frequently in the environment where the application of exhaustification is obligatory. That is, the prediction is that children should interpret disjunction conjunctively more frequently with complex form of disjunction than with their simple counterpart.

Tieu et al. (2017) conducted the experiment in French (complex *soit. . . soit* vs. simple *ou*) and in Japanese (complex *ka. . . ka* vs. simple *ka*), testing 28 French-speaking children (3;07–6;06, M=4;05) and 18 Japanese-speaking children (4;07–6;06, M=5;05). They replicated the earlier finding in English by Singh et al. (2016) in both French and Japanese and found that young children accepted the use of disjunction in both the context where one but not both disjuncts are true (1DT) and the context where both conjuncts are true (2DT). Furthermore, there were children who accepted the use of disjunction in 2DT contexts while rejecting its use in 1DT contexts. Contrary to the prediction, however, there was no difference between the simple and the complex disjunctions.

As discussed above, Singh et al. (2016) and Tieu et al. (2017) argue that the difference between children's and adults' interpretation is that children do not have the conjunction as the alternative to disjunction. It is puzzling, however, why we do not observe any difference between simple and complex disjunction. This is one of the reasons why we tested two types of disjunctions in German, *oder* and *entweder. . . oder*, which have different morpho/syntactic characteristics from French *soit. . .soit/ou* and Japanese *ka. . .ka/ka*. German complex disjunction is marked by *entweder* 'either' before the first disjunct. *Entweder* uniquely marks complex disjunction in German, without occurring in any other environment. Japanese *ka*, on the other hand, is at least homophonous, if not the same lexical item, with the question particle *ka*, as well as the existential quantificational particle *ka*.

Before moving to the next section, let us discuss Skordos et al. (2020). Skordos et al. extended the study by Tieu et al. (2017) by constructing three conditions: (i) replication of Tieu et al.'s study, (ii) modification of Tieu et al. regarding the lead-in sentences (modified script version), (iii) modification of Tieu et al. by adding a third object that is not mentioned (three alternatives version). In the original study by Tieu et al. (2017), the experimenter describes what really happened before they ask whether what the puppet predicted was right. Skordos et al. shortened the lead-in sentence by taking out the experimenter's description. The other modification, the addition of the third object, was introduced to the context in order to make the context more felicitous for the puppet's prediction. Their results show that overall, there was no difference among conditions for 2DT contexts (mixed effects logistic regression: $\chi^2$=3.721, *p*=.156), but there was an effect of condition for 1DT contexts–children who were tested with the three alternatives condition were more likely to accept 1DT items than children who were tested with Tieu et al.'s original design, while children who were tested with the modified script version did not differ from either of the other two groups.

In the present study, we use the design by Tieu et al., in order to compare the Japanese data and German data.

# 3 Experiment: Truth-value judgment task

## 3.1 Predictions

We make the following predictions:

(8) a. If 4- to 5-year-old children across languages do not have conjunction as an alternative to disjunction, we expect not to see a difference between Japanese and German children. We should observe, then, that German children also show conjunctive interpretation.
b. As Tieu et al. (2017) did not observe difference between simple and complex disjunctions in French and Japanese, we should not see an effect of complexity of disjunction in German, either.
c. Some other factor (such as uniqueness of the disjunction morpheme) influences the availability of conjunction as an alternative, however, we may observe language variation.

We tested German children's interpretation of two types of disjunctions to examine whether these predictions are supported. Furthermore, we tested larger age-range of children to see if we observe a developmental pattern.

## 3.2 Method and procedure

We adapted the design of Tieu et al. (2017) to German. This allows us to compare the results from the two studies. Furthermore, we followed the protocol for conducting the experiment that Tieu et al. developed. The experiment used the truth-value judgment task in the prediction mode (Chierchia et al. 1998). Each experimental set-up was introduced as a story, which was followed by a prediction by the puppet stating what would happen next. During the experiment, the participant and the experimenter sat next to each other with a tablet computer in front of them. The stories were presented on the tablet, using a presentation software. All the audio stimuli were pre-recorded by a native speaker, and played from the tablet. Each story consisted of three slides. On the first slide, the story was told. In the story, there were always two possible theme argument candidates. On the second slide, a puppet appeared on the monitor and predicted what might have happened next. The prediction contained a disjunction for the target and the control items. The third slide was the end of story, depicting three types of ending: (i) one disjunct was true (1DT), (ii) two disjuncts were true (2DT), or (iii) neither disjunct was true (0DT). After puppet made the prediction, the experimenter described what happened on the third slide, in order to make sure that the participant understood what happened. Participants were then asked whether the puppet made the right prediction. They put a stamp on the answer sheet, depending on whether they judged the prediction that the puppet made matched what actually happened (under a smiley face when matched) or not (under a sad face). When the story and the prediction did not match, the participants were asked to state what was wrong with the puppet's prediction. The whole experiment was audio-recorded, and later checked for the responses children made.

## 3.3 Participants

Total of 89 monolingual German speaking children (3;11–8;8, *M*=6;2) and 21 adults participated in this study. There were two lists per type of disjunction (simple and complex), in order to verify the order of presentation affecting the interpretation. Participants were randomly assigned to one of the four lists.

The child participants were recruited at a day care center and two public schools in Berlin, Germany. The adult speakers were recruited from the participant pool of Humboldt University, Berlin. Child participants received a sticker for their participation. Adult participants received 5 euro to compensate for the time they spent for taking part in this study.

## 3.4 Material

There were two experimental conditions: only one of the disjuncts is true (1DT) and both of the disjuncts are true (2DT). Let us illustrate this using an example from the experiment. Sentences in (9) and (10) are examples, using two types of disjunctions.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| (9) | Das | Huhn | hat | das | Flugzeug | oder | den | Bus | geschubst. |
| | the | chicken | has | the | plane | or | the | bus | pushed |

‘The chicken has pushed the plan or the bus.’

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| (10) | Das | Huhn | hat | entweder | das | Flugzeug | oder | den | Bus | geschubst. |
| | the | chicken | has | either | the | plane | or | the | bus | pushed. |

‘The chicken has pushed either the plane or the bus.’

For sentence (11), the first scene shows a chicken with two toys: an airplane and a bus. The two scenarios that describe what happened in 1DT and 2DT contexts are described in (12a) and (12b).

(11) Das Huhn hat das Flugzeug oder den Bus geschubst.
the chicken has the plane or the bus pushed
'The chicken pushed the plane or the bus.'

(12) a. 1DT: The chicken pushed only the plane, not the bus
b. 2DT: The chicken pushed both the plane and the bus

We used simple (*oder*) and complex (*entweder. . . oder*) disjunctions in German, and constructed two versions of experiments, using only one type of disjunction within each version. Participants were randomly assigned to the list.

The experiment contained four items each of 1DT and 2DT conditions, 2 control items (0DT), and 3 fillers. Filler items had two potential endings, with the third scene either true or false. The experimenter showed the false-ending when the participant mostly accepted the puppet's prediction as correct, or vice versa, in order to balance the number of times puppet predicted correctly/incorrectly. The orders of the items were identical to the order used by Tieu et al.

## 3.5 Results

In the analysis below, we include participants who responded correctly to the filler items at least two out of three trials. Adult speakers correctly responded to the filler items 100% of the time. There were 11 children that responded correctly to filler items only once, and will be excluded from the analysis.

In addition, participants were excluded if they did not reject the puppet's utterance in the 0DT condition. One adult and 8 child participants were excluded for this reason. Below, data from 20 adult participants and 70 child participants are analyzed. The number of participants for each list is shown in Table 2.

**Table 2:** The number of participants per list.

| | Adults | Children |
|---|---|---|
| Simple | 11 | 32 |
| Complex | 9 | 38 |

Throughout this chapter, we analyze the data from the experiment by fitting a generalized linear mixed-effects models (glmer) using the lme4 package in R (Bates et al. 2015). In what follows, the dependent variable is always the response type (*agree* or *disagree*). The fixed effect we use for each analysis will be specified each time.

Let us first consider the responses for the 1DT contexts. Adult speakers responded that the puppet correctly predicted the outcome of the story 100% of the time (36 out of 36 trials) when the puppet used the complex disjunction (*entweder. . . oder*), and 90.9% of the time (40 out of 44 trials) when the simple disjunction (*oder*) was used. Children, on the other hand, said the puppet correctly predicted the outcome of the story 82% of the time (105 out of 128 trials) with the simple disjunction, and 75.5% of the time (114 out of 151 trials) with the complex disjunction.

To check the effect of age group (child vs adult) and the effect of complexity of disjunctions (simple vs. complex) on how participants responded to the items in the 1DT condition, we constructed a mixed-effects logit model with the response type (agree vs. disagree) as the dependent variable, and the age group (child vs. adult) and complexity (simple *oder* vs. complex *entweder. . . oder*) as the predictors, and participant as a random effect (using the lme4 package in R, Bates et al. 2015). The model with the two predictors show that the age group is a significant effect (z-value=–2.439, $p$<0.05), whereas the complexity is not (z-value=–0.124, p=0.9012). A comparison between a model with the age group and without the age group show that the model with the age group is significantly different from the one without ($\chi^2$=7.1605, $p$<.01).

Next let us divide child participants into two groups: 4–5-year-olds, similar age range as in Tieu et al.'s study, and 6-year-olds and older. This divides child participants into two groups with similar number of participants, as shown on Table 3.

**Table 3:** Number of participants, divided into three groups.

| | **Adult** | **4–5-year-olds** | **6 & older** |
|---|---|---|---|
| Simple | 11 | 13 | 19 |
| Complex | 9 | 19 | 19 |

We observe that 4- and 5-year-old children accepted the puppet's use of disjunction in 1DT condition 63.5% of the time (33 out of 52 trials) with simple disjunction, and 68% of the time (51 out of 75 trials) with a complex disjunction. The 6-year-olds and older, on the other hand, accepted the use of simple disjunction 94.7% of the time (72 out of 76 trials) and 82.9% of the time (63 out of 76 trials) with complex disjunction, as summarized in Table 4.

We fitted a mixed-effects logit model with the age group (4–5-year-olds vs. 6–8-year-olds vs. adults) as a predictor. The analysis shows that, while there is no difference between adults and older children (z-value=–1.562, p=.11824), there was a significant difference in how participants responded between 4–5-year-olds

**Table 4:** Proportion of "agree" response for the 1DT condition.

| | Adults | Children (overall) | 4–5-year-olds | 6 & older |
|---|---|---|---|---|
| Simple | 100 | 82.0 | 63.5 | 94.7 |
| Complex | 90.9 | 75.5 | 68.0 | 82.9 |

and adults (z-value=–2.944 $p<.00324$). A comparison between a model with the age group and without the age group show that the model with the age group is significantly different from the one without ($\chi^2=12.554$, $p<.01$).

Let us now turn to the 2DT condition. Recall that the expected response by adult speakers, if they arrive at exclusive interpretation, is that they reject the use of disjunction in 2DT condition. This is, in fact, what we obtained. Adult participants rejected the use of disjunction 90.9% of the time (40 out of 44 trials) with simple disjunction, and 100% of the time (36 out of 36 trials) with complex disjunction. Child participants, on the other hand, rejected the use of disjunction only 40.9% of the time (52 out of 127 trials) with the simple disjunction, and 57.9% of the time (88 out of 162 trials) with the complex disjunction.

We fitted a linear mixed model with the complexity of disjunction and age group (child vs. adults) as predictors, and observe that, as is the case with the 1DT, the age group is a significant factor (z-value=4.453, $p<.01$), while complexity is not (z-value=−1.797, p=.0723). A comparison between a model with the age group and without the age group show that the model with the age group is significantly different from the one without ($\chi^2=33.228$, $p<.01$).

We, again, divided the child participants into two groups based on age, and check the average rate of rejection of the use of disjunction for both simple and complex disjunctions. Overall summary of rejection rates for each age group is shown in Table 5.

**Table 5:** Proportion of "disagree" response for the 2DT condition.

| | Adults | Children (overall) | 4–5-year-olds | 6 & older |
|---|---|---|---|---|
| Simple | 90.9 | 40.9 | 19.6 | 55.3 |
| Complex | 100 | 57.9 | 57.9 | 57.9 |

A model with the age group (4–5-year-olds vs. 6–8-year-olds vs. adults) as a predictor shows that adults differed significantly from both 4–5-year-olds (z-value=4.526, $p<.01$) and 6-year-olds and older (z-value=4.024, $p<.01$).

# 4 Comparing Japanese and German

Let us now compare our results with the Japanese data from Tieu et al. We first examine how consistently participants from different groups responded for 1DT and 2DT items. In order to do so, we calculated how many times each participant accepted or rejected the puppet's prediction as being correct in 1DT and 2DT conditions separately. Figures 1 and 2 show how participants responded, based on how many times they accepted the use of disjunction in 1DT and in 2DT conditions. In these graphs, the proportion of children who accepted the use of disjunction 4 out of 4 trials are indicated as black, and the proportion of children who rejected the use of disjunction 4 out of 4 trials are indicated as lightest grey.

On Figure 1, the results from simple disjunction are shown. Comparing these three graphs, we observe that large proportion of German children who are 6 and older accepted the use of disjunction 4 out of 4 trails of the 1DT condition, whereas

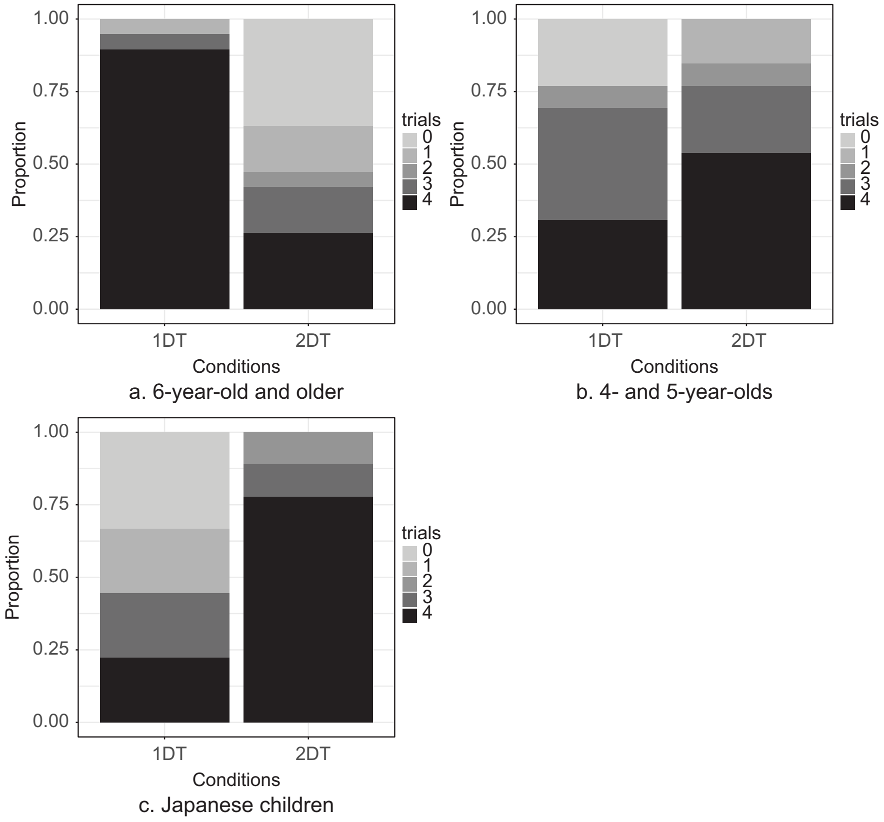



**Figure 1:** Number of agree responses per participant (simple).

German 4- and 5-year-olds and Japanese speaking 4- and 5-year-old children did not do so as frequently. As for the 2DT condition, large proportion of Japanese 4- and 5-year-old children accepted the use of disjunction, German 4- and 5-year-olds only around 50% of the group, and 6-year-olds and older German children did around 25%.

Figure 2 shows what happened when the disjunction was complex. When the disjunction was complex, the proportion of children who accepted the use of disjunction in 2DT conditions seem to decrease, compared to when the disjunction was simplex.

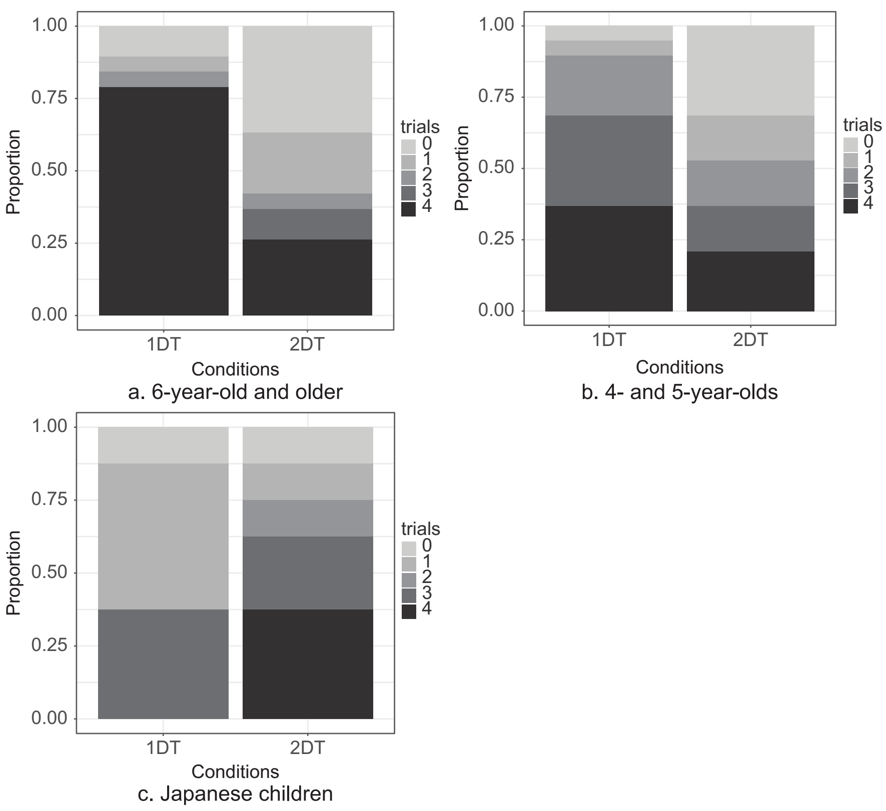



**Figure 2:** Number of agree responses per participant (complex).

Let us next combine the results from both conditions. Following figures show the distribution of participants, based on how many times they accepted the use of disjunction in 1DT context (X-axis) and that of 2DT context (Y-axis). Recall that inclusive speakers accept disjunction in both 1DT and 2DT contexts (top right corner of the chart), whereas exclusive speakers accept disjunction in 1DT but not in 2DT contexts (bottom right corner). We expect that most adult speakers should be an exclusive speaker, occupying the bottom right corner. According to Tieu et al.'s study, children interpret disjunction inclusively or conjunctively, and hence, should be in the upper half of the graph.

As can be seen in Figures 3 and 4, adult participants of both languages are mostly concentrated in the lower right corner, as expected, with either simple or complex disjunction was used.

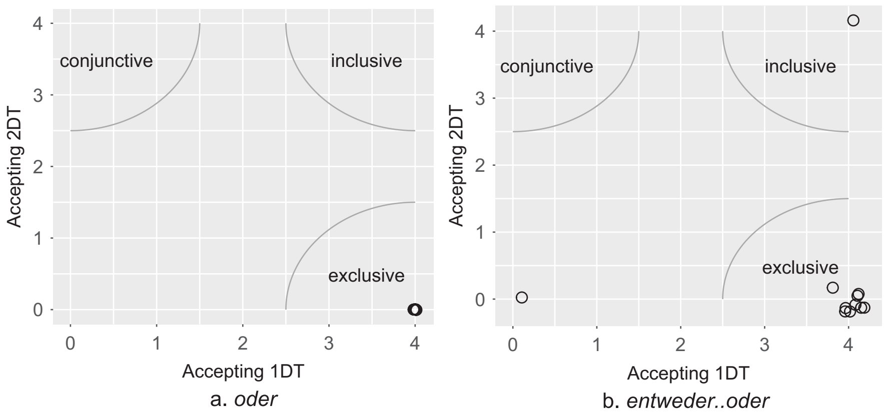



**Figure 3:** Distribution of German adult participants.

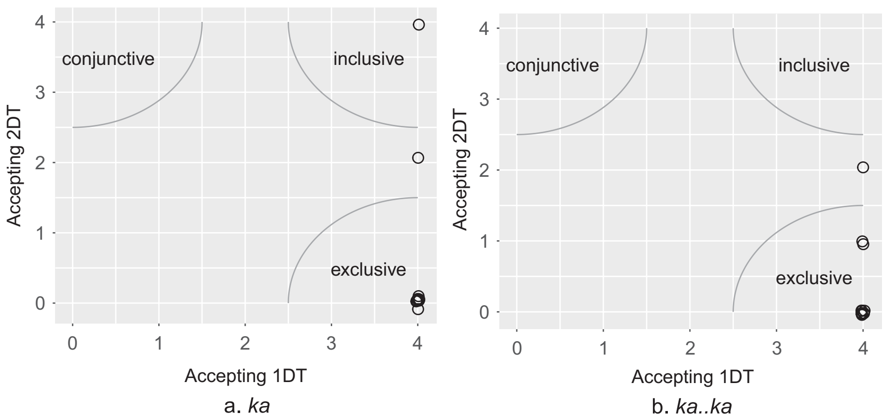



**Figure 4:** Distribution of Japanese adult participants.

Let us now see the distribution of children. As can be seen in Figures 5 and 6, children are distributed over three different areas. With the simple disjunction, there are only 2 children who are in the conjunction interpretation area among German 4- and 5-year-old children. Instead it seems that the 4–5 year old children most frequently access the inclusive interpretation for *oder*. This contrasts with the distribution for Japanese children.[3] There were no children who interpreted disjunction exclusively. German children who are 6 years old and older, on the other hand, are more concentrated in the exclusive interpretation area.

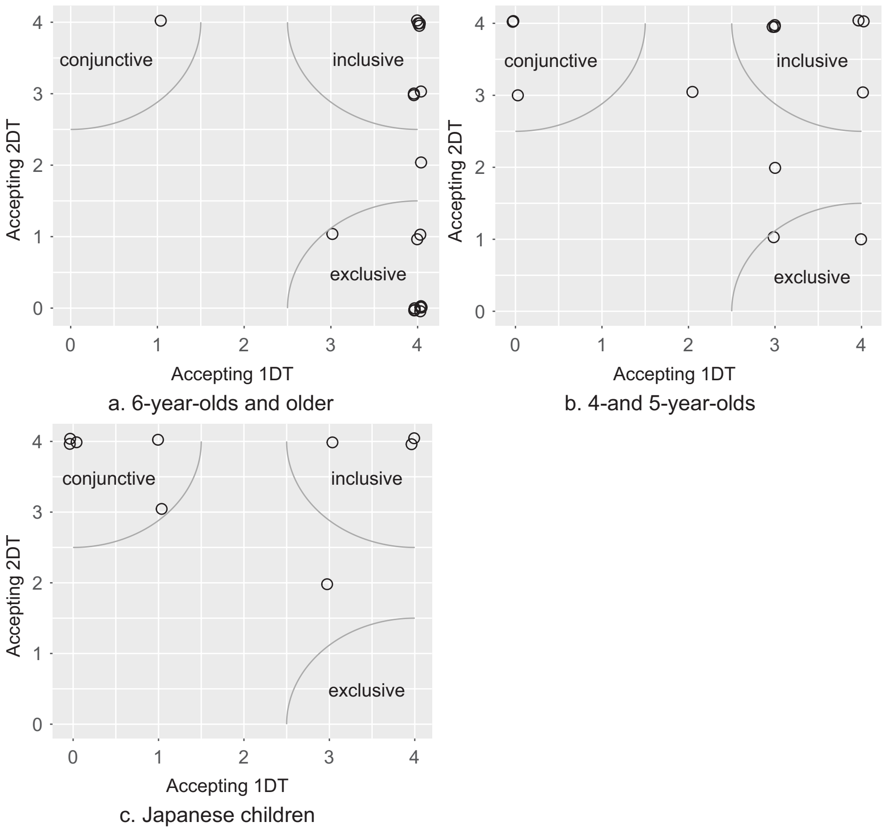



**Figure 5:** Distribution of children (simple disjunctions).

3 The graphs are created from the dataset of Tieu et al., available at https://semanticsarchive.net/Archive/mE4YmYwN/TYCRSC-AcqDisj.html

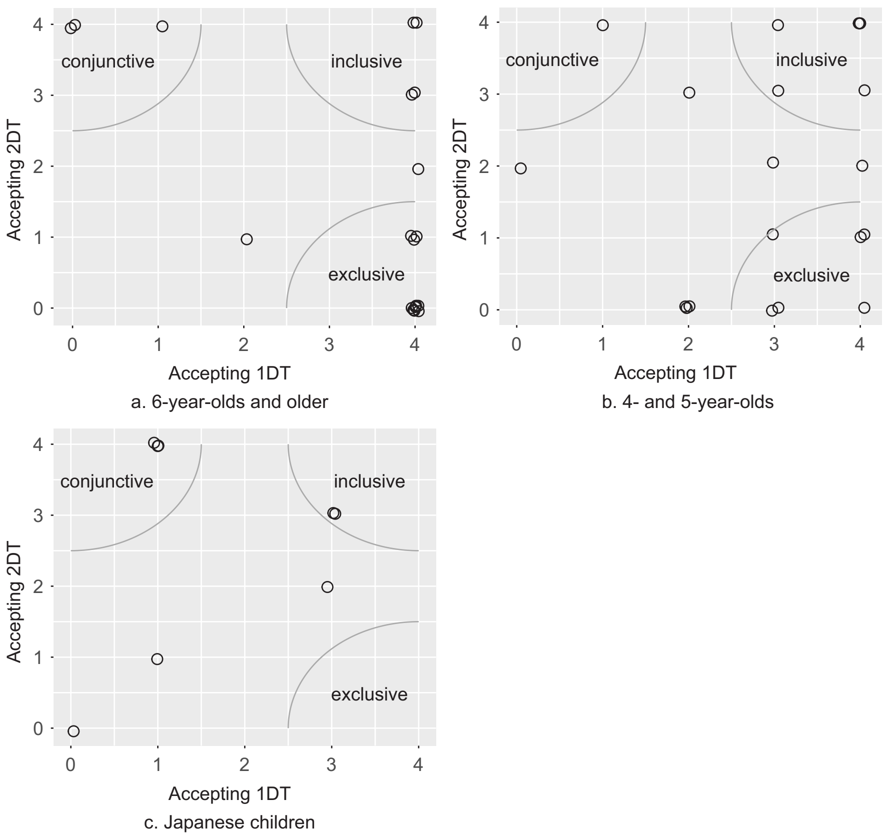



**Figure 6:** Distribution of children (complex disjunctions).

A summary of the distribution of children is shown in Table 6.

**Table 6:** Number of participants categorized by interpretation.

| | G. 4 & 5-yo | | G. 6-yo & older | | J. 4 & 5-yo | |
|---|---|---|---|---|---|---|
| | **simple** | **complex** | **simple** | **complex** | **simple** | **complex** |
| exclusive | 2 | 6 | 10 | 10 | 0 | 0 |
| inclusive | 6 | 5 | 7 | 4 | 3 | 2 |
| conjunctive | 3 | 1 | 1 | 3 | 5 | 3 |

We first tested whether we observe an effect of complexity of disjunction by checking whether there is a difference in ratios between exclusive and conjunctive interpretation within language and age group, using Fisher's exact test. Because we did not find a difference, we collapsse the two conditions (simple and complex).

When we compare the 4- and 5-year-old German speaking children and Japanese children, there is a difference between the ratio of children showing exclusive and conjunctive interpretations (Fisher's exact test: $p<0.05$). Two ages groups of German children's interpretation did not differ significantly, however (Fisher's exact test: $p=.3974$).

This is puzzling, if the exhausitifcation is obligatory when the disjunction is complex across languages. If this is the case, and if the reason that Japanese children arrive at the conjunctive interpretation is due to the unavailability of conjunction as an alternative, we expect the same from the German children.

## 5 Summary

In this chapter, we compared data from children's interpretation of disjunction in Japanese and German. Japanese was previously discussed in Tieu et al. (2016). We observe a different pattern between Japanese and German speaking children in terms of how they interpret disjunction in an upward entailing context: more children interpreting disjunction exclusively among German children.

Although more has to be investigated regarding what the source of the difference could be between the languages, one potential source may be the morphological difference we observe between German and Japanese complex disjunctions. Whereas Japanese complex disjunction is created by doubling of disjunction *ka*, German complex disjunction uses an expression *entweder* (equivalent to 'either' in English). *Entweder* has no other use but as being a part of a complex disjunction. This predicts, then, that whenever a language has a unique particle (such as *entweder* in German) that marks disjunction, children may be able to access its alternative (conjunction). Further crosslinguistic investigation is needed, however.

## References

Bates, Douglas, Martin Mächler, Ben Bolker, & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, Articles* 67. 1–48. https://www.jstatsoft.org/v067/i01.

Bowler, Margit. 2014. Conjunction and disjunction in a language without 'and'. *Semantics and Linguistic Theory* 24. 137–155.

Chierchia, Gennaro, Stephen Crain, Maria Teresa Guasti, Andrea Gualmini, & Luisa Meroni. 2001. The acquisition of disjunction: evidence for a grammatical view of scalar implicatures. *Proceedings of the Boston University Conference on Language Development* 25. 157–168. Somerville, MA: Cascadilla Press.

Chierchia, Gennaro, Stephen Crain, Maria Teresa Guasti, & Rosalind Thornton. 1998. “Some” and “or”: A study on the emergence of logical form. *Proceedings of the Boston University Conference on Language Development* 22. 97–108. Somerville, MA: Cascadilla Press.
Chierchia, Gennaro, Danny Fox, & Benjamin Spector. 2012. The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In Paul Portner, Claudia Maienborn, & Klauf von Heusinger (eds.), *Semantics: An International Handbook of Natural Language Meaning*, Vol. 3, 2297–2332. Berlin: Mouton de Gruyter.
Fox, Danny. 2007. Free choice and the theory of scalar implicatures. In Uli Sauerland & Penka Stateva (eds.), *Presupposition and Implicature in Compositional Semantics*, 71–120. Berlin, Germany: Springer.
Goro, Takuya. 2007. *Language-specific constraints on scope interpretation in first language acquisition*. College Park, Maryland: University of Maryland dissertation.
Goro, Takuya & Sachie Akiba. 2004. The acquisition of disjunction and positive polarity in Japanese. *Proceedings of the 23rd west coast conference on formal linguistics*. 251–264. Somerville, MA: Cascadilla Press.
Grice, H. Paul. 1975. *Logic and Conversation*. Cambridge, MA: Academic Press.
Horn, Laurence. 1972. *On the Semantic Properties of Logical Operators in English*. Los Angeles, CA: University of California, Los Angeles dissertation.
Meyer, Marie-Christine. 2015. Generalized free choice and missing alternatives. *Journal of Semantics* 33. 703–754.
Nicolae, Andreea C. & Uli Sauerland. 2016. A contest of strength: or versus either–or. In Nadine Bade (ed.), *Proceedings of Sinn und Bedeutung* 20. 551–568.
Sauerland, Uli. 2004. Scalar implicatures in complex sentences. *Linguistics and Philosophy* 27. 367–391.
Singh, Raj, Ken Wexler, Andrea Astle-Rahim, Deepthi Kamawar, & Danny Fox. 2016. Children interpret disjunction as conjunction: consequences for theories of implicature and child development. *Natural Language Semantics* 24. 305–352.
Skordos, Dimitrios, Roman Feiman, Alan Bale, & David Barner. 2020. Do children interpret ‘or’ conjunctively? *Journal of Semantics* 37. 247–267.
Spector, Benjamin. 2014. Global positive polarity items and obligatory exhaustivity. *Semantics and Pragmatics* 7. 1–61.
Su, Yi Esther & Stephen Crain. 2013. Children’s knowledge of disjunction and universal quantification in mandarin Chinese. *Language and Linguistics* 14. 599–631.
Sudo, Yasutada. 2014. Higher-order scalar implicatures of ‘ya’ in Japanese. Paper presented at the 9th International Workshop on Theoretical East Asian Linguistics (TEAL) at the University of Nantes, Nantes, France, 25–26 September, 2014.
Tieu, Lyn, Jacopo Romoli, Peng Zhou, & Stephen Crain. 2016. Children’s knowledge of free choice inference and scalar implicatures. *Journal of Semantics* 33. 269–298.
Tieu, Lyn, Kazuko Yatsushiro, Alexandre Cremers, Jacopo Ramoli, Uli Sauerland, & Emmanuel Chemla. 2017. On the role of alternatives in the acquisition of simple and complex disjunctions in French and Japanese. *Journal of Semantics* 34. 127–152.

Noriaki Yusa, Cornelia D. Lupsa, Naoki Kimura, Kensuke Emura, Jungho Kim, Kuniya Nasukawa, Masatoshi Koizumi, and Hiroko Hagiwara

# Chapter 11 Effects of annual quantity of second language input on pronunciation in EFL environments

## 1 Introduction

It is generally agreed that early and substantial second language (L2) exposure leads to success in the acquisition of L2 pronunciation, i.e., to the acquisition of a nativelike pronunciation, one that is indistinguishable from a native speaker's (e.g., Oyama 1976; Flege, Munro, and MacKay 1995). In this respect, "earlier is better" has become nothing short of a slogan. However, this is agreed to be the case for second language, not foreign language environments (for English, English as a Second Language, or ESL, environments versus English as a Foreign Language, or EFL, environments). The question we would like to address in this chapter is: does the *earlier-is-better* rule of thumb apply to EFL? One of the significant differences between ESL and EFL is the quantity of L2 exposure. To identify its effects, we followed our participants' changes in Voice Onset Time (VOT) for four years. To the best of our knowledge, this is the first study to observe the VOT values of Japanese-English bilingual children in an EFL environment over a period of four years. We argue that not only the total quantity of L2 input but also the annual quantity affects the production of L2 VOT, suggesting the importance of L2 input that is both continuous and consistent in quantity.

## 2 Age and L2 pronunciation

Stated broadly, early L2 learners enjoy greater success in second language pronunciation than late L2 learners. Based on a literature review of the effects of the age of learning on the acquisition of L2 pronunciation, Long (1990) estimates that the latest age at which the acquisition of nativelike pronunciation is possible is 6 years. In between 6 and 12 years, the results may vary, and after 12 years of age, it is virtually impossible to acquire nativelike pronunciation. Flege, Munro,


https://doi.org/10.1515/9783110778946-011

and MacKay (1995) conducted a large-scale study on the speech of immigrants to Canada and their perceived foreign accents, and arrived at the same conclusion: The earlier people are exposed to English, the more likely they are to acquire nativelike pronunciation. This study was done on 240 Italian immigrants, whose age of arrival (in this case coinciding with the age of learning or first exposure) was between 2 and 23 years. By the time they were tested, they had resided in Canada for an average of 32 years (the shortest length of residence being 15 years). The longer their length of residence, the longer their exposure to English, but in the end, length of residence was found to be only a very small factor. Other factors that were found to influence the acquisition of L2 pronunciation were gender and language use or frequency of use (at work, for social interaction, at home), but the most crucial factor was age. The more exigent of the native speakers who rated the English spoken by the Italian immigrants thought they detected foreign accents in immigrants who were as young as 3 or 5 when they first arrived in Canada. The conclusion is that for the acquisition of second language pronunciation, *earlier is better*, and this finding has been replicated in other studies as well (e.g., Kang and Guion 2006), and is the consensus in the field.[1] It is, however, essential to bear in mind that early learners are not always good learners of L2 pronunciation. In a follow-up study of Flege, Munro, and MacKay (1995), Flege, Frieda, and Nozawa (1997) provided interesting evidence that early learners who used Italian more often and English less often spoke English with significantly heavier foreign accents than those who used English more often and Italian less often. One possible interpretation of the finding is that successful early learners were exposed to a larger amount of L2 input as a result of the frequent use of L2. It might also be the case that early learners with detectable foreign accents often received input with Italian-accented English. The finding clearly shows that not all early immigrants necessarily receive abundant and adequate input from native speakers, suggesting the importance of the quality and quantity of input in the development of L2 pronunciation, even in L2 environments.

Most studies on the acquisition of pronunciation have been done in L2 speaking countries, that is, ESL environments. In ESL environments, learners are exposed to English in a naturalistic setting, have substantial input from native speakers at school, at work or in everyday life, and have a strong incentive to adapt and communicate. However, most learners of EFL are first exposed to it in their non-English

**1** The earlier-is-better rule of thumb does not mean that late L2 learners cannot achieve nativelike pronunciation (See Birdsong (2007, 2018) for compounding factors influencing nativelike pronunciation).

speaking home countries, in a school environment. This is what is known as an EFL environment. This means first exposure to English at maybe around 7–8 years of age (or even later), for one or two hours a week, with teachers that are not native speakers of English. As far as pronunciation goes, EFL learners have everything stacked against them: age, quality of input, and quantity of input, and as a result, a foreign accent is unavoidable for these learners.

Apart from growing up in a bilingual household, the best that an EFL environment can offer, as far as the acquisition of pronunciation goes, is an immersion program. Although immersion is still a classroom environment, it is closest to a naturalistic setting (Harada 1999, 2007). If it is an early immersion program, i.e., if it starts in kindergarten, it starts early enough for a successful or nativelike acquisition of pronunciation. The teachers are native speakers or bilingual speakers, and there is substantial, almost daily input of the target language. All these conditions create a reasonable expectation for successful acquisition/nativelike pronunciation. In other words, this is as good as it gets for an EFL environment. The questions are: Is it enough? How good can it get? At least for pronunciation, immersion studies report contradictory results (for a review of this topic, see Harada 2007; Jiang 2018).

# 3 Laryngeal-source contrasts in English and Japanese

## 3.1 Voice Onset Time

There are several ways to analyze and rate pronunciation, but the measure we used in this study is the amount of aspiration, technically known as Voice Onset Time (VOT). VOT refers to the duration of the time interval between the release of a stop closure and the onset of vocal fold vibration, i.e., voicing (Lisker and Abramson 1964), used by L2 research as an effective measurement for L2 learners' level of phonological acquisition. Take the articulation of the word "panda" for example. When "pa" in "panda" is articulated, the vibration of the vocal folds or voicing does not occur until the vowel [a] is produced (see Figure 1). After the [p] is released, there may be a short time lag before the vowel or voicing, or there may be a longer time lag. VOT measures this time lag in milliseconds, and the amount of time lag or aspiration can vary across languages. For example, English [p] in "panda" is not quite the same as Japanese [p] in its Japanese equivalent "panda", which is reflected in the difference in VOT values between English and Japanese (see Table 1).

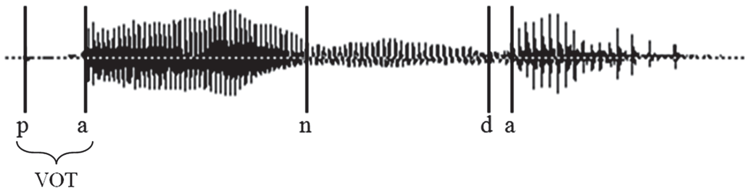



**Figure 1:** VOT in “panda”.

Pronouncing [p] in “panda” with a short time lag without aspiration will reveal that the speaker is not a native speaker of English, though aspiration is not distinctive in English. Aspiration refers to the presence of “voicelessness after the stop articulation and before the start of the voicing for the vowel” (Ladefoged 2006: 56). An English voiceless consonant stop is aspirated when it appears immediately before a stressed vowel (ex., *p*anda), except when it immediately follows an [s] (ex., s*p*ot). The aspiration rule is an unconscious rule that native speakers of English acquire as children (Knight 2012). For the purpose of this chapter, it is important to note the VOT differences between English and Japanese. Table 1 lists the mean VOT values (ms) in English and Japanese word-initial (prevocalic) stops produced by monolingual adults.

**Table 1:** The mean VOT in English and Japanese for word-initial (prevocalic) stops by produced by monolingual adults (a simplified version of Harada 2007: 372).

| | **closure** | | | **release** | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| English | | | | | | | | | | | |
| | | | | *b*<br>(7) | *d*<br>(19) | *g*<br>(22) | | | *p*<br>(68) | *t*<br>(80) | *k*<br>(88) |
| Japanese | | | | | | | | | | | |
| | *b*<br>(–27) | *d*<br>(–34) | *g*<br>(1) | | | *p*<br>(24) | *t*<br>(26) | *k*<br>(42) | | | |
| | voicing-lead<br>–VOT | | | neutral | | | | | | voicing-lag<br>+VOT | |

The English voiceless stops /p, t, k/ in word-initial (prevocalic) positions, which are called voiceless aspirated or positive VOTs, are produced with a relatively long time lag between closure release and the onset of voicing. On the other hand, the English ‘voiced’ stops /b, d, g/ exhibit a relatively short time lag between closure release and the onset of voicing, so that they are termed voiceless unaspirated or neutral VOTs. By contrast, the Japanese voiceless stops /p, t, k/ are produced with VOT values similar to those found in the English 'voiced' stops, having neutral VOT values, while the Japanese stops /b, d, g/, which are described as prevoiced stops, display a relatively long lead time (i.e., negative VOT values) between the onset of voicing and stop release, where voicing begins during the closure of stops (before the burst of stops). The question we would like to address here is whether Japanese-English bilingual children in an immersion environment can acquire the VOT values for English voiceless stops.

## 3.2 Two case studies of VOT in ESL and JFL environments

Flege (1991) examined the VOT values of English /t/ produced by Spanish learners in the United States. Note here that VOT durations for English /t/ are much longer than those for Spanish /t/ (Lisker and Abramson 1964). 10 early Spanish learners, 10 late Spanish learners, and 10 monolingual speakers of English and Spanish participated in this study. The average age of arrival (AoA) for early learners was two years and that for late learners was 20 years. Flege found that VOT values for English /t/ produced by the early learners did not significantly differ from those of the English monolinguals, while the late learners' production fell between the VOT values of the English /t/ produced by English monolinguals and Spanish /t/ of the Spanish monolinguals. This finding suggests that VOT values become less nativelike as AoA increases. The younger AoA advantage has been replicated in Flege and Eefting (1987), Flege (1987), MacKay et al. (2001), and Kang and Guion (2006).

Harada (1999), which is closest in approach to our research, was a study of native English children learning Japanese as a foreign language, in other words: Japanese as a Foreign Language or JFL environment, in an immersion program in the United States, starting from Grade 1, i.e., 6 years of age. The participants were 19 elementary school students from grade 1 (age 6), grade 3 (age 8) and grade 5 (age 10). Their VOT values for Japanese were measured and compared with the VOT values of Japanese monolingual children. The result shows that the children in the Japanese immersion program produced longer Japanese VOT values than those of the Japanese monolingual children and those of their teachers: a value which is intermediate between English VOT values and Japanese VOT values. This means that the Japanese pronunciation of the children in the immersion program had an English accent since English VOT values for initial stop consonants /p, t, k/ in pre-vocalic positions are longer than Japanese VOT values for their counterparts. This indicates that even in an immersion program, in which the amount of L2 input is substantial, children cannot make progress in achieving nativelike pronunciation. These intermediate or compromise values (between the norms for L1 and L2) are actually typical results for immersion education.[2] However, the cut-off age for nativelike pronunciation may be even lower than 6. As mentioned, Flege, Munro, and MacKay (1995) found that foreign accents can be detected in speakers that immigrated to Canada when they were as young as 3 or 5. The children in Harada's study were tested twice, with an interval of two and a half months, but they showed no improvement after the interval. As mentioned in Harada (1999), this short a time between measurements was clearly not sufficient for the children to make any progress. Therefore, it is worth pursuing this approach with younger children in an immersion program, over a longer period of time. This is what we did in our experiment, presented in the next section.

# 4 Experiment

To probe the effects of the quantity of L2 input on VOT, we measured the VOT values for the word-initial voiceless stops /p, t, k/ produced by Japanese children in an English immersion program that they started at the age of 3. We followed the children for 4 years and measured their VOT values near the end of each academic year, starting when they were around 4 years of age. In our study, the intervals between measurements were of approximately 1 year, long enough to expect the children to show progress at subsequent measurements.

## 4.1 Participants

The participants (see Table 2) were 39 Japanese-speaking children learning English at a kindergarten in Japan, who were first exposed to English at the age of 3. They attended kindergarten five days a week and spent about four hours a day with American English-speaking teachers in the first and second years (5 and 6 yrs). This means 824 hours of exposure for each of these years. Kindergarten ended at the end of the second year, and they went on to elementary school, in which the language of instruction was Japanese. They were still required to come to the immersion program four days a week after school and to spend about two hours a day speaking English. The total amount of English input received by the bilinguals over the four-year period of the immersion program amounted to 2778 hours, 824

2 Even proficient bilinguals tend to exhibit intermediate or "compromise" VOT values (for key findings in the VOT literature, see Zampini 2008; Jing 2018): L2 VOT values tend to be attracted toward L1 VOT values, while L1 VOT values of bilinguals tend toward L2 values (Birdsong 2006).

hours during each of the first and second years (5 and 6 yrs), and 565 hours during each of the third and fourth years (7 and 8 yrs). In addition to the Japanese-English bilingual children, 15 Japanese monolingual children in Japan and 29 English monolingual children in the United States also participated in this study as control groups (see Table 2). They were the same age as the Japanese-English bilingual children at the first measurement.

**Table 2:** Participants.

| Group | Number | Age |
|---|---|---|
| Japanese-English bilinguals | 39 | 5;5 |
| Japanese monolinguals | 15 | 5;5 |
| English monolinguals | 29 | 5;0 |

We started this experiment with two expectations. One was that the bilinguals' VOT values of English voiceless stops would increase over a period of four years to approximate nativelike VOT values, since early L2 learners or bilinguals produce VOT values that are similar to those produced by native speakers (Flege 1991; Kang and Guion 2006). Considering that L2 affects L1 even for early bilinguals (Yeni-Komshian et al. 2000),[3] the other expectation was that their VOT values of Japanese voiceless stops would also increase due to the effect of English long-lag VOT values, making their Japanese pronunciation less nativelike.

## 4.2 Procedure

In order to elicit the target consonants, one of the experimenters used a picture card depicting the target word and asked the participants to pronounce each target word three times, as in the following:

- Experimenter: (Showing a cue picture depicting a panda) "What's this?"
- Participant: (Looking at the picture) "Panda, panda, panda."

Only measurements of the second of the three repetitions were used as data. The data from all participants were recorded with a PCM-D1 Sony Digital Audio recorder. The VOT data were digitized and analyzed at a sampling rate of 96 kHz and stored for reviewing and listening. This experiment was carried out in both Japanese and

**3** Early bilinguals have underdeveloped L1 representations, which are more susceptible to restructuring as a result of L2 learning than those of late bilinguals (Jiang 2018).

English for the Japanese-English bilingual participants, in Japanese for the Japanese monolingual participants, and in English for the English monolingual participants. In order to minimize cross-linguistic influences, Japanese and English data for the bilingual children were recorded on different days. The bilinguals' data for the 1st measurement were collected in March 2007, those for the 2nd measurement in February 2008, those for the 3rd measurement in January and February 2009, those for the 4th measurement in January and February 2010. The Japanese monolinguals' data for the 1st measurement were collected in February 2007, those for the 2nd measurement in February 2008, and those for the 3rd measurement in July 2009. The English monolinguals' data were collected once in September 2009 in the United States.

## 4.3 Stimuli

There are no Japanese words that completely match English words in the same stop-vowel environment (Harada 2007), and we know that VOT values for stop consonants are affected by several factors such as the following vowels and stress placement. Nevertheless, such confounding factors can be minimized, as in Harada (1999), for which reason we decided to use the same words. There were nine cue words starting with voiceless stop consonants in both Japanese and English, three each for initial /p/, /t/ and /k/ (see Table 3). According to Harada (2007), the Japanese and English words were selected based on the following criteria: "(1) the following vowel quality ([a] for Japanese words and [æ] for English words; the high vowels were excluded because they are likely to be devoiced in Japanese), (2) disyllabic words, (3) the same pitch accent or stress pattern (HL for Japanese VOT data, LH for singletons and LHH for geminates, and stress on the first syllable for English VOT data) and (4) concrete words" (p.364). In this way, the factors affecting VOT values were controlled.

**Table 3:** VOT corpus.

| English VOT Corpus | | | Japanese VOT Corpus | | |
|---|---|---|---|---|---|
| **/p/** | **/t/** | **/k/** | **/p/** | **/t/** | **/k/** |
| panda | tablet | carrot | papa (papa) | tako (octopus) | kame (turtle) |
| parrot | tadpole | camel | pari (Paris) | tane (seed) | kata (shoulder) |
| package | taxi | candy | | tate (length) | kasa (umbrella) |

## 4.4 Analysis of the data

The data were analyzed using Multi-speech (KayPENTAX, Model 3700) software. The total number of target words analyzed was 2808 tokens for the Japanese-English bilingual participants (3 words × 3 repetitions × 39 participants × 2 languages × 4 years), 405 tokens for the Japanese monolingual participants (3 words × 3 repetitions × 15 participants × 3 years), and 261 tokens for the English monolingual participants (3 words × 3 repetitions × 29 participants). The VOT value measurements for all tokens were made using the Multi-Speech analysis program. The VOT values of the word-initial stops were measured by finding the nearest millisecond from "the beginning of the release burst to the onset of voicing energy in F2 formants" (Harada 2007: 365) in the display of the waveform. The mean of the VOT values of the second of the three repetitions was used as the VOT value for each of the three word-initial stop sounds /p/, /t/ and /k/. The mean VOT values obtained for each participant in the test were submitted to a 4 × 2 ANOVA in order to compare the effects on VOT values of two factors: years (1$^{st}$ measurement, 2$^{nd}$ measurement, 3$^{rd}$ measurement and 4$^{th}$ measurement), and languages (English and Japanese).

## 4.5 Results

### 4.5.1 Japanese VOT values produced by bilinguals and monolinguals

With regard to Japanese VOT values, the Japanese-English bilingual children produced an average of 27.8 ms (1$^{st}$ measurement = 25.4 ms, 2$^{nd}$ measurement = 27.6 ms, 3$^{rd}$ measurement = 30.0 ms, 4$^{th}$ measurement = 28.3 ms), and the Japanese monolingual children, 27.4 ms (1$^{st}$ measurement = 28.9 ms, 2$^{nd}$ measurement = 27.7 ms, 3$^{rd}$ measurement = 25.6 ms). Figure 2 provides the Japanese VOT values produced by the bilinguals over the four years ($F(3, 114) = 2.078$, $p = 0.107$, $\eta_p^2 = 0.052$). The mean VOT values obtained for each participant were submitted to an ANOVA of (1) Group (bilinguals and monolinguals) and (2) Interval (1$^{st}$, 2$^{nd}$, and 3$^{rd}$ measurements). The results showed that no significant Group and Interval main effects were found (Group, $F(1, 52) = 0.026$, $p = 0.873$, $\eta_p^2 = 0.000$; Interval, $F(2, 104) = 0.059$, $p = 0.943$, $\eta_p^2 = 0.001$). This means that the bilinguals' VOT values did not differ from those of the monolinguals, and that their VOT values did not change at all during the four-year period. Second language input generally affects the VOT values of the first language (Flege 1987, Harada 1999), so we expected that Japanese VOT values would increase, i.e., they would become longer and more like English, because of the increasing quantity of English exposure. However, Japanese VOT values did not change significantly during the four years, contrary to our expectation. These results suggest that the bilingual children established the same L1 phonetic category as the monolinguals', and that the English input did not affect their L1 during the four years. That is, earlier L2 input does not affect children's first language. This finding does not support the claim by Yeni-Komshian et al. (2000) that L2 affects L1 for early bilinguals. However, the contradictory results may be due to the difference in environment: EFL for our participants, ESL for those in Yeni-Komshian et al. (2000).

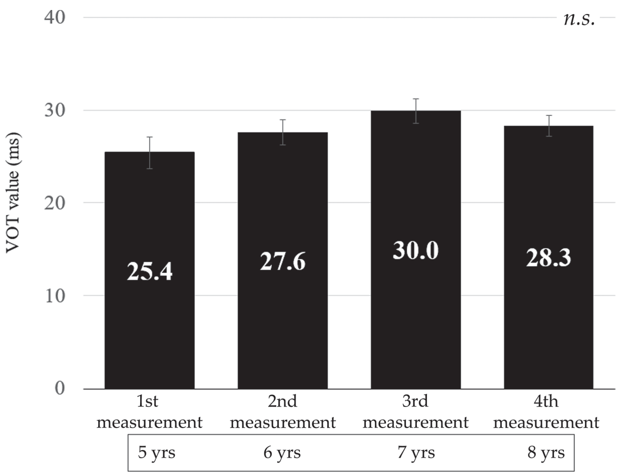



**Figure 2:** Japanese total VOT values produced by bilinguals.
(*n.s.*, not significant ($p > 0.05$))

### 4.5.2 English VOT values produced by bilinguals and monolinguals

With regard to English VOT values, the Japanese-English bilinguals produced an average of 42.3 ms ($1^{st}$ measurement = 39.1 ms, $2^{nd}$ measurement = 53.8 ms, $3^{rd}$ measurement = 39.9 ms, $4^{th}$ measurement = 36.5 ms), and the American English monolinguals, 87.0 ms. Figure 3 shows the English VOT values produced by the bilinguals over the four years. In order to compare the differences in English VOT values between the Japanese-English bilinguals and English monolinguals, we conducted a direct comparison using a *t*-test, which shows a significant difference between the two groups ($t = 9.814$, $df = 46.645$, $p = 0.000$, $d = 2.406$). This means that English monolinguals produced significantly longer English VOT values than the bilinguals. This finding supports the claim by Harada (1999) that, although bilinguals in immersion environments receive abundant L2 exposure, their L2 pronunciation does not achieve a native phonetic norm, at least in terms of VOT values. In other words, they establish an incomplete phonological category for L2, which is likely to be a result of the influence of their L1.

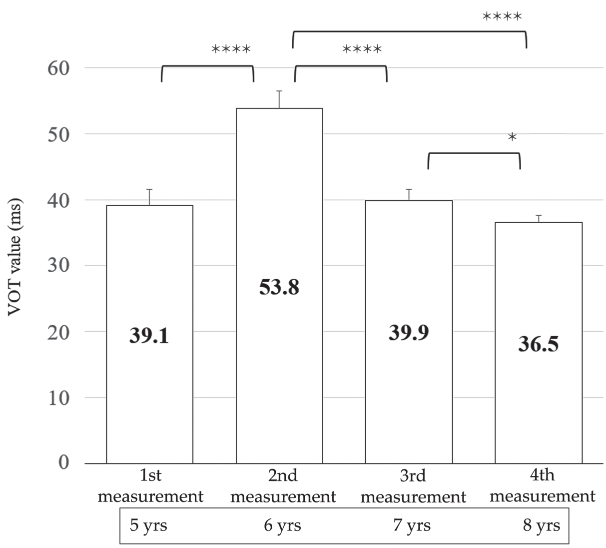



**Figure 3:** English total VOT values produced by bilinguals.
(*$p < 0.05$, ****$p < 0.001$)

In addition, a one-way ANOVA also showed a significant effect of Interval on the bilingual group ($F(3, 114) = 16.669$, $p = 0.000$, $\eta_p^2 = 0.305$), which turns out to be the most interesting finding of this study.

A post hoc analysis using a Bonferroni test revealed a significant difference between the first and second measurements ($p = 0.000$), which means that their pronunciation of English improved significantly. This in turn means that the increase in English input from the first measurement to the second measurement affected their pronunciation. As for the third measurement, our expectation was that English VOT values would continue to increase because the amount of English input also continued to increase. However, contrary to our expectation, VOT values actually decreased: there were significant differences between the second and the third measurements ($p = 0.000$), between the second and fourth measurements ($p = 0.000$) and between the third and fourth measurements ($p = 0.019$) but no significant differences between the first and the third measurements ($p = 1.000$) and between the first and the fourth measurements ($p = 1.000$). At the first measurement, the subjects' average English VOT value was 39.1ms, and at the second measurement, 53.8ms. The VOT values increased significantly. We can say that this is because the total amount of English input also increased. At the third and fourth measurements, as the total amount continued to increase, we expected the VOT values to be longer than those at the second measurement. However, the VOT values at the third measurement actually decreased as the amount of annual input decreased. At the fourth measurement, VOT values show the same regression tendency, as they decrease slightly but not significantly. This unexpected regression in the third measurement can only be explained by the decrease in the amount of exposure.

Figure 4 shows that English VOT values decreased following the decrease in the annual amount of English exposure the Japanese-English bilingual children received. They attended kindergarten five days a week and spent about 4 hours a day with American English-speaking teachers in the first and second years (5 and 6 yrs). That meant 824 hours of exposure for each of these years. The children finished kindergarten at the end of the second year (6 yrs), and went on to elementary school, in which the language of instruction was Japanese. They were still required to come to the immersion program four days a week after school and to spend about two hours a day speaking English, but the amount of exposure decreased to 565 hours per year (7 and 8 yrs). This fluctuation turned out to have a significant effect on their English pronunciation.[4]

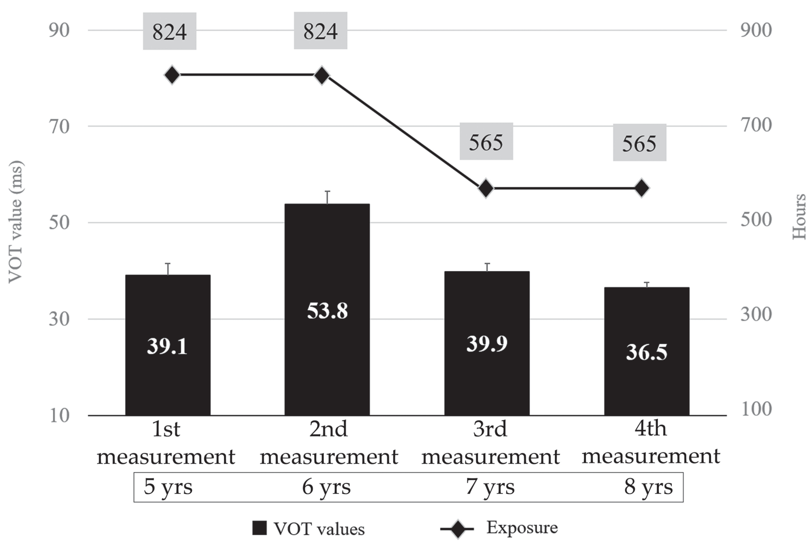



**Figure 4:** English VOT values and the annual amount of English exposure.

4 A reviewer points out that the relative proportion of time bilingual children spend on the respective languages after entering elementary school might have resulted in the decrease of the VOT values at the third and fourth measurements (7 and 8 yrs). It is true that “once they enter elementary school, the proportion of the time they spend in social activities in Japanese surpasses the time they spend using English in the after-school programs,” but in an EFL environment, they spend more time using Japanese than English *even before* they go on to elementary school. Therefore, the annual quantity of input is the most plausible factor in the decrease in VOT values.

There are two ways to look at the quantity of second language input. One is the total amount of English exposure; The other is the annual amount of English exposure. Our results tell us that not only the total quantity of input but also the annual quantity is important for the acquisition of English pronunciation. In an ESL environment, the amount of English exposure is potentially abundant, so that the learners, depending on AoA, the quantity and quality of input, and the length of residence, either improve and reach nativelike pronunciation, or reach a plateau. However, in the type of EFL environment that is immersion, the amount is subject to fluctuations. 565 hours of English exposure in a year is quite a good number for foreign language practice. But it is not enough to maintain the gain in VOT values shown in the second year.

#### 4.5.3 Japanese and English VOT values produced by bilinguals

The results for the Japanese and English VOT values clearly reveal that while the bilingual children reached Japanese monolinguals' phonetic norms, they did not reach English monolinguals' phonetic norms. Figure 5 shows the VOT values for both Japanese and English produced by the bilingual children over the four years. The expectation was that the Japanese VOT values produced by the bilinguals would be shorter than their English VOT values. As expected, the bilingual children produced shorter Japanese VOT values than their English VOT values. The mean VOT values obtained for the bilingual children were submitted to an ANOVA of (1) Interval ($1^{st}$, $2^{nd}$, $3^{rd}$, and $4^{th}$ measurements), and (2) Language (Japanese and English), which yielded significant Interval and Language main effects (Interval, $F(3, 114) = 9.055$, $p = 0.000$, $\eta_p^2 = 0.192$; Language, $F(1, 38) = 127.89$, $p = 0.000$, $\eta_p^2 = 0.771$). There was also a significant interaction between Interval and Language effects ($F(3, 114) = 17.578$, $p = 0.000$, $\eta_p^2 = 0.316$). This significant difference between Japanese VOT values and English VOT values was found for each year ($1^{st}$ measurement, $F(1, 38) = 28.294$, adjusted $p = 0.000$, $\eta_p^2 = 0.427$; $2^{nd}$ measurement, $F(1, 38) = 104.449$, adjusted $p = 0.000$, $\eta_p^2 = 0.733$; $3^{rd}$ measurement, $F(1, 38) = 15.047$, adjusted $p = 0.002$, $\eta_p^2 = 0.284$; $4^{th}$ measurement, $F(1, 38) = 10.237$, adjusted $p = 0.016$, $\eta_p^2 = 0.212$). The resulting *p*-values were adjusted with the Bonferroni correction for multiple comparisons. This means that the bilingual children produced much longer English VOT values than their Japanese VOT values, and that the ability to distinguish the two languages was maintained for the entire period of four years. These results regarding the two languages support the claim by Harada (1999) that "the immersion students are making a phonetic distinction in VOT values between Japanese and English" (p.57).

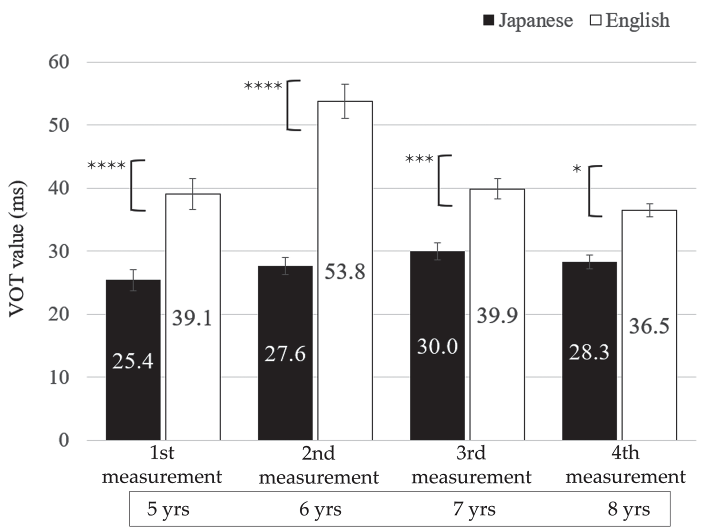



**Figure 5:** Japanese and English VOT values produced by bilinguals.
(*$p < 0.05$, ***$p < 0.005$, ****$p < 0.001$)

# 5 Discussion and conclusions

In this study, we investigated the effects of the amount of early English exposure on the acquisition of pronunciation in EFL environments. The results show that the bilingual children produced longer English VOT values than their Japanese VOT values, making a distinction between the two languages during the period of four years. This result is compatible with Harada (1999), who claims that early bilingual children are sensitive to phonetic distinctions after they are exposed to L2 for as little as a year in immersion environments and that the sensitivity can be maintained while they receive L2 exposure.

Though the bilingual children succeeded in making phonetic distinctions between the two languages, they failed to acquire the phonetic properties of the English monolingual children. In order to explain what affects nativelike perception and production of L2 phonology, Flege (1995) developed the Speech Learning Model (SLM). The SLM is based on three assumptions: Accurate perception of L2 sounds leads to accurate production; many L1 and L2 sounds exist in a "common phonological space" (Flege 1995: 238); L2 learners of all ages retain all the capacities children have in developing nativelike phonology in L2. The very fact that the bilingual children made a distinction in VOT production means that they were successful in perceiving the difference in VOT values between English voiceless stops and Japanese ones. However, it also means that successful perception is necessary but not sufficient for the successful production of L2 sounds. In our case, the quantity of L2 input the bilinguals received for four years in an EFL environment might have been insufficient for the successful mastery of L2 production.

The decrease in the annual amount of exposure after the second measurement led to a decrease in the VOT values of English stops. The VOT values did not change between the third and fourth measurements (7 and 8 yrs), in which the annual amount of exposure was the same. This indicates that although the total amount of L2 exposure increased, the bilingual children did not continue to improve their L2 pronunciation and that the annual amount of L2 exposure had a significant effect on the acquisition of L2 pronunciation. It can be said, then, that long-term success in L2 development is affected by the annual quantity of L2 input. This is similar to cases in ESL environments where the quantity of L2 input cannot be properly measured by length of residence, as shown in Flege, Frieda, and Nozawa (1997): early immigrants do not necessarily receive abundant and adequate input from native speakers and so they do not attain the nativelike phonetic norm.

In this study, we have arrived at two important conclusions. First, the amount of L2 exposure does not affect L1 pronunciation, but only L2 pronunciation in EFL environments. Second, not only the total amount of L2 exposure but also the annual amount of L2 exposure is important for the acquisition of L2 pronunciation in EFL environments. It has been widely believed that *earlier is better* for the acquisition of L2 pronunciation. However, this study clearly shows that the acquisition of L2 pronunciation is related to not only the age factor but also the amount of L2 exposure, and that abundant and consistent L2 input exposure is important in EFL environments. Even if L2 learning starts early, the early-onset advantage diminishes unless abundant and consistent input is provided, suggesting that early learners might not always outperform late learners. One must consider examining the quality and quantity of L2 input before drawing the conclusion that "earlier is better" in L2 acquisition.

**Acknowledgments**: This chapter was supported by the JST Grant-in-Aid for the promotion of "Brain Science and Education, Type II," (PI: Hiroko Hagiwara). The results have not been published since the project finished in 2010. We are privileged to dedicate the chapter to the late Hiroko Hagiwara, who played an invaluable role in the project. We thank S. Yamaguchi for her indispensable role in data collection. For her own personal reasons, she declined to be listed as one of the co-authors but agreed to let us use her unpublished paper and the data, which we re-analyzed. We are also indebted to David Birdsong for helping us collect the VOT data of the American-English speaking children. Finally, we would like to thank the two reviewers and the audience at the International Symposium on Issues in Japanese Psycholinguistics from Comparative Perspectives for invaluable comments and suggestions. All remaining errors are our own.

## References

Birdsong, David. 2006. Age and second language acquisition and processing: A selective overview. *Language Learning* 56. 9–49.

Birdsong, David. 2007. Nativelike pronunciation among late learners of French as a second language. In Ocke-Schwen Bohn & Murray J. Munro (eds.), *Language Experience in Second Language Speech Learning*, 99–116. Amsterdam: John Benjamins.

Birdsong, David. 2018. Plasticity, variability and age in second language acquisition and bilingualism. *Frontiers in Psychology* 9. 81. http://doi: 10.3389/fpsyg.2018.00081

Flege, James E. 1987. The production of 'new' and 'similar' phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15, 47–65.

Flege, James E. 1991. Age of learning affects the authenticity of voice onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America* 89(1). 395–411.

Flege, James E. 1995. Second-language speech learning: Theory, findings, and problems. In Winifred Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, 233–277. Baltimore: York Press.

Flege, James E. & Wieke Eefting. 1987. Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication* 6. 185–202.

Flege, James E., Elaina M. Frieda, & Takeshi Nozawa. 1997. Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics* 25. 169–186.

Flege, James E., Murray J. Munro, & Ian R. A. MacKay. 1995. Effects of age of second-language learning on the production of English consonants. *Speech Communication* 16. 1–26.

Harada, Tetsuo. 1999. *The acquisition of segmental timing by children in a Japanese immersion program*. Los Angeles, CA: UCLA dissertation.

Harada, Tetsuo. 2007. The production of voice onset time (VOT) by English-speaking children in a Japanese immersion program. *International Review of Applied Linguistics in Language Teaching* 45. 353–378.

Jiang, Nan. 2018. *Second Language Processing: An Introduction*. New York/London: Routledge.

Kang, Kyoung-Ho & Susan G. Guion. 2006. Phonological systems in bilinguals: Age of learning effects on the stop consonant systems of Korean-English bilinguals. *Journal of the Acoustical Society of America* 119. 1672–1683.

Knight, Rachael. 2012. *Phonetics: A Coursebook*. Cambridge: Cambridge University Press.

Ladefoged, Peter. 2006. *A Course in Phonetics*. Boston: Thomson Wadsworth.

Lisker, Leigh & Arthur Abramson, S. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20. 384–422.

Long, Michael H. 1990. Maturational constraints on language development. *Studies in Second Language Acquisition* 12. 251–285.

MacKay, Ian R. A., James E. Flege, Thorsten Piske, & Carlo Schirru. 2001. Category restructuring during second language speech acquisition. *The Journal of the Acoustical Society of America* 110. 516–528.

Oyama, Susan. 1976. A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research* 5. 261–283.
Yeni-Komshian, Grace H., James E. Flege, & Serena Liu. 2000. Pronunciation proficiency in the first and second languages of Korean-English bilinguals. *Bilingualism: Language and Cognition* 3. 131–149.
Zampini, Mary L. 2008. L2 speech production research: Findings, issues, and advances. In Jette G. Hansen Edwards & Mary L. Zampini (eds.), *Phonology and Second Language Acquisition*, 219–249. Amsterdam: John Benjamins.

Kexin Xiong, Keiyu Niikuni, Toshiaki Muramoto, and Sachiko Kiyama

# Chapter 12
# Asymmetric effects of sub-lexical orthographic/phonological similarities on L1-Chinese and L2-Japanese visual word recognition

## 1 Introduction

About half of the world's population can use more than one language (Grosjean 2010). Learning how bilinguals process the languages in their minds could shed light on language processing architecture and provide helpful information for learning a second language. There are ample data suggesting that when bilinguals read in either their first language (L1) or their second language (L2), both languages are automatically activated (i.e., language non-selective activation; e.g., Dijkstra, Grainger, and Van Heuven 1999; Duyck et al. 2007; Hsieh et al. 2021; Mulder, Dijkstra, and Baayen 2015; Peeters, Dijkstra, and Grainger 2013; Vanlangendonck et al. 2020; Woumans, Clauws, and Duyck 2021). Consequently, translation equivalents with overlapping word forms between L1 and L2 (i.e., cognates) are processed more quickly than matched control words that do not share word forms (i.e., non-cognates), depending on the given task. For instance, in an L2-English lexical decision task (LDT) conducted with (late) Dutch-English bilinguals, cognates with identical orthographic forms (e.g., *lamp* in English and Dutch) required shorter reading times compared to the matched control words (e.g., *song* in English but *lied* in Dutch; Dijkstra et al. 2010). This cognate facilitation effect has consistently been reported in previous studies using languages with shared scripts (e.g., Dutch-English: Cop et al. 2017; French-English: Libben and Titone 2009; Spanish-English: Hoshino and Kroll 2008) and languages with cross-scripts (Arabic-Hebrew: Degani, Prior, and Hajajra 2018; Chinese-English: Zhang et al. 2019; Japanese-English: Nakayama et al. 2013).

**Acknowledgments:** This research was supported by the Grant-in-Aid for FY2018 JSPS Postdoctoral Fellowship for Research in Japan (Standard, 18F18302) and the Frontier Research Institute for Interdisciplinary Sciences (FRIS), Tohoku University. We thank Lingfeng Song and Ya Zhuo for their help with data collection.


https://doi.org/10.1515/9783110778946-012

Because cognates exist in both of bilinguals' languages and can vary in their degree of orthographic and phonological overlap between L1 and L2, they have been a useful tool for investigating bilinguals' mental representation and the mechanism of lexical processing. Studies have shown that orthographic similarity across languages is associated with cognate reading. The magnitude of the cognate effects tends to be larger for those with greater orthographic overlap in L2-LDTs (e.g., Dijkstra et al. 2010; Van Assche et al. 2011). In terms of phonological overlap, research that use languages with cross-scripts has shown that a higher degree of phonological similarity results in a stronger cognate effect (e.g., Miwa, Dijkstra, et al. 2014). These findings are discussed within a localist framework – the bilingual interactive activation plus model (BIA +; Dijkstra and Van Heuven 2002), which is a successor to the bilingual interactive activation model (BIA; Dijkstra and Van Heuven 1998). From the interactive activation viewpoint, visual presentation of a word will activate all candidates with orthographic form overlap, including those in the non-target language for bilingual readers. The visual features of the input word are predicted to activate orthographic, phonological, and semantic representations across the two languages, with the amplitude of activation depending on the degree of overlaps. However, the matched non-cognates only activate one of the bilingual's languages. Besides cross-linguistic overlaps, word frequency also has a significant effect on lexical access. In late (or unbalanced) bilinguals who read more often in their native language, cognates' subjective frequency in L1 is higher than that in L2, thus leading to a smaller frequency effect on word processing in L1 than in L2 (e.g., Gollan et al. 2008). In the same vein, L2 word processing benefits more from cross-linguistic characteristics compared to L1 (Duyck 2005; Peleg et al. 2020; Zhang et al. 2019).

Most research has focused on the role of orthographic and phonological overlap at the lexical level, with only a few studies exploring similarities between languages at the sub-lexical level. A recent study using a masked priming lexical decision task in L2 revealed that cross-language similarities of words' morphemic constitutes affected European Portuguese-English bilingual participants' lexical processing (Comesaña et al. 2018), indicating that cross-language similarities at the sub-lexical level play an essential role in visual word processing. However, as the authors mention, given that cognate effects in the forward direction (L1-L2) are usually more pronounced than in the backward direction (L2-L1), sub-lexical effects may differ in L1 and L2. Additionally, because both European Portuguese and English use alphabetic scripts, the cross-language similarity refers to both orthographic and phonological overlap, making it difficult to distinguish between these two kinds of effects. Therefore, it remains unclear how cross-language orthographic and phonological similarities at the sub-lexical level affect bilinguals' word processing.

This question can be addressed by examining cognates containing two characters in Chinese and Japanese that use logographic scripts. Approximately 72% of the words in Chinese (Cui et al. 2021) and 70% of the entries in a Japanese-language dictionary (Yokosawa and Umeda 1988) consist of two characters (hereafter, two-character compound words). Many overlap at the orthographic, phonological, or semantic levels. In a database for Japanese, Korean, Chinese, and Vietnamese, which contains 2,058 Japanese two-character compound words (JKCV database; Park, Xiong, and Tamaoka 2014; Yu and Tamaoka 2015; available online at http://kanji-godb.herokuapp.com/), 56.51% of the words have identical or similar orthographic and semantic meanings (e.g., Chinese-Japanese cognates; Xiong 2018). Unlike typical cognates in languages with alphabetic scripts, their pronunciations are not always similar in Chinese and Japanese. For example, the cognate 景色 ("scenery") is orthographically identical in Chinese and Japanese, whereas it is pronounced as /ke.shiki/ in Japanese but /jing.se/ in Chinese. More importantly, each character could be considered a morpheme, and the degree of orthographic or phonological overlap at the sub-lexical (i.e., character) level can vary between the two languages.

Evidence suggests that orthographic information is crucial in logographic reading. Using an L2-Japanese LDT with event-related potentials (ERP), Xiong, Verdonschot, and Tamaoka (2020) found that orthographically identical cognates required decreased reaction times (RTs) and N250 (a component potentially involved in word formation during visual word recognition), indicating that orthographic similarity between L1-Chinese and L2-Japanese facilitated cognate reading. On the other hand, the contribution of phonology remains debatable. Several psycholinguistic studies have shown that visual recognition of two-character compound words in Chinese (Wong, Wu, and Chen 2014) and Japanese (Chen et al. 2007) does not require phonological activation. Others have demonstrated that phonological information is automatically activated during visual word recognition, as the initial characters' number of homophones affected the processing of two-character compound words in a Japanese LDT conducted with native Japanese speakers (Tamaoka 2005). These studies often use RTs to evaluate word processing mechanisms; however, it is difficult to capture all cognitive processes in end-point responses because they include the entire recognition and decision-making process (Miwa, Libben, and Ikemoto 2017). Even isolated visual word recognition requires multiple fixations. Using eye-tracking technology allows us to build a timeline when certain factors are involved in visual word recognition (e.g., Kuperman et al. 2009; Miwa, Libben, et al. 2014). Indeed, an eye-tracking study (Miwa, Libben, et al. 2014) reported that phonological effects during visual processing of Japanese compound words occurred only in a late stage of processing (i.e., second fixation duration and RTs), whereas orthographic effects occurred at the beginning of processing (i.e., first fixation duration).

To summarize, in this chapter, we aimed to clarify (1) the extent to which orthographic and phonological information at the sub-lexical level contribute to bilinguals' cognate reading and (2) whether sub-lexical information affects L2 cognate reading in a similar manner as it affects L1. We conducted an L2-Japanese LDT (Experiment 1) and an L1-Chinese LDT (Experiment 2) together with eye-tracking. Using the first and second fixation durations, as well as RTs during visual word recognition, we addressed the temporal locus of the effects of orthographic and phonological overlap that arise at the sub-lexical level when bilinguals read cognates in their L2 (Japanese) and L1 (Chinese). Because visual word recognition studies have consistently demonstrated the effects of orthographic form overlap (e.g., Dijkstra et al. 2010), whereas the role of phonological information in processing logographic scripts still remains controversial (Chen et al. 2007; Miwa, Libben, et al. 2014), we predicted a strong facilitating effect of orthographic similarity but only a limited effect of phonological similarity at the sub-lexical level. Considering that cross-linguistic similarities benefit L2 more than L1, we expected these effects to be more pronounced in L2.

# 2 Experiment 1: L2-Japanese LDT

## 2.1 Method

### 2.1.1 Participants

We recruited a total of 31 native Mandarin speakers (L1; seven males; mean age = 25.95 years, standard deviation (*SD*) = 1.74) proficient in Japanese (L2) from Tohoku University in Japan. On average, they have learned Japanese for more than five years (*M* = 70.03 months, *SD* = 23.51) and lived in Japan for more than one year (*M* = 30.84 months, *SD* = 13.03). All participants passed the most challenging level (N1) of the Japanese-Language Proficiency Test administered by the Japan Foundation and Japan Educational Exchanges and Services' joint organization. Using a 7-point Likert scale, all participants completed self-assessment questionnaires regarding their L1-Chinese and L2-Japanese proficiency and frequency of regular use of each language (Table 1).

To evaluate their level of Japanese proficiency at the time of the experiment, we also administered the Tsukuba Test – Battery of Japanese (TTBJ, http://ttbj-tsukuba.org/) online. It includes a Simple Performance-Oriented Test (SPOT; testing the ability of communication in Japanese), a test for grammar, and a test for kanji knowledge (Kanji-SPOT). Based on the TTBJ standard, participants' communication abilities in Japanese (SPOT score: *M* = 74.94, *SD* = 6.11) and grammar knowledge (*M* = 70.74, *SD* = 7.34) were at intermediate levels. However, their Japanese kanji-words knowledge was at an advanced level (Kanji-SPOT: *M* = 46.87, *SD* = 1.62). Participants

**Table 1:** Mean (SD) self-ratings of L1-Chinese and L2-Japanese proficiency and frequency of use with a 7-point Likert scale (Proficiency: 1 = unable to 7 = excellent; Frequency: 1 = rarely to 7 = very often).

| | Listening | Speaking | Reading | Writing |
|---|---|---|---|---|
| Proficiency | | | | |
| L1-Chinese | 7.00 (0.00) | 6.97 (0.18) | 6.97 (0.18) | 6.87 (0.34) |
| L2-Japanese | 5.06 (0.76) | 4.32 (0.96) | 5.35 (0.78) | 4.26 (0.95) |
| Frequency | | | | |
| L1-Chinese | 5.29 (1.71) | 5.39 (1.58) | 5.16 (2.16) | 4.10 (2.05) |
| L2-Japanese | 5.71 (1.22) | 4.58 (1.48) | 4.97 (1.62) | 3.81 (1.61) |

provided informed consent before the experiment. This study was approved by the ethical committee of the Graduate School/Faculty of Arts and Letters at Tohoku University.

### 2.1.2 Materials

We used four types of Chinese-Japanese cognates as target stimuli, with each type containing 36 words: (1) cognates with identical orthographic forms,[1] (2) cognates with the same initial characters, (3) cognates with the same final characters, and (4) those in which both characters have similar word forms in Chinese and Japanese. All cognates originated from the JKCV database.

Participants rated the degree of orthographic and phonological similarity between L2-Japanese and L1-Chinese after the experiments using a 7-point Likert scale (1 = not similar at all, 7 = same). Table 2 summarizes the ratings for each type. We collected word frequencies and frequencies for each character in Japanese from an online database (www.kanjidatabase.com; Tamaoka et al. 2017), and we collected word/character frequencies in Chinese from the newspaper genre of the BCC corpus (http://bcc.blcu.edu.cn/index.php; Xun et al. 2016). Instead of using the cognate types as a factor, we treated the ratings of cross-language orthographic

**1** Regarding the identical cognates, we carefully selected words that will not be affected by the fonts commonly used in each language. That is, characters such as '包' in Japanese and '包' in Chinese were not included in the identical cognate group. However, some participants rated identical cognates as extremely similar but not identical, possibly because the fonts used in each language were different in appearance. For example, the word 'liquid' in Japanese font '液体' appeared thinner in Chinese font '液体'. Therefore, in the data analysis, we used the ratings of orthographic similarity by participants rather than the four predefined groups.

and phonological overlap as well as the frequencies in L1 and L2 as continuous variables in data analysis, leveraging mixed-effects regressions (Baayen 2008). In addition to target stimuli, we included 36 non-cognates (i.e., fillers) and 180 non-words in the experiments. To ensure that no identical stimuli appeared in both experiments, we divided the stimuli into two lists of 180 items each. The stimuli list and the order of the two experiments were counterbalanced.

**Table 2:** Mean (SD) ratings of orthographic and phonological similarity for each character.

| Types | Examples | | Orthographic similarities | | Phonological similarities | |
|---|---|---|---|---|---|---|
| | Japanese | Chinese | Initial character | Final character | Initial character | Final character |
| 1 | 湿度 (/shitsu.do/) | 湿度 (/shi.du/) | 6.92 (0.09) | 6.95 (0.04) | 3.46 (1.13) | 3.77 (1.50) |
| 2 | 音楽 (/on.gaku/) | 音乐 (/yin.yue/) | 6.93 (0.10) | 4.09 (0.94) | 3.25 (1.14) | 3.15 (1.52) |
| 3 | 動物 (/do.butsu/) | 动物 (/dong.wu/) | 4.23 (1.00) | 6.89 (0.07) | 3.74 (1.28) | 3.87 (1.33) |
| 4 | 鉛筆 (/em.pitsu/) | 铅笔 (/qian.bi/) | 4.07 (0.86) | 4.28 (0.90) | 3.43 (1.45) | 3.31 (1.16) |

*Note*: Types one to four refer to cognates with identical orthographic forms, cognates with the same initial characters, cognates sharing the same final characters, and cognates in which both characters have similar word forms. The similarities were rated using a 7-point Likert scale.

### 2.1.3 Apparatus and procedure

Participants performed an L2-Japanese LDT combined with eye-tracking in a sound-proof room. Each trial began with a white fixation mark on a black screen for 500 ms, followed by a stimulus containing two kanji characters. The stimulus disappeared immediately once participants entered a response. All stimuli were presented on a 21” display (FlexScan T965, Eizo, Japan), in 48-point white Mincho font (an old-fashioned font of Japanese) on a black background. The WinPython-PyGaze-0.5.1 software package (Dalmaijer, Mathôt, and Van der Stigchel 2014) controlled the presentation of stimuli and the collection of response latencies and accuracy for each trial. An EyeLink 1000 Plus desktop mounted system (SR Research, Canada) recorded eye movements at a rate of 1000 Hz. Participants’ eyes were calibrated using a 9-point calibration.

Participants were instructed to decide whether the presented stimulus was a real Japanese word as quickly and as accurately as possible. They indicated their decisions by pressing specified buttons with their right middle finger (real words) or their left middle finger (non-words). Participants completed 20 practice trials to gain familiarity with the task prior to the experimental trials. Excluding the practice and calibrations, the experiment lasted approximately 10 minutes.

### 2.1.4 Data analysis

We analyzed the first fixation duration, the second fixation duration, and the total RTs of the target words,[2] using linear mixed-effects models (Baayen 2008) with the lme4 package (version 1.1-21, Bates et al. 2015) in R version 3.6.0 (R core Team 2019). We eliminated incorrect responses for target words (16 trials, 0.72% of trials for target words), trials with RTs over 3,000 ms, and those with absolute standardized residuals exceeding 2.5 standard deviations (80 trials, 3.75% of trials) before constructing any models. Fixation durations were analyzed with trials of target words that required multiple fixations (84.36% of trials: 53.46% and 30.90% for exact two fixations and three or more fixations, respectively), and the fixation durations with absolute standardized residuals above 2.5 SD (2.97% for the first fixation and 2.27% for the second fixation) were excluded from the analyses. In these analyses, we treated the two variables of interest as fixed-effects: cross-language (1) orthographic similarity and (2) phonological similarity for each character. Word frequency, the frequency of each character in both L1-Chinese and L2-Japanese, and the trial numbers were included in the analyses as covariate fixed effects. We constructed models with the main effects and the interactions among these factors. The residuals of the word frequency and the frequency for each character in L2-Japanese were used to avoid multicollinearity. As for the analysis of second fixation durations, besides the abovementioned predictors, the first fixation durations were also included as a covariate. The random effects were structured with random intercepts for participants and items. The effects were selected by backward elimination and the optimal models were selected based on Akaike's Information Criterion (AIC). All data were *z*-transformed and the final list of fixed-effects of the optimal models are summarized in Table 3.

2 The fixation mark was placed at the middle of the two-character compound stimulus, and over half of the words had only two fixations, which make it difficult to analyze the fixation times on each character. Therefore, we analyzed the first and second fixations on the entire word, using sub-lexical features of each character as index to investigate the time-course of two-character compound word processing.

**Table 3:** RTs, first fixation durations, and second fixation durations for cognate reading in L2-Japanese.

| | β | Std. Error | *df* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|---|
| *RTs* | | | | | | |
| (Intercept) | − .230 | .018 | 132 | −12.673 | < .001 | [− .265, − .195] |
| InitialC orthographic similarity | − .104 | .019 | 134 | −5.578 | < .001 | [− .140, − .068] |
| InitialC phonological similarity | − .069 | .019 | 133 | −3.677 | < .001 | [− .104, − .033] |
| L2 word frequency | − .053 | .018 | 134 | −2.884 | .005 | [− .088, − .018] |
| L1 word frequency | − .098 | .019 | 135 | −5.050 | < .001 | [− .136, − .061] |
| InitialC L1 frequency | − .061 | .019 | 134 | −3.167 | .002 | [− .097, − .024] |
| FinalC L1 frequency | − .040 | .019 | 131 | −2.136 | .035 | [− .076, − .004] |
| Trials | − .040 | .014 | 2031 | −2.794 | .005 | [− .068, − .012] |
| *First fixation durations* | | | | | | |
| (Intercept) | − .093 | .052 | 30 | −1.785 | .084 | [− .196, .011] |
| InitialC orthographic similarity | .076 | .018 | 1666 | 4.205 | < .001 | [ .041, .112] |
| InitialC L1 frequency | .034 | .018 | 1667 | 1.893 | .059 | [− .001, .070] |
| *Second fixation durations* | | | | | | |
| (Intercept) | − .086 | .068 | 30 | −1.261 | .217 | [− .223, .050] |
| L2 word frequency | − .062 | .018 | 1606 | −3.435 | < .001 | [− .097, − .026] |
| InitialC L1 frequency | − .043 | .018 | 1604 | −2.409 | .016 | [− .078, − .008] |
| Previous fixation durations | − .251 | .019 | 1624 | −13.080 | < .001 | [− .288, − .213] |

*Note:* CI = confidence interval; InitialC = initial character; FinalC = final character.

## 2.2 Results

The average accuracy for cognates (i.e., targets; *M* = 99.28%, *SD* = 0.08), Japanese non-cognates (i.e., fillers; *M* = 89.43%, *SD* = 0.31), and non-words (*M* = 96.99%, *SD* = 0.17) exceeded 85%. Cognates were responded to more accurately than non-cognates ($z = 6.101$, $p < .001$). The orthographic similarity (β = − .104, $t = -5.578$, $p < .001$) and phonological similarity (β = − .069, $t = -3.677$, $p < .001$) of the initial characters, the character frequency in L1-Chinese (Initial character: β = − .061, $t = -3.167$, $p = .002$; final character: β = − .040, $t = -2.136$, $p = .035$), and the word frequency in both L1-Chinese (β = − .098, $t = -5.050$, $p < .001$) and L2-Japanese (β = − .053, $t = -2.884$, $p = .005$) significantly affected RTs for cognates. As shown in Figure 1, RTs declined along with increasing similarities, for both orthography and phonology of the initial character.

To assess the time course of the form overlapping effect at the character level more precisely, we further analyzed the first fixation and second fixation durations for cognates. As a result, the initial character's orthographic similarity was significant (β = .076, $t = 4.205$, $p < .001$) in the beginning of processing (i.e., first fixation

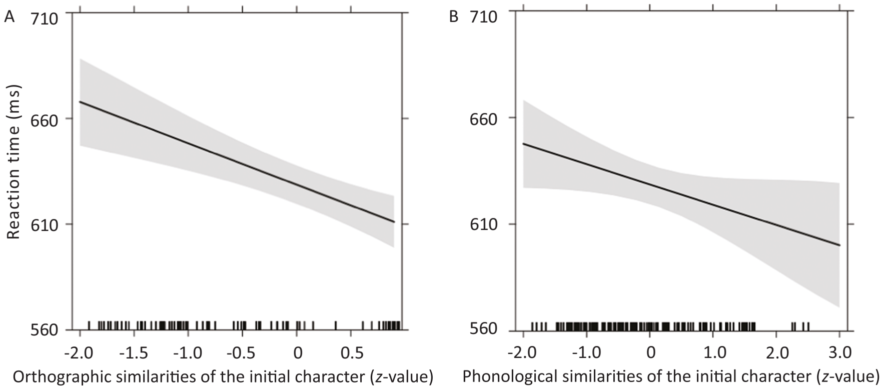



**Figure 1:** The effects of orthographic similarities (A) and phonological similarities (B) of the initial character on L2-Japanese reading in terms of reaction time. The grey shaded area represents a 95% confidence interval, and the rugs on the x-axis represent the marginal distribution of the predictor.

duration). Contrary to the facilitating effect we observed for RTs, the first fixation duration increased as orthographic similarity increased for the initial character (Figure 2). Additionally, we found no significant effect of form similarity but did find significant effects of the L1-Chinese initial character frequency ($\beta = -.043$, $t = -2.409$, $p = .016$) and L2-Japanese word frequency ($\beta = -.062$, $t = -3.435$, $p < .001$) regarding second fixation durations.

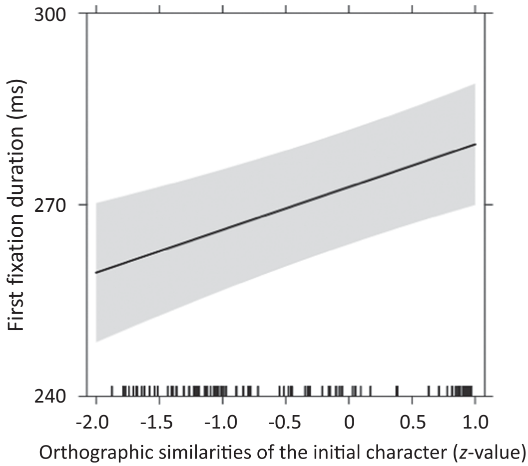



**Figure 2:** The effects of orthographic similarities of the initial character on L2-Japanese reading during first fixation period. The grey shaded area represents a 95% confidence interval, and the rugs on the x-axis represent the marginal distribution of the predictor.

## 2.3 Discussion

In Experiment 1, cognates were more accurately comprehended than non-cognates, suggesting that L1-Chinese knowledge activated automatically and facilitated lexical reading in L2-Japanese. More importantly, we found both orthographic and phonological form overlaps at the character level to be associated with L2-Japanese cognate reading. Participants tended to process stimuli with high degrees of orthographic and phonological similarity more rapidly than those with less form overlaps. Our results provide further evidence for the character-driven processing model (Miwa, Libben, et al. 2014), suggesting that not only monolingual but also bilingual readers process compound words based on the features of the characters in logographic scripts.

Surprisingly, the orthographic similarity of the initial character had an inhibitory effect on cognate reading in the early stage of processing (i.e., long first fixation durations), which contradicted our expectations as well as previous findings from eye-tracking studies conducted with native Japanese monolingual individuals. For example, Miwa, Libben, and Ikemoto (2017) administered a Japanese LDT to native Japanese speakers using trimorphemic compound words (i.e., words that consist of three kanji-characters), finding that initial characters with more visually complex features elicited longer first fixation durations but also that middle and final characters with more complex visual features required shorter fixation durations. They interpreted these findings following Hyönä and Bertram (2004), demonstrating that visual feature complexity effects interfere with character processing at the foveal region but facilitate recognition at the parafoveal region. If this is the case, then in the present study, initial characters with lower orthographic similarity should have had longer first fixation durations, while the final characters with less orthographic overlap should have facilitated cognate reading, which seems to be the opposite. Note that in Miwa, Libben, and Ikemoto (2017), the fixation mark was placed on the initial character/element of the compound words; however, it was located in the middle of the stimulus in the present study, which may explain the contradictory results. We discuss these findings in more detail in the *General discussion*.

Given that word processing in L2 usually takes advantage of cross-linguistic features more than in L1, this may also be true for sub-lexical effects. In Experiment 2, we investigated this question using an L1-Chinese LDT.

# 3 Experiment 2: L1-Chinese LDT

## 3.1 Method

### 3.1.1 Participants

The participants were the same as those in Experiment 1.

### 3.1.2 Materials

The cognates and non-words were the same as those used in Experiment 1, but they were presented using L1-Chinese (see the difference of orthographic form between L1-Chinese and L2-Japanese in Table 2). Non-cognates were Chinese translation equivalents of the fillers that were used in Experiment 1.

### 3.1.3 Apparatus and procedure

All the apparatus and procedures were identical to those of Experiment 1 except that this task was conducted in L1-Chinese. All the stimuli were displayed using the Chinese font Simsun. Participants were instructed to decide whether the presented stimulus was a Chinese word as quickly as accurately as possible.

### 3.1.4 Data analysis

We eliminated incorrect responses for target words (37 trials, 1.66% of trials), RTs shorter than 300 ms (one trial) or longer than 3,000 ms (three trials), and trials with absolute standardized residuals exceeding 2.5 standard deviations (59 trials, 2.87% of trials) before constructing any models. Those trials requiring multiple fixations were analyzed (77.20% of trials: 55.32% and 21.88% for exact two fixations and three or more fixations, respectively). We eliminated fixation durations that had absolute residuals greater than 2.5 SD (1.41% for the first fixation and 2.45% for the second fixation). The analyses of RTs, first fixation durations, and second fixation durations that relied on target words were identical to those of Experiment 1. We summarized the fixed-effects of all optimal models in Table 4.

**Table 4:** RTs, first fixation durations, and second fixation durations for cognate reading in L1-Chinese.

| | β | Std. Error | *df* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|---|
| *RTs* | | | | | | |
| (Intercept) | – .202 | .022 | 136 | −9.211 | < .001 | [– .244, – .160] |
| InitialC orthographic similarity | – .051 | .023 | 137 | −2.239 | .027 | [– .094, – .007] |
| L1 word frequency | – .111 | .022 | 136 | −5.029 | < .001 | [– .154, – .069] |
| InitialC L2 frequency | – .034 | .023 | 135 | −1.522 | .130 | [– .078, .009] |
| FinalC L2 frequency | – .036 | .023 | 135 | −1.574 | .118 | [– .081, .008] |
| InitialC : FinalC L2 frequency | .081 | .026 | 140 | 3.131 | .002 | [ .031, .131] |
| *First fixation durations* | | | | | | |
| (Intercept) | – .049 | .073 | 30 | −0.668 | .509 | [– .193, .096] |
| InitialC phonological similarity | – .045 | .018 | 2003 | −2.492 | .013 | [– .080, – .010] |
| *Second fixation durations* | | | | | | |
| (Intercept) | – .057 | .057 | 30 | −0.992 | .329 | [– .171, .057] |
| FinalC phonological similarity | – .030 | .016 | 2121 | −1.840 | .066 | [– .061, .002] |
| L1 word frequency | – .066 | .016 | 2127 | −4.054 | < .001 | [– .097, – .034] |
| Previous fixation durations | – .365 | .016 | 2126 | −22.294 | < .001 | [– .397, – .332] |

*Note:* CI = confidence interval; InitialC = initial character; FinalC = final character.

## 3.2 Results

Participants exhibited high accuracy in recognizing cognates ($M$ = 98.34%, $SD$ = 0.13), L1-Chinese non-cognates ($M$ = 96.24%, $SD$ = 0.19), and non-words ($M$ = 98.03%, $SD$ = 0.14). Like the results in Experiment 1, the average accuracy for cognates was higher than that of non-cognates ($z$ = 2.467, $p$ = .036). Regarding the RTs for cognates, we found a significant effect of orthographic similarities of the initial characters (β = − .051, $t$ = − 2.239, $p$ = .027). As shown in Figure 3, RTs declined with increasing form overlaps. There was also a significant Chinese word frequency effect (β = − .111, $t$ = − 5.029, $p$ < .001) and an interaction effect between the L2-Japanese frequency of the initial character and the final character (β = .081, $t$ = 3.131, $p$ = .002).

As for the analyses of fixation durations, unlike in L2-Japanese processing, there was no significant effect of orthographic form overlap, but the effect of phonological similarity of the initial characters (β = − .045, $t$ = − 2.492, $p$ = .013) reached significance during the first fixation. As shown in Figure 4, a higher degree of phonological similarity led to a shorter fixation duration. Additionally, there was no significant effect of word form overlap, but L1-Chinese word frequency (β = − .066, $t$ = − 4.054, $p$ < .001) significantly contributed to word reading during the second fixation.

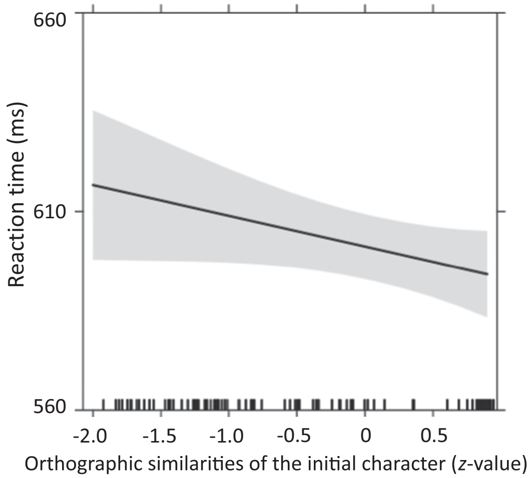



**Figure 3:** The effects of orthographic similarities of the initial character on L1-Chinese reading in terms of reaction time. The grey shaded area represents a 95% confidence interval, and the rugs on the x-axis represent the marginal distribution of the predictor.

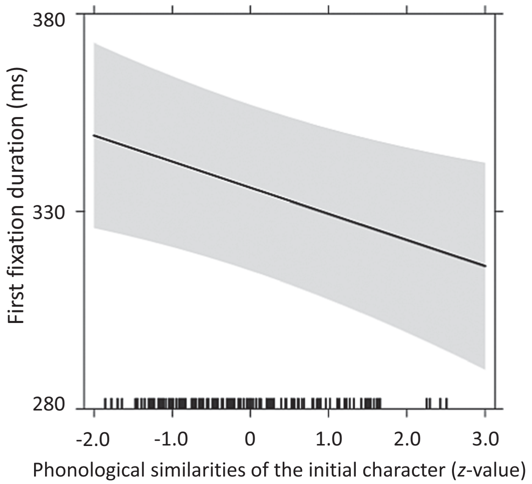



**Figure 4:** The effect of phonological similarities of the initial character on L1-Chinese reading during first fixation period. The grey shaded area represents a 95% confidence interval, and the rugs on the x-axis represent the marginal distribution of the predictor.

### 3.3 Discussion

Participants processed cognates with a higher degree of orthographic and phonological similarity at the character level more rapidly, suggesting that both types of form overlap facilitated cognate reading in L1-Chinese. Although some studies have shown that phonological information may not contribute to lexical access, by analyzing the time course of word processing more precisely, this study successfully detected phonology's facilitation effect on lexical access in early processing (i.e., first fixation duration). However, there were no significant effects of orthographic similarities during this timeframe. Because it was a visual word recognition task, both phonological and semantic activation should be guided by visual features. It is reasonable to assume that orthographic similarity was associated with cognate processing, but somehow, we failed to capture it. This finding further supported our assumption that the inhibitory effect found during the first fixation of L2-Japanese cognate reading was not simply caused by the parafoveal effect, because if this was the case, we should have observed a similar effect in L1-Chinese cognate reading.

## 4 General discussion

The present study attempted to address two related questions: (1) whether effects of orthographic and phonological similarities emerge at the sub-lexical (i.e., character) level when bilinguals process cognates in L2-Japanese and L1-Chinese and (2) whether the effects on L2 reading are similar to those on L1. For both L2-Japanese and L1-Chinese, we conducted lexical decision tasks with eye-tracking. In line with previous studies (e.g., Duyck et al. 2007; Vanlangendonck et al. 2020; Woumans, Clauws, and Duyck 2021), our results suggested that both L1 and L2 activated automatically during cognate reading in a single context. Sub-lexical orthographic and phonological similarities affected both L2-Japanese and L1-Chinese reading; however, the effects differed.

In L2-Japanese cognate processing, the orthographic similarity of the initial characters hindered L2-Japanese reading in early processing (i.e., long first-fixation duration; Figure 2) but together with phonological similarity facilitated the late stage of processing (i.e., short RTs; Figure 1). Such effects of orthographic features have been reported in several eye-tracking studies, but only at the about-to-fixed region. More specifically, when reading trimorphemic compound words, the second and final characters with simple visual features located in the parafoveal region resulted in longer fixations (e.g., Miwa, Libben, and Ikemoto 2017). However, we found an effect of the first character only. One possibility is that there was competition between L2-Japanese and L1-Chinese at the character level. According to the BIA+ model, cognates with similar word forms are likely represented in L1 and L2 orthographic representations that are connected with an inhibitory link (e.g., Dijkstra et al. 2010). The visually presented word would activate all candidates that exhibit form overlap; thus, both the word in the target language (L2-Japanese) and its cognate pair (L1-Chinese) would be activated and compete. Considering that activation at character level is crucial for processing two-character compound words in Chinese and Japanese (e.g., Yan et al. 2006; Miwa, Libben, et al. 2014), the inhibitory link may also occur at the character level, which elicited an inhibitory effect from orthographic similarity in early processing.

Note that orthographic similarities turned into a facilitating effect, together with phonological overlap, contributed to the L2 cognate reading in late processing. Given that cognates have the same meaning in L2-Japanese and L1-Chinese, the word forms further activated shared semantic representations, which facilitated cognate processing (i.e., RTs decreased with increasing orthographic/phonological form overlap). This change in the direction of the effect further supported our hypothesis that the interference that occurred during the first fixation was due to the inhibitory link between L1 and L2.

However, we did not find any significant orthographic effects, but a phonological facilitation effect on L1-Chinese cognate reading in early processing (i.e., short first fixation durations; Figure 4). This result contradicts the view that word formation starts from orthographic features in visual word recognition tasks. This is surprising because for bilinguals who use logographic scripts, the activation of phonological information may not be required (e.g., Wong, Wu, and Chen 2014); rather, individuals rely on orthographic features in word formation (Xiong, Verdonschot, and Tamaoka 2020). This result may be because the activation of orthographic information at the character level occurs too rapidly for eye-tracking technology to capture. Particularly, all the stimuli are frequently used in Chinese and Japanese, facilitating lexical access for participants. While there is not yet consensus regarding phonological activation during L1-Chinese cognate processing, this study suggests that phonological information contributes to two-character compound word processing in L1-Chinese (and in L2-Japanese). The absence of this phonological effect in previous research may be due to the precision of the measurement tools used.

It is interesting to note that although the fixation mark was positioned in the middle of the words – that is, the distance from the first eye fixation to both characters was equal – only form overlaps of initial characters showed significant contributions to L2-Japanese and L1-Chinese cognate processing. Given that the first element may hold a perceptual advantage when reading from left to right (Kush et al. 2019), research has found the first constituent's effect on compound word reading in the early stages of processing in many languages (e.g., Hyönä and Pollatsek 1998; Kuperman et al. 2009). As for the second constituent, evidence from previous eye-tracking studies on compound word processing suggests that although later than the first character, the effect of the second element emerges during word processing (e.g., Miwa, Libben, et al. 2014). On the other hand, there is also evidence suggesting that frequent compound words are processed as single entities, resulting in the absence of character effects (Yan et al. 2006; Cui et al. 2021). It is apparently not applicable to our findings, as we uncovered effects of the initial character.

Future research must clarify whether the orthographic overlaps contribute to L1-Chinese cognate reading and whether the orthographic/phonological overlaps of final characters play a role in reading cognates in L1-Chinese and L2-Japanese. The use of ERPs could be helpful for investigating sub-lexical components during word processing, as they have better time resolution.[3] Future research should also consider the effect of the semantic transparency of compound words, as it has been shown that, besides frequency effects at the morphemic and lexical levels, semantic transparency also has a robust effect on compound word reading (Schmidtke, Van Dyke, and Kuperman 2020).

In sum, in line with previous studies on monolingual individuals (e.g., Miwa, Libben, et al. 2014), our results suggest that Chinese-Japanese bilinguals' processing of two-character compound words is also driven by processing at the character level. The effects of orthographic and phonological overlap on L2-Japanese and L1-Chinese differ. In the L2-Japanese task, orthographic similarity inhibited word processing in early processing, then along with phonological overlap facilitated cognate recognition during late-stage processing. Regarding L1-Chinese reading, sub-lexical phonological information facilitated cognate reading during very early processing, which disappeared in the late stage. Contrary to those who have claimed that phonological activation is not mandatory for Chinese processing (Wong, Wu, and Chen 2014; Chen et al. 2007), our results indicate that phonological activation at the sub-lexical level is crucial for Chinese cognate reading.

To our knowledge, the current study is the first to address the effects of orthographic and phonological information at the character level in both L2-Japanese and L1-Chinese cognate reading. We have demonstrated that cognate processing with logographic scripts is driven by sub-lexical form overlap, as the phonological information activates automatically even though it is not required for visual word recognition.

---

**3** Alternately, as the reviewers suggested, the questions might also be addressed by placing the fixation mark outside the stimulus (e.g., at the top center of the screen) and leaving a space between the two characters, so that the entire word does not fall within foveal vision and the fixations that locate at each character can be examined independently.

# References

Baayen, R. Harald. 2008. *Analyzing Linguistic Sata: A Practical Introduction to Atatistics Using R*. New York, NY: Cambridge University Press.

Bates, Douglas, Martin Mächler, Ben Bolker, & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48.

Chen, Hsin-Chin, Takashi Yamauchi, Katsuo Tamaoka, & Jyotsna Vaid. 2007. Homophonic and semantic priming of Japanese kanji words: A time course study. *Psychonomic Bulletin & Review* 14(1). 64–69.

Comesaña, Montserrat, Paulin Bertin, Helena Oliveira, Ana Paula Soares, Juan Andrés Hernández-Cabrera, Séverine Casalis. 2018. The impact of cognateness of word bases and suffixes on morpho-orthographic processing: A masked priming study with intermediate and high-proficiency Portuguese-English bilinguals. *PLoSONE* 13(3): e0193480.

Cop, Uschi, Nicolas Dirix, Eva Van Assche, Denis Drieghe, & Wouter Duyck. 2017. Reading a book in one or two languages? An eye movement study of cognate facilitation in L1 and L2 reading. *Bilingualism: Language and Cognition* 20(4). 747–769.

Cui, Lei, Jue Wang, Yingliang Zhang, Fengjiao Cong, Wenxin Zhang, & Jukka Hyönä. 2021. Compound word frequency modifies the effect of character frequency in reading Chinese. *Quarterly Journal of Experimental Psychology* 74(4). 610–633.

Dalmaijer, Edwin S., Sebastiaan Mathôt, & Stefan Van der Stigchel. 2014. PyGaze: An open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments. *Behavior Research Methods* 46(4). 913–921.

Degani, Tamar, Anat Prior, & Walaa Hajajra. 2018. Cross-language semantic influences in different script bilinguals. *Bilingualism: Language and Cognition* 21(4). 782–804.

Dijkstra, Ton, Jonathan Grainger, & Walter J. B. van Heuven. 1999. Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language* 41(4). 496–518.

Dijkstra, Ton & Walter J. B. van Heuven. 1998. The BIA model and bilingual word recognition. In Jonathan Grainger & Arthur M. Jacobs (eds.), *Localist Connectionist Approaches to Human Cognition*, 189–225. Mahwah, NJ: Lawrence Erlbaum Associates.

Dijkstra, Ton & Walter J. B. van Heuven. 2002. The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition* 5(3). 175–197.

Dijkstra, Ton, Koji Miwa, Bianca Brummelhuis, Maya Sappelli, & Harald Baayen. 2010. How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language* 62(3). 284–301.

Duyck, Wouter. 2005. Translation and associative priming with cross-lingual pseudohomophones: Evidence for nonselective phonological cctivation in bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31(6). 1340–1359.

Duyck, Wouter, Eva Van Assche, Denis Drieghe, & Robert J. Hartsuiker. 2007. Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access. *Journal of Experimental Psychology-Learning Memory and Cognition* 33(4). 663–679.

Gollan, Tamar H., Rosa I. Montoya, Cynthia Cera, & Tiffany C. Sandoval. 2008. More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language* 58(3). 787–814.

Grosjean, François. 2010. *Bilingual: Life and Reality*. Cambridge, MA: Harvard University Press.

Hoshino, Noriko & Judith F. Kroll. 2008. Cognate effects in picture naming: Does cross-language activation survive a change of script? *Cognition* 106(1). 501–511.

Hsieh, Ming-Che, Hyeonjeong Jeong, Motoaki Sugiura, & Ryuta Kawashima. 2021. Neural evidence of language membership control in bilingual word recognition: An fMRI study of cognate processing in Chinese–Japanese bilinguals. *Frontiers in Psychology* 12. 643211.

Hyönä, Jukka & Raymond Bertram. 2004. Do frequency characteristics of nonfixated words influence the processing of fixated words during reading? *European Journal of Cognitive Psychology* 16(1/2). 104–127.

Hyönä, Jukka & Alexander Pollatsek. 1998. Reading Finnish compound words: Eye fixations are affected by component morphemes. *Journal of Experimental Psychology: Human Perception and Performance* 24(6). 1612–1627.

Kuperman, Victor, Robert Schreuder, Raymond Bertram, & R. Harald Baayen. 2009. Reading polymorphemic Dutch compounds: Toward a multiple route model of lexical processing. *Journal of Experimental Psychology: Human Perception and Performance* 35(3). 876–895.

Kush, Dave, Brian Dillon, Ragnhild Eik, & Adrian Staub. 2019. Processing of Norwegian complex verbs: Evidence for early decomposition. *Memory & Cognition* 47(2). 335–350.

Libben, Maya R. & Debra A. Titone. 2009. Bilingual lexical access in context: Evidence from eye movements during. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 35(2). 381–390.

Miwa, Koji, Ton Dijkstra, Patrick Bolger, & R. Harald Baayen. 2014. Reading English with Japanese in mind: Effects of frequency, phonology, and meaning in different-script bilinguals. *Bilingualism: Language and Cognition* 17(3). 445–463.

Miwa, Koji, Gary Libben, Ton Dijkstra, & Harald Baayen. 2014. The time-course of lexical activation in Japanese morphographic word recognition: Evidence for a character-driven processing model. *Quarterly Journal of Experimental Psychology* 67(1). 79–113.

Miwa, Koji, Gary Libben, & Yu Ikemoto. 2017. Visual trimorphemic compound recognition in a morphographic script. *Language, Cognition and Neuroscience* 32(1). 1–20.

Mulder, Kimberley, Ton Dijkstra, & R. Harald Baayen. 2015. Cross-language activation of morphological relatives in cognates: The role of orthographic overlap and task-related processing. *Frontiers in Human Neuroscience* 9. 1–18.

Nakayama, Mariko, Christopher R. Sears, Yasushi Hino, & Stephen J. Lupker. 2013. Masked translation priming with Japanese–English bilinguals: Interactions between cognate status, target frequency and L2 proficiency. *Journal of Cognitive Psychology* 25(8). 949–981.

Park, Sunju, Kexin Xiong, & Katsuo Tamaoka. 2014. Dookei nizi kanzigo no hinsisei ni kansuru nikkantyuu deeta beesu [A database of grammatical categories for orthographically-similar two-kanji compound words among Japanese, Korean and Chinese]. *Kotoba no Kagaku* 27. 53–111.

Peeters, David, Ton Dijkstra, & Jonathan Grainger. 2013. The representation and processing of identical cognates by late bilinguals: RT and ERP effects. *Journal of Memory and Language* 68(4). 315–332.

Peleg, Orna, Tamar Degani, Muna Raziq, & Nur Taha. 2020. Cross-lingual phonological effects in different-script bilingual visual-word recognition. *Second Language Research* 36(4). 653–690.

Schmidtke, Daniel, Julie A. Van Dyke, & Victor Kuperman. 2021. CompLex: An eye-movement database of compound word reading in English. *Behavior Research Methods* 53(1). 59–77.

Tamaoka, Katsuo. 2005. The effect of morphemic homophony on the processing of Japanese two-kanji compound words. *Reading and Writing* 18(4). 281–302.

Tamaoka, Katsuo, Shogo Makioka, Sander Sanders, & Rinus G. Verdonschot. 2017. www.kanjidatabase.com: A new interactive online database for psychological and linguistic research on Japanese kanji and their compound words. *Psychological Research* 81(3). 696–708.

Van Assche, Eva, Denis Drieghe, Wouter Duyck, Marijke Welvaert, & Robert J. Hartsuiker. 2011. The influence of semantic constraints on bilingual word recognition during sentence reading. *Journal of Memory and Language* 64(1). 88–107.

Vanlangendonck, Flora, David Peeters, Shirley-Ann Rueschemeyer, & Ton Dijkstra. 2020. Mixing the stimulus list in bilingual lexical decision turns cognate facilitation effects into mirrored inhibition effects. *Bilingualism: Language and Cognition* 23(4). 836–844.

Wong, Andus, Yan Wu, & Hsuanchih Chen. 2014. Limited role of phonology in reading Chinese two-character compounds: Evidence from an ERP study. *Neuroscience* 256. 342–351.

Woumans, Evy, Robin Clauws, & Wouter Duyck. 2021. Hands down: Cognate effects persist during written word production. *Frontiers in Psychology* 12. 647362.

Xiong, Kexin. 2018. *Cognitive Processing of Orthographically and Semantically Similar Two-Kanji Compound Words by Native Chinese Speakers Learning Japanese*. Nagoya: Nagoya University dissertation.

Xiong, Kexin, Rinus G. Verdonschot, & Katsuo Tamaoka. 2020. The time course of brain activity in reading identical cognates: An ERP study of Chinese-Japanese bilinguals. *Journal of Neurolinguistics* 55. 100911.

Xun, Endong, Gaoqi Rao, Xiaoyue Xiao, & Jiaojiao Zang. 2016. The construction of the BCC Corpus in the age of big data. *Corpus Linguistics* 3. 93–118.

Yan, Guoli, Hongjie Tian, Xuejun Bai, & Keith Rayner. 2006. The effect of word and character frequency on the eye movements of Chinese readers. *British Journal of Psychology* 97(2). 259–268.

Yokosawa, Kazuhiko & Michio Umeda. 1988. Processes in human kanji-word recognition. In *Proceedings of the 1988 IEEE International Conference on Systems, Man, and Cybernetics* 1, 377–380.

Yu, Shaoyun & Katsuo Tamaoka. 2015. Nikkantyuu dookei nizi kanzigo no hinsisei web-kensaku enzin [A web-accessible search engine for grammatical category of orthographically-similar two-kanji compound words between Japanese, Korean and Chinese]. *Kotoba no Kagaku* 29. 43–61.

Zhang, Juan, Chenggang Wu, Tiemin Zhou, & Yaxuan Meng. 2019. Cognate facilitation priming effect is modulated by writing system: Evidence from Chinese-English bilinguals. *International Journal of Bilingualism* 23(2). 553–566.

Jungho Kim

# Chapter 13
# Cortical neural activities related to processing Japanese scrambled sentences by Japanese L2 learners: An fMRI study

## 1 Introduction

In languages in which word order serves as the most critical clue in understanding the meanings of sentences, such as English, interchanging the subject and the object with each other completely alters the meaning of who did what to whom as in (1).

(1) a. John praised Mary.
    b. Mary praised John.

Unlike English, however, Japanese is a language that allows for relatively free word order. For example, in a Japanese sentence with a transitive verb, it is possible to swap the subject and the object in the sentence without changing the essential meaning it conveys, as shown in (2).

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| (2) | a. | Hanako | ga | Asuka | o | home | ta |
| | | Hanako | NOM | Asuka | ACC | praise | PAST |
| | | 'Hanako praised Asuka.' | | | | | |
| | b. | Asuka | o | Hanako | ga | home | ta |
| | | Asuka | ACC | Hanako | NOM | praise | PAST |
| | | 'Hanako praised Asuka.' | | | | | |

In Japanese generative linguistics, sentences with a Subject–Object–Verb (SOV) word order like (2a) are referred to as "canonical sentences," and those with an Object–Subject–Verb (OSV) word order like (2b) as "scrambled sentences." Scrambled sentences such as (2b) are generated derivatively when the object noun phrase

**Acknowledgments:** The author would like to thank two anonymous reviewers for their valuable comments on the manuscript and audiences at the 14th Japanese Language Association of Korea and the 12th Annual Meeting of the Organization for Human Brain Mapping for their helpful comments. Parts of the results in this chapter were published in Kim (2012). This study was supported in part by JSPS KAKENHI Grant Numbers JP19H05589, JP17K02754 and JP21K00536.


https://doi.org/10.1515/9783110778946-013

(in this case “*Asuka-o*”) is moved before the subject (“*Hanako-ga*”) in one way or another. This movement process is called “scrambling.” Scrambled sentences that are derived from such phrase movements basically represent the same meanings as their canonical sentence counterparts.

(3) a. [S Hanako-ga [VP Asuka-o home-ta]]
    b. [S Asuka-o$_i$ [S Hanako-ga [VP $t_i$ home-ta]]]

*scrambling*

In theoretical linguistics, it is assumed that when an object is scrambled to a position that precedes a subject, it leaves “a trace” in its original position and creates “a filler-gap dependency” (Saito 1985; Hoji 1985). Accordingly, compared with the canonical sentence (3a), which does not involve the syntactic processing of taking the object noun phrase at the beginning of the sentence and filling it in at the original trace position (*t*), the scrambled sentence (3b), which requires gap-filling parsing, demands more complicated syntactic processing.

(4) [S Asuka-o$_i$ [S Hanako-ga [VP $t_i$ home-ta]]]

*gap-filling parsing*

Chujo (1983) conducted an experiment with a sentence correctness decision task using verbs whose objects are human (animate nouns) (e.g., “to hire”) and verbs whose objects are not human (inanimate nouns) (e.g., “to close”), and reported that the reaction time in cases with the OSV word order was statistically significantly longer than that with the SOV word order. To determine primary factors behind the differences in the processing costs for comprehending Japanese SOV and OSV sentences (OSV sentences involve longer reaction time and higher error rates than their SOV counterparts), Tamaoka et al. (2005) focused on three possible factors (thematic roles, case particles, and grammatical functions) and conducted five experiments (active sentences with transitive verbs, active sentences with ditransitive verbs, passive sentences with transitive verbs, potential sentences, and causative sentences). Consequently, the following conclusion was made: the primary source of the scrambling effects was found in grammatical functions such as the subject and the object. In addition to experiments just mentioned, Mazuka et al. (2002) conducted an experiment measuring eye movement. In this experiment, the participants were asked to read SOV and OSV sentences, and their eye movements were compared between when they read the first noun phrases in SOV sentences and when they read the second noun phrases in the OSV sentences. Mazuka et al.

(2002) reported that they recorded statistically significantly longer gaze times and the participants exhibited more regressive eye movements when they read the SUBJECT noun phrase in O**S**V than in **S**OV sentences. Furthermore, multiple studies have attempted to clarify the syntactic construction differences between the two-word order structures in Japanese transitive verb sentences by observing priming effects. "Priming effects" denotes a phenomenon in which words that are seen or heard several times are easier to remember than those that are seen or heard only once. Miyamoto and Takahashi (2002) conducted an experiment with a probe-word recognition task and reported that priming effects were more remarkable in the case of the OSV than the SOV structure. They argued that this was because the subjects needed to verify the presented probe ("mondai" in the example mentioned below) twice, first at the beginning of the sentence and second at the trace position (*t*) near the end of the sentence with the OSV structure, in contrast to the SOV structure without any traces.

(5) a. Canonical condition

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Gakkoo | de | <u>mondai</u> | o | dashita | kooshi | *ga* | totemo |
| school | LOC | question | ACC | asked | lecturer | NOM | very |
| kashikoi | gakusei | *o* | mita. | | | | |
| smart | student | ACC | saw | | | | |

"The lecturer who asked the question at school saw the extremely smart student."

b. Scrambling condition

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Gakkoo | de | [<u>mondai</u> | o | dashita | kooshi]$_i$ | o | totemo |
| school | LOC | question | ACC | asked | lecturer | ACC | very |
| kashikoi | gakusei | ga | <gap$_i$> | mita. | | | |
| smart | student | NOM | | saw | | | |

"The extremely smart student saw the lecturer who asked the question at school."

(Partly revised from Miyamoto and Takahashi 2002)

Behavioral experiments have demonstrated that such scrambling effects as noted in the case of Japanese transitive verb sentences are present not only in native Japanese speakers but also in native Korean and Chinese speakers who are learning Japanese as a second language at the advanced level (Tamaoka 2005; Kim and Koizumi 2007).

Hagiwara and Caplan (1990) conducted an experiment involving aphasic patients and reported that the correct answer rate with the OSV structure only reached 64%, which was close to the chance level. This can be interpreted as a result of difficulty in OSV syntactic processing (comprehension) arising from certain damage incurred in Broca's area in the brain. In other words, compared with healthy individuals who can perform gap-filling parsing with relative ease, aphasic patients might find it difficult to adequately process the relationship between fillers and gaps in scrambled sentences, which explains significantly lower correct answer rates.

In a series of studies using functional magnetic resonance imaging (fMRI) conducted by Kinno et al. (2008) and Kim et al. (2009), higher cortical activation was observed in the left frontal cortex when subjects were processing scrambled sentences compared with their canonical sentences.

As described above, studies on the comprehension of the OSV sentence structure generated from the scrambling of Japanese active transitive sentences have been pursued in various fields, including theoretical linguistics, psycholinguistics, and cognitive neuroscience. However, most of those studies were conducted with native Japanese speakers. Currently, there have been only a few cognitive neuroscientific studies on sentence processing by people who are learning Japanese as a second language (Mueller 2005; Jeong et al. 2007; Kim 2012). This chapter reports on cortical neural activities related to processing scrambled Japanese sentences by Japanese L2 learners using functional magnetic resonance imaging.

# 2 Target sentences and predictions

## 2.1 Target sentences

In this study, two syntactic types of Japanese transitive verb sentences were used as target sentences: canonical word order (SOV) and scrambled word order (OSV).

(6) a. Canonical condition (Subject–Object–Verb)

| | | | | | |
|---|---|---|---|---|---|
| Gakusei | ga | syukudai | o | wasure | ta |
| Gakusei | NOM | syukudai | ACC | forget | PAST |

"The student forgot the homework."

b. Scrambled condition (Object–Subject–Verb)

| | | | | | |
|---|---|---|---|---|---|
| Syukudai | o | gakusei | ga | wasure | ta |
| Syukudai | ACC | gakusei | NOM | wasure | PAST |

"The student forgot the homework."

As mentioned above, Kinno et al. (2008) and Kim et al. (2009) indicated that native Japanese speakers display higher cortical activation in the left frontal cortex when processing scrambled sentences than when processing canonical sentences.

## 2.2 Predictions

Similar to Japanese, Korean also has SOV word order as the basic sentence structure. It is also possible to move the object to the beginning of the sentence to construct a sentence with OSV word order, as shown in (7). In contrast, Chinese, like English, relies on word order as the most critical clue in understanding the meaning of sentences. Nevertheless, besides the basic SVO word order, Chinese allows for OSV word order in topicalized sentence structures, as shown in (8) (Ernst and Wang 1995).

(7) a. Canonical sentence (Subject–Object–Verb)

| | | | | | |
|---|---|---|---|---|---|
| Chelsoo | ga | Tosilak | ul | mek | ess-ta |
| Chelsoo | NOM | box lunch | ACC | eat | PAST-DEC |

"Chelswu ate the box lunch."

b. Scrambled sentence (Object–Subject–Verb)

| | | | | | |
|---|---|---|---|---|---|
| Tosilak | ul | Chelsoo | ga | mek | ess-ta |
| box lunch | ACC | Chelsoo | NOM | eat | PAST-DEC |

(8) a. SVO

| | | |
|---|---|---|
| Wo | he | Jiu. |
| I | drink | liquor |

"I drink liquor."

b. OSV

| | | |
|---|---|---|
| Jiu, | wo | he. |
| liquor | I | drink |

"Liquor, I drink."

(Ernst and Wang 1995: 241)

**Native Korean speaker group:** Similar to Japanese, Korean allows for relatively free word order. As shown in (7b), the object is permitted to precede the subject in a sentence. Therefore, if a native Korean speaker is an advanced learner of the Japanese language, they are likely to devise similar strategies to native Japanese speakers. Specifically, they are expected to exhibit scrambling effects, where reaction time is longer and error rates are higher in cases of OSV than in cases of SOV. As in the case of the results in native Japanese speakers reported by Kinno et al. (2008) and Kim et al. (2009), higher cortical activation is expected to be observed in the Broca's area in a direct comparison between OSV and SOV. This is attributable to the fact that scrambled word order sentence is considered to have a more complex syntactic structure than its canonical counterpart in theoretical linguistics and psycholinguistics.

**Native Chinese speaker group:** In a behavioral experiment involving native Chinese speakers learning Japanese at high proficiency, scrambling effects were reportedly observed in their reaction times and error rates for Japanese transitive verb sentences (Tamaoka 2005). Furthermore, in a neuroimaging study focusing on Chinese, L.IFG reportedly had a strong relationship to syntactic processing (Wang et al. 2008; Bulut et al. 2017). These studies support that L.IFG is the critical region to handle syntactic complexity irrespective of the influence of crosslinguistic typological distance. Considering the above-mentioned findings and the fact that the Chinese language also displays the phenomenon of the object preceding the subject in a sentence (topicalization), similar results are predicted for the native Chinese speaker group.

# 3 Methods

## 3.1 Participants

The fMRI experiment involved 13 international exchange students studying at Tohoku University (7 and 6 native Korean and Chinese speakers, respectively). The two groups of participants included those who speak Korean and Chinese as the only mother tongue, respectively. The participants were all non-disabled individuals with a mean age of 28.3 years (20–35 years; SD = 5.17; 10 females). The mean age of the native Korean speakers was 30.6 years (20–35 years; SD = 5.41; 5 females), whereas that of the native Chinese speakers was 25.7 years (22–31 years; SD = 3.67; 5 females). No statistically significant intergroup differences were noted in terms of the duration of staying in Japan and the duration of exposure to L2. Table 1 summarizes the duration of the students staying in Japan and that of the exposure to Japanese.

**Table 1:** Duration of living in Japan (months) and duration of exposure to Japanese.

| | **Duration of staying in Japan (months)** | **Duration of exposure to Japanese (months)** |
|---|---|---|
| KOR ($n$ = 7) | 53.6 (21.9) | 92.6 (39.7) |
| CHI ($n$ = 6) | 42.0 (17.0) | 92.0 (41.4) |
| | $F$(1,11) = 1.1, $p$ = .32 | $F$(1,11) = 0.001, $p$ = .98 |

“KOR” and “CHI” represent the native Korean and Chinese speakers, respectively.

The participants in the experiment were recruited from individuals who were certified at level N1 of the Japanese Language Proficiency Test (JLPT). The JLPT is a test organized by the Japan Foundation and Japan Educational Exchanges and Services. To pass N1, the applicant needs to acquire high-level knowledge of Japanese grammar, kanji (approx. 2000 characters), and vocabulary (approx. 10,000 words), and is expected to understand Japanese used in a variety of circumstances. To reach this level, the applicant needs approximately 900–1200 hours of study. No statistically significant differences could be observed in the JLPT scores between the groups. The mean starting age of learning Japanese as a second language was 22.6 years (16–28 years) and 19.3 years (13–22 years) in the Korean and the Chinese group, respectively (Table 2); both of these time periods were beyond the so-called "critical period of language acquisition." Therefore, all the participants were advanced learners of Japanese who started learning beyond the critical period, known as LAHP (Late Acquisition High Proficiency).

The participants were examined for their dominant hands using the Edinburgh Handedness Inventory (Oldfield 1971), which revealed that all of them were right-handed (LQ = 90.4). Prior to the experiment, the experimental content and the fMRI apparatus safety were fully explained to the participants in accordance with the guidelines stipulated by Tohoku University. Written informed consent was obtained from each participant.

**Table 2:** Age of starting Japanese learning and JLPT score.

| | AoA (SD) | JLPT score (SD) |
|---|---|---|
| KOR (n = 7) | 22.6 (5.0) | 334 (26.4) |
| CHI (n = 6) | 19.3 (3.4) | 332 (19.8) |
| | $F$ (1,11) = 1.8, $p$ = .21 | $F$ (1,11) = 0.01, $p$ = .92 |

AoA stands for the age of acquisition L2, which is the age of starting Japanese learning. The JLPT changed its scoring system in 2010. All the participants had a JLPT score before the change of the system, where the maximum score was 400: Writing-Vocabulary (100 points); Listening (100 points); and Reading-Grammar (200 points). After the change in 2010, the full marks for N1 became 180: Vocabulary-Grammar (60 points); Reading (60 points); and Listening (60 points).

## 3.2 Materials and procedures

A session consisted of four tasks as follows: canonical sentence (CS) task, scrambled sentence (SS) task, word list (WL) task, and rest (R) task (see Table 3, Figure 1).

Each task comprised 42 sentences or 42 word lists. Among the 42 sentences, there were 28 semantically plausible ("correct") transitive sentences and 14 semantically implausible ("incorrect") transitive sentences. Participants were instructed to judge whether a given sentence was a semantically plausible or implausible transitive sentence and to press the "correct" button with their right index finger or the "incorrect" button with their right middle finger as accurately and swiftly as possible when a "+" sign was displayed following the appearance of the third verb. The third verb and the ensuing "+" sign were displayed for the duration of 2.6 seconds. If a participant pressed the button before the third verb appeared or after the 2.6 seconds elapsed, it was counted as "incorrect." The WL task contained 42 sequences of words. The participants were instructed to read lists consisting of three words and press the "correct" button with their right index finger when a list contained three different words and the "incorrect" button with their right middle finger when the same word reappeared at least once. The word lists used in the WL task were prepared using the same words applied under the CS and SS tasks.

**Table 3:** Samples of stimuli used in the CS, SS, WL, and R tasks.

| | | | |
|---|---|---|---|
| **(1) Canonical Sentence (CS) task** | | | |
| 28 correct SOV sentences | Gakusei ga | shukudai o | wasureta |
| | student-NOM | homework-ACC | forgot |
| | "The student forgot his homework." | | |
| 14 incorrect SOV sentences | Hanaya ga | soji o | tsutsunda |
| | florist-NOM | cleaning-ACC | wrapped |
| | "The florist wrapped the cleaning." | | |
| **(2) Scrambled Sentence (SS) task** | | | |
| 28 correct OSV sentences | Kaisha o | shujin ga | yasunda |
| | work-ACC | my husband-NOM | absent |
| | "My husband was absent from work." | | |
| 14 incorrect OSV sentences | Megane o | shinshi ga | tazuneta |
| | glasses-ACC | gentleman-NOM | visited |
| | "The gentleman visited the glasses." | | |
| **(3) Word List (WL) task** | | | |
| 28 correct word lists | Shacho ga | Gakusha ga | Doryo ga |
| | boss-NOM | scholar-NOM | colleague-NOM |
| 14 incorrect word lists | Migaita | Kudaita | Migaita |
| | brushed | spalled | brushed |
| **(4) Rest (R) task** | | | |
| | A fixation cross is displayed at the center of the screen. | | |

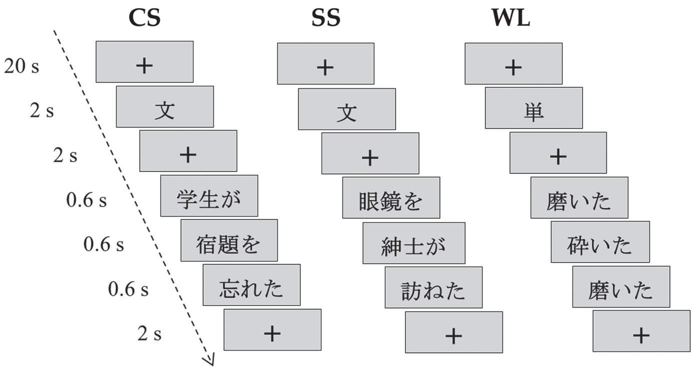



**Figure 1:** Design of the three tasks for the experiment.

In the R task block, participants were asked to look at a fixation cross and to rest for 20 s. At the beginning of each block, either ("文", sentence task) or ("単", word task) was presented to let the participants know the type of task that follows in the blocks. A period (".") was not displayed after the verb in the sentence tasks (CS and SS). However, sentence tasks were comprised of three phrases; and it was carefully explained to participants that sentences would be completed by the verb phrase which appeared in the third position. Furthermore, participants' understanding of this was confirmed via the practice task.

Words used in the JLPT in the past 10 years (1990–2000) were categorized by mora count. The stimulus sentences were selected from this word list in descending order of frequency. The data did not include the JLPT questions from 1998 as they were unavailable. The vocabulary used in N3 and N4 tests was employed in the stimulus sentences, with its choice priority given to N4. For the fMRI experiment, a blocked design was applied (Figure 2).

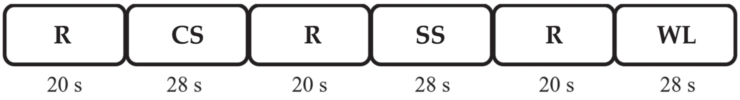



**Figure 2:** Time course by blocked design.
R, Rest task; CS, Canonical sentence task; SS, Scrambled sentence task; WL, Word list task.

In each block, the task title (either "文" (sentence task) or "単" (word list task)) was displayed for 2 seconds following a rest period (20 seconds) to inform the participants of whether the presented task would be a sentence or a word task. Each phrase was displayed at the center of the screen for 0.5 seconds. To let the participants recognize gaps between the phrases and/or words, intervals of 0.1 seconds were intercalated. The order of the task presentation was counter-balanced among the participants. Each task was presented to each participant seven times, with each block consisting of either six sentences or word lists. Moreover, to prevent the participant being distracted from the test in the MRI apparatus by factors such as nervousness, a block for practice (four sentences) that would not be included in the data analysis was added at the beginning of each session. A sentence-plausibility judgment task was administered using E-prime (Psychology Software Tools, Inc.). Prior to the main experiment, the subjects were requested to take practice tasks designed based on the same principle as the actual tasks. Practice tasks consisted of 4 CS sentences, 4 SS sentences, and 4 word lists. When a participant's score was low, they were asked to do the practice tasks twice.

## 3.3 fMRI data acquisition

Data recordings were obtained by fMRI using a 1.5 Tesla Siemens Symphony scanner (Siemens, Erlangen, Germany) at Tohoku University, under the preliminarily determined condition in Echo Planar Imaging (TR: 2000 ms; TE: 50 ms; FOV: 192 mm; slices: 22; thickness: 6 mm; data matrix: 64×64 voxels; flip angle; 90°). The twelve initial scans were dummy scans used to calibrate the state of magnetization and were excluded from the analysis. After the attainment of functional imaging, anatomical T1-weighted MDEFT images (thickness: 1 mm; FOV: 256 mm; data matrix: 192 × 224; TR: 1900 ms; TE: 3.93 ms) were also acquired from all participants.

## 3.4 fMRI data analysis

The fMRI data was analyzed through the same procedures as described by Kim et al. (2009). Pre-processing included correction for slice-time differences and spatial alignment to the first volume of the image series to adjust for head movements during the course of the experiment. Subsequently, the realigned datasets were spatially normalized to the Montreal Neurological Institute (MNI) standard brain template, and then performed smoothing of images using a 12 mm Gaussian filter. An analysis of the tasks for each participant was conducted at the first statistical stage and group statistical analysis at the second stage. Moreover, to eliminate fluctuations in brain activity owing to the difficulty factors of the questions (complexity of the syntactic structure) during the statistical processing of each participant, an analysis of covariance (ANCOVA) was performed using the reading time for all three tasks (CS, SS, and WL) and the error rate by each participant. Regarding groups of native speakers of Korean and Chinese groups, statistical inferences were made at the voxel level threshold of $p < .001$ (uncorrected) with a cluster extent threshold of $k > 20$ voxels.

# 4 Results

## 4.1 Behavioral data

The elapsed time from the presentation of a verb or the third word to when the button was pressed was measured. In each participant, values that deviated beyond ± 2.5 SD from the mean of each task were replaced with boundary values (mean ± 2.5 × SD). This data editing process of reaction times has been commonly used in statistical analysis (e.g., Tamaoka et al. 2005; Godfroid 2019). A series of one-way analyses of variance (ANOVA) with repeated measures in the three tasks (CS, SS, and WL) were conducted on reaction times and error rates. Table 4 summarizes the means and SDs of the reaction times and error rates by task in the Korean and Chinese groups.

**Table 4:** Reaction times (ms) and error rates (%) in the Korean and Chinese groups.

| | **Reaction time (ms)** | **Error rate (%)** |
|---|---|---|
| KOR ($n$ = 7) | | |
| CS | 1272 (170.6) | 7.1 (3.6) |
| SS | 1315 (152.3) | 11.2 (6.1) |
| WL | 953 (94.8) | 4.4 (6.4) |
| CS vs. SS vs. WL | $F$ (2,12) = 20.68, $p < .01$** | $F$ (2,12) = 3.16, $p = .08$ *n.s.* |
| CS vs. SS | $F$ (1,6) = 5.39, $p = .06$ *n.s.* | $F$ (1,6) = 4.85, $p = .07$ *n.s.* |
| CHI ($n$ = 6) | | |
| CS | 1145 (106.1) | 6.0 (3.8) |
| SS | 1259 (125.2) | 14.7 (5.7) |
| WL | 928 (93.9) | 3.6 (3.6) |
| CS vs. SS vs. WL | $F$ (2,10) = 100.47, $p < .01$** | $F$ (2,10) = 18.52, $p < .01$** |
| CS vs. SS | $F$ (1,5) = 15.71, $p < .01$** | $F$ (1,5) = 15.92, $p < .05$* |

KOR, Korean native speaker group; CHI, Chinese native speaker group; CS, Canonical sentence; SS, Scrambled sentence; WL, Word list; *n.s.*, no significant difference. * $p < .05$, ** $p < .01$

The Korean group revealed that the main effects were significant in the reaction time between the CS, SS, and WL tasks. However, no significant differences were observed between the CS and SS tasks. In the error rate, the main effects between the CS, SS, and WL tasks were not significant. The Chinese group revealed that the main effects between the CS, SS, and WL tasks were significant in the reaction time and error rate. The comparison between the CS and SS task showed significant differences of 1% and 5% in the reaction time and error rate, respectively. There was no significant difference between the two groups in terms of reaction times (CS: $F(1,11) = 2.113$, $p = 0.17$; SS: $F(1,11) = 0.434$, $p = 0.52$; WL: $F(1,11) = 0.2$, $p = 0.66$) and error rates (CS: $F(1,11) = 0.282$, $p = 0.61$; SS: $F(1,11) = 0.934$, $p = 0.35$; WL: $F(1,11) = 0.07$, $p = 0.79$) in each task.

## 4.2 Imaging data

To discover what effect typological differences such as word order would have on sentence processing of Japanese transitive verb sentences, direct comparisons were made within each group of [SS – CS]. The results in both groups of native Chinese speakers and native Korean speakers were compared, and activation of L.IFG was observed in both. However, subtracting [CS – SS] showed no activation in the relevant brain region (Figure 3, Table 5).

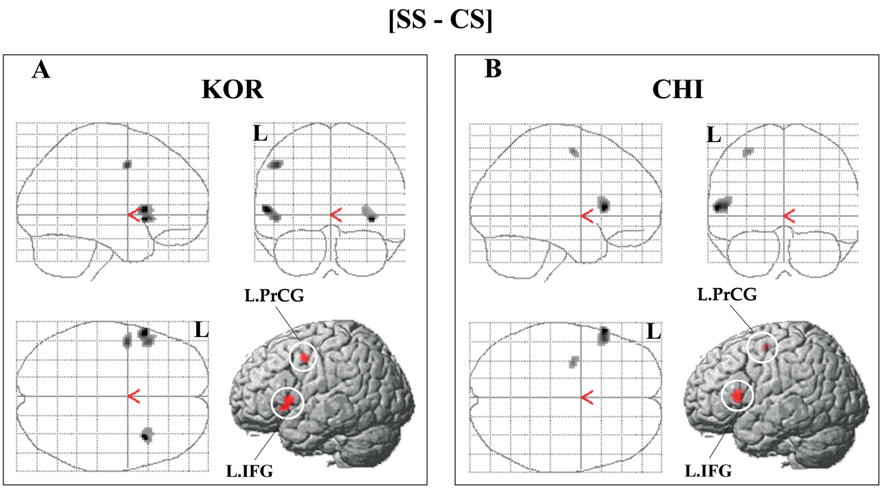



**Figure 3:** Activated brain regions observed in the [SS – CS] direct comparison.
A: Regions identified by [SS – CS] in Korean native speaker group. B: Regions identified by [SS – CS] in Chinese native speaker group. Statistical inferences were made at the voxel level threshold of $p < .001$ (uncorrected), with a cluster extent threshold of 20 voxels. L, left hemisphere; KOR, Korean native speaker group; CHI, Chinese native speaker group; L.IFG, left inferior frontal gyrus; L.PrCG, left precentral gyrus.

**Table 5:** Functional Imaging Results.

| Brain regions | BA | Side | *t* | *x* | *y* | *z* |
|---|---|---|---|---|---|---|
| [CS – SS] | | | No significant activation | | | |
| KOR | | | | | | |
| CHI | | | | | | |
| [SS – CS] | | | | | | |
| KOR | | | | | | |
| Insula | 13 | R | 11.81 | 38 | 16 | −4 |
| Insula | 13 | R | 6.41 | 32 | 16 | 2 |
| IFG | 44/45 | L | 11.34 | −56 | 14 | 4 |
| IFG | 45/44 | L | 8.41 | −50 | 18 | −2 |
| PrCG | 6 | L | 9.15 | −48 | 0 | 46 |
| CHI | | | | | | |
| IFG | 45/44 | L | 18.77 | −60 | 20 | 8 |
| PrCG | 6 | L | 9.72 | −32 | −8 | 58 |

For each area, the coordinates (x, y, z) of the activation peak in MNI space and peak T-value are shown for Chinese native subject ($n$ = 6) and Korean native subject ($n$ = 7). BA, Brodmann's area; L, left hemisphere; R, right hemisphere; CS, canonical sentence; SS, scrambled sentence; IFG, inferior frontal gyrus; PrCG, precentral gyrus.

Figure 4 shows activated brain regions observed in the [CS – WL] and [SS – WL] comparisons for all subjects (KOR+CHI). Both word-orders activated similar regions, including Broca's and Wernicke's areas. These results were largely consistent with those of native Japanese speakers (Kim et al. 2009), indicating that most of the cognitive processes involved in the comprehension of canonical sentences are common with the comprehension of scrambled sentences.

More importantly, in the direct comparison of [SS – CS] a significant brain activity was detected in the inferior frontal gyrus (BA 44/45) (Figure 5, Table 6).

# 5 Discussion

Similar to Japanese, Korean is a language with relatively free word order. Moreover, both are agglutinative languages with grammatical functions (subject, object, etc.) determined by case particles. Meanwhile, Chinese is a language mainly spoken in China and Southeast Asian countries. Unlike Korean and Japanese, it is an isolating language, belonging to the Sino-Tibetan family. In Chinese, as in English, word order serves as the most critical clue in understanding the meaning of sentences.

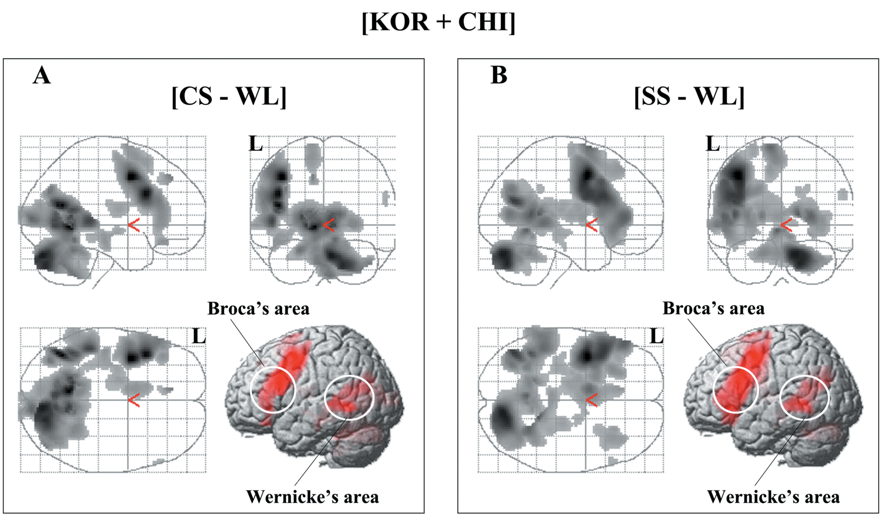



**Figure 4:** Activated brain regions in the [CS – WL] and [SS – WL] task comparisons for all subjects (KOR+CHI).
The threshold was set at a voxel-level correction of $p < .05$ FDR. CS, canonical sentence; SS, scrambled sentence; WL, word list.

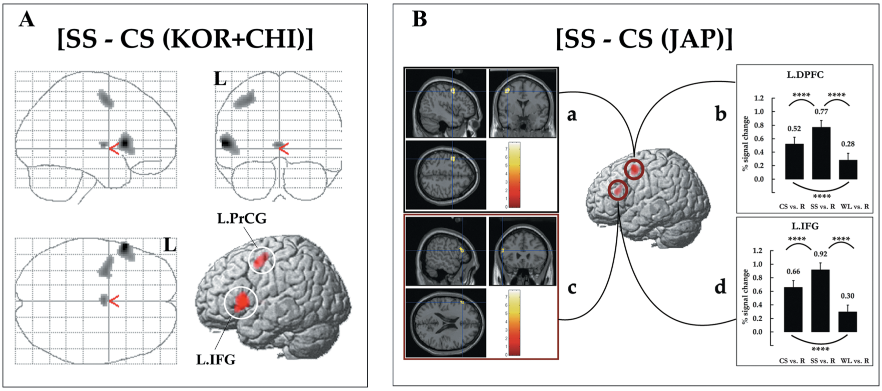



**Figure 5:** A: Comparisons of brain activation [SS – CS] (CHI+KOR). The threshold was set at a voxel-level correction of $p < .05$ FDR. B: Main effect of task [SS – CS] as reported by Kim et al. (2009).
KOR+CHI, all subjects of Chinese and Korean ($n = 13$); JAP, Japanese native subject; CS, canonical sentence; SS, scrambled sentence; L.PrCG, left precentral gyrus; L.DPFC, left dorsal prefrontal cortex; L.IFG, left inferior frontal gyrus.

**Table 6:** Cortical Regions Identified by [CS – SS] and [SS – CS] Tasks.

| | **BA** | **Side** | *t* | *x* | *y* | *z* |
|---|---|---|---|---|---|---|
| [CS – SS] | | | No significant activation | | | |
| [SS – CS] | | | | | | |
| IFG | 45/44 | L | 9.70 | −58 | 16 | 6 |
| IFG | 45/47 | L | 5.43 | −48 | 22 | −2 |
| PrCG | 6 | L | 6.28 | −32 | −4 | 52 |
| PrCG | 6 | L | 5.81 | −42 | 0 | 48 |
| Thalamus | | | 6.19 | 0 | −6 | 4 |

Results of all participants ($n = 13$). CS, canonical sentence; SS, scrambled sentence; BA, Brodmann's area; L, left hemisphere; IFG, inferior frontal gyrus; PrCG, precentral gyrus.

However, despite the syntactically typological distance between the languages, the reaction time and the error rate for scrambled sentences of the Chinese participants were almost the same as those of native Japanese speakers. The participants in this experiment were advanced learners with considerably high Japanese proficiency. In fact, the accuracy rate in the SS task exceeded 85% both in the Korean and Chinese groups. In the Korean group, no statistically significant differences were observed in the reaction time and error rate between the CS and SS task. In the Chinese group, it was found that CS was understood more quickly and accurately than SS (Table 4). These findings, as well as the results of preceding studies (Tamaoka et al. 2001; Tamaoka 2005), corroborate the fact that advanced learners of the Japanese language execute the same syntactic processing as native Japanese speakers, namely retaining the object at the beginning of a sentence until the appearance of the subject, then embedding the object subsequent to the subject (gap-filling parsing).

In the [CS – WL] and [SS – WL] comparisons for all subjects (KOR+CHI, Figure 4), the fMRI data analysis revealed activation in Broca's and Wernicke's areas, regions traditionally considered to be related to language processing. This suggests that most cognitive processes involved in the comprehension of scrambled sentences are also involved in the comprehension of canonical sentences. Finally, it is imperative to note that a direct comparison of [CS – SS] and [SS – CS] was performed to determine the brain regions that are involved in the processing of scrambled sentences. The former comparison showed no activation in any region, whereas the latter revealed activations in the left precentral gyrus (L.PrCG), which overlaps with the left dorsal prefrontal cortex (L.DPFC) in the previous study (Kim et al. 2009). The inferior frontal gyrus, including Broca's area, is considered to be a region deeply connected with language comprehension, according to previous studies (Caplan et al. 1998; Dapretto and Bookheimer 1999; Fiebach et al. 2004; Friederici 2002; Musso et al. 2003). The results of this study were compatible with those regarding native Japanese speakers (Kim et al. 2009; Koizumi et al. 2012). These findings all substantiate the theoretical linguistic and psycholinguistic views that a scrambled sentence is syntactically more complex than its canonical counterpart in that the former contains a filler-gap dependency absent in the latter. Based on the results of this study and Kim et al. (2009), the following interpretation patterns could be drawn for the intracerebral processing mechanism for Japanese transitive verb sentences by L1 and L2 speakers:

**Interpretation 1:** In the [SS – CS] direct comparison that was made to identify brain regions inherently associated with processing OSV sentences, the L.IFG (a region related to syntactic processing) was activated in the advanced learners of the Japanese language, similar to native Japanese speakers (Kim et al. 2009). This supports the argument of Musso et al. (2003: 778) that the inferior frontal gyrus, including Broca's area, is a specific region deeply connected with syntactic processing irrespective of the AoA or mother tongue of the subject.

**Interpretation 2:** Reports indicate that stronger activation is detected in the posterior left inferior frontal gyrus (pLIFG) when the syntactic structure of a presented sentence is more complex (Koizumi et al. 2012; Fiebach et al. 2004). For instance, in an fMRI experiment using Japanese ditransitive sentences conducted by Koizumi et al. (2012), whereas the subjects showed activation in the anterior left inferior frontal gyrus (aLIFG) with short-scrambling sentences, they showed activation in the pLIFG with middle-scrambling sentences with longer movement distance. Moreover, Fiebach et al. (2004) conducted an fMRI experiment using German scrambling sentences with three complexity levels and reported increased activation in the inferior portion of Broca's area. The results of this study also suggest that L1 and high proficiency L2 speakers of the Japanese language might rely on different intracerebral mechanisms for processing Japanese scrambled sentences.

# 6 Conclusion

First, despite the typological differences between the languages (e.g., SOV vs. SVO), the direct comparison of [SS – CS] in Korean and Chinese groups revealed cortical activation in the L.IFG. This result supported the proposal by Musso et al. (2003) that the Broca's area is a syntactically modulated region both in L1 and L2. Second, the direct [SS – CS] comparison for all subjects (KOR+CHI) displayed the posterior part of the L.IFG (pLIFG) was activated in Broca's area. This result supported the fact that pLIFG is activated when the syntactic structure of a presented sentence is more complex. Therefore, the L2 speakers (with high levels of Japanese proficiency) might have encountered greater loads while processing scrambled sentences than native Japanese speakers.

# References


Bulut, Talat, Yi-Hui Hung, Ovid Tzeng, & Denise H. Wu. 2017. Neural correlates of processing sentences and compound words in Chinese. *PLOS ONE* 12(12). e0188526.

Caplan, David, Nathaniel Alpert, & Gloria Waters. 1998. Effects of syntactic structure and propositional number on patterns of regional cerebral blood flow. *Journal of Cognitive Neuroscience* 10(4). 541–552.

Chujo, Kazumitsu. 1983. Nihongo tanbun-no rikai katei – Bunrikai sutoratezii no sougo kankei [The interrelationships among strategies for sentence comprehension]. *Japanese Journal of Psychology* 54. 250–256.

Dapretto, Mirella & Susan Y. Bookheimer. 1999. Form and content: Dissociating syntax and semantics in sentence comprehension. *Neuron* 24. 427–432.

Ernst, Thomas & Chengchi Wang. 1995. Object preposing in Mandarin Chinese. *Journal of East Asian Linguistics* 4. 235–260.

Fiebach, Christian J., Matthias Schlesewsky, Ina D. Bornkessel, & Angela D. Friederici. 2004. Distinct neural correlates of legal and illegal word-order variations in German: How can fMRI inform cognitive models of sentence processing. In Manuel Carreiras & Charles Clifton Jr. (eds.), *The On-Line Study of Sentence Comprehension: Eyetracking, ERPs and Beyond*, 357–370. Hove: Psychology Press.

Friederici, Angela D. 2002. Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences* 6. 78–84.

Godfroid, Aline. 2019. *Eye Tracking in Second Language Acquisition and Bilingualism: A Research Synthesis and Methodological Guide*. New York: Routledge.

Hagiwara, Hiroko & David Caplan. 1990. Syntactic comprehension in Japanese aphasics: Effects of category and thematic role order. *Brain and Language* 38. 159–170.

Hoji, Hajime. 1985. *Logical Form Constraints and Configurational Structures in Japanese*. Seattle, WA: University of Washington dissertation.

Jeong, Hyeonjeong, Motoaki Sugiura, Yuko Sassa, Tomoki Haji, Nobuo Usui, Masato Taira, Kaoru Horie, Shigeru Sato, & Ryuta Kawashima. 2007. Effect of syntactic similarity on cortical activation during second language processing: A comparison of English and Japanese among native Korean trilinguals. *Human Brain Mapping* 28(3). 194–204.

Kim, Jungho. 2012. Kankokuzin nihongo zyokyusya-to bogowasya-no bunsyori mekanizumu-ha onazinanoka – Kakimagozyunbun-o reitosite [Parsing Strategies in L1(Japanese native speaker) and L2(Korean advanced learners of Japanese) Sentence Processing: A case study of Japanese scrambled sentences]. *Journal of Japanese Language and Literature* 82(1). 189–209.

Kim, Jungho, Masatoshi Koizumi, Naho Ikuta, Yuichiro Fukumitsu, Naoki Kimura, Kazuki Iwata, Jobu Watanabe, Satoru Yokoyama, Shigeru Sato, Kaoru Horie, & Ryuta Kawashima. 2009. Scrambling effects on the processing of Japanese sentences: An fMRI study. *Journal of Neurolinguistics* 22(1). 151–166.

Kim, Jungho & Masatoshi Koizumi. 2007. Nihongo L2 gakusyusyaniyoru kakimaze bunrikai-no mekanizumu – Nihongo zyokyusya (kankokuzin/tyugokuzin)-o taisyoni [Second Language Learners of Japanese: A case study on Korean and Chinese advanced learners]. *Journal of Japanese Language* 18. 95–111.

Kinno, Ryuta, Mitsuru Kawamura, Seiji Shioda, & Kuniyoshi L. Sakai. 2008. Neural correlates of noncanonical syntactic processing revealed by a picture-sentence matching task. *Human Brain Mapping* 29. 1015–1027.

Koizumi, Masatoshi, Jungho Kim, Naoki Kimura, Satoru Yokoyama, Shigeru Sato, Kaoru Horie, & Ryuta Kawashima. 2012. Left inferior frontal activations differentially modulated by scrambling in ditransitive sentences. *Open Medical Imaging Journal* 6. 70–79.

Mazuka, Reiko, Kenji Itoh, & Tadahisa Kondo. 2002. Costs of scrambling in Japanese sentence processing. In Mineharu Nakayama (ed.), *Sentence Processing in East Asian languages* 131–166. Stanford, CA: CSLI Publications.

Miyamoto, Edson T. & Shoichi Takahashi. 2002. Antecedent reactivation in the processing of scrambling in Japanese. In Tania Ionin, Heejeong Ko, & Andrew Nevins (eds.), *Proceedings of HUMIT 2001, MIT Working Papers in Linguistics* 43, 123–138. Cambridge, MA: MIT Working Papers in Linguistics.

Mueller, Jutta L, Anja Hahne, Yugo Fujii, & Angela D. Friederici. 2005. Native and Nonnative Speakers' Processing of a Miniature Version of Japanese as Revealed by ERPs. *Journal of Cognitive Neuroscience* 17(8). 1229–1244.

Musso, Mariacristina, Andrea Moro, Volkmar Glauche, Michel Rijntjes, Jürgen Reichenbach, Christian Büchel, & Cornelius Weiller. 2003. Broca's area and the language instinct. *Nature Neuroscience* 6. 774–781.

Oldfield, R. C. 1971. The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* 9. 97–113.

Saito, Mamoru. 1985. *Some Asymmetries in Japanese and Their Theoretical Implications*. Cambridge, MA: MIT dissertation.

Tamaoka, Katsuo. 2005. Tyugokugo-o bogotosuru nihongo gakusyusyaniyoru seizyun/ kakimaze-gozyun-no nodobun-to kanobun-no rikai [Comprehension of Japanese active and potential sentences with canonical and scrambled word orders by native Chinese speakers learning the Japanese language]. *Journal of Japanese Grammar* 5(2). 92–109.

Tamaoka, Katsuo, Jun-ichiro Kawahara, Hyunjung Lim, & Yayoi Miyaoka. 2001. Is there 'default' in word order? Scrambling effects on the processing of Japanese and Korean Sentences. Paper presented at the 5th conference of the Cognitive Neuropsychology Society, Nagoya University, Japan, 2–3 August.

Tamaoka, Katsuo, Hiromu Sakai, Jun-Ichiro Kawahara, Yayoi Miyaoka, Hyunjung Lim, & Masatoshi Koizumi. 2005. Priority information used for the processing of Japanese sentences: Thematic roles, case particles or grammatical functions? *Journal of Psycholinguistic Research* 34. 281–332.

Wang, Suiping, Zude Zhu, John X Zhang, Zhaoxin Wang, Zhuangwei Xiao, Huadong Xiang, & Hsuan-Chih Chen. 2008. Broca's area plays a role in syntactic processing during Chinese reading comprehension. *Neuropsychologia* 46(5). 1371–1378.

Akinori Ito, Satoru Mizuochi, and Takashi Nose

# Chapter 14
# Spoken term detection from utterances of minority languages

## 1 Introduction

There are 3,000 to 6,000 languages globally (Kim 2001), most of which are in danger of extinction (Krauss 1992; Janse 2003; Simons and Lewis 2013). As it is difficult to prevent the extinction of endangered languages, much effort has been made to record them. For example, Yang and Rau (2005) described an effort to archive the speech of Yami, an aboriginal language spoken in Taiwan. A similar effort has been made by Laoire (2008) for Scottish Gaelic, Ćavar, Cavar, and Cruz (2016) for Chatino, and Batibo (2009) for Naro. Since archiving a language is very costly, the use of computing technologies such as speech recognition and forced alignment has been examined (Palmer et al. 2010; Gerstenberger et al. 2016; Foley et al. 2018). However, it is generally difficult to develop speech recognizers for endangered languages because there is no large language resource for them. Languages with no language resources are called "zero-resource languages" (Jansen, Church, and Hermansky 2010).

Without a language resource for model training for zero-resource languages, we need to develop a practical method for documenting the language that does not require model training using any language resource. One established approach is to use "query by example spoken term detection" (QbE-STD, also called "word spotting"), where the system searches for pronunciations in the speech database that sound similar to a given query speech.

QbE-STD has been investigated since the very beginning of speech processing (Medress et al. 1978). The basic technique of QbE-STD is dynamic time warping (DTW), also called dynamic programming (DP) matching (Myers, Rabiner, and Rosenberg 1980; Nakagawa 1984), which is a non-linear matching method between two sequences with variable lengths. The DTW-based method has been widely applied to QbE-STD for zero-resource languages (Muscariello, Gravier, and Bimbot 2011; Mantena and Prehallad 2013; Gracia, Anguera, and Binefa 2014; Ito and Koizumi 2018; Ram, Asaei, and Bourlard 2018).

**Acknowledgments:** Prof. Masatoshi Koizumi, Graduate School of Arts and Letters, Tohoku University, provided the Kaqchikel speech database. This study was supported by Yotta Informatics Project by MEXT, Japan, and JSPS Kakenhi JP19H05589.


https://doi.org/10.1515/9783110778946-014

Conventional methods of QbE-STD extract acoustic feature vectors such as mel-frequency cepstral coefficients (MFCC) from speech and match the query and database features. A problem with this framework is that an acoustic feature varies not only with the linguistic content of speech but also with speaker variation. Therefore, we need a feature that depends only on the linguistic content of speech (i.e., phonemes) and is independent of the speaker. The phonetic posteriorgram (PPG), a vector of posterior probabilities of phonemes, is one such feature (Hazen, Shen, and White 2009). However, the drawback of using the PPG here is that a speech recognizer of the target (minority) language is needed in order to calculate the posteriorgram, because phoneme inventories are language specific. This chapter presents a new method we developed to use PPGs of multiple major languages (English and Japanese) for the QbE-STD of a minority language.

## 2 Spoken term detection based on DTW

In this section, we briefly introduce the algorithm of QbE-STD based on DTW. The DTW method is a technique to find the best non-linear matches to a short sequence from a long sequence (Nakagawa 1984).

Let the sequence of feature vectors of a database speech be $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_I$ and that of the query speech be $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_J$. In general, the length of the query is shorter than the database speech ($I > J$). A feature vector is typically a frequency-based feature such as MFCC or speaker-independent (and language-dependent) features such as PPG. The purpose of QbE-STD is to find the set of non-linear matching $\Phi = \{\Phi_1, \ldots, \Phi_I\}$,

$$\Phi_n = \phi_1^n \phi_2^n \ldots \phi_{K_n}^n \tag{1}$$

$$\phi_k^n = \left(i_k^n, j_k^n\right) \tag{2}$$

$$i_1^n \geq 1, i_k^n \in \left\{i_{k-1}^n + 1, i_{k-1}^n + 2\right\}, i_{K_n}^n \leq I \tag{3}$$

$$j_1^n = 1, j_k^n \in \left\{j_{k-1}^n + 1, j_{k-1}^n + 2\right\}, j_{K_n}^n = J \tag{4}$$

where $\Phi_n$ is the best non-linear matching that ends at position $n$ of the database speech. To determine the best non-linear match, the dynamic programming algorithm is used (Sakoe and Chiba 1978; Nakagawa 1984),

$$D = \left\{d_{i,j}\right\}, d_{i,j} = \|\boldsymbol{x}_i - \boldsymbol{y}_j\|^2 \tag{5}$$

$$G = \{g_{i,j}\} \tag{6}$$

$$g_{ij} = \infty \text{ for } i, j \le 0 \tag{7}$$

$$g_{i,1} = d_{i,1} \tag{8}$$

$$g_{i,j} = \min \begin{cases} g_{i-1,j-1} + d_{i,j} \\ g_{i-2,j-1} + (d_{i-1,j} + d_{i,j})/2 \\ g_{i-1,j-2} + d_{i,j-1} + d_{i.j} \end{cases} \tag{9}$$

The best match $\Phi_n$ is determined by tracing back the best choice in Eq. (9).

Figure 1 shows the distance matrix *D* and the matching matrix *G*. We can see dark-colored slanted lines in the lower panel of Figure 1, which are the candidates of occurrence of the query speech in the database speech.

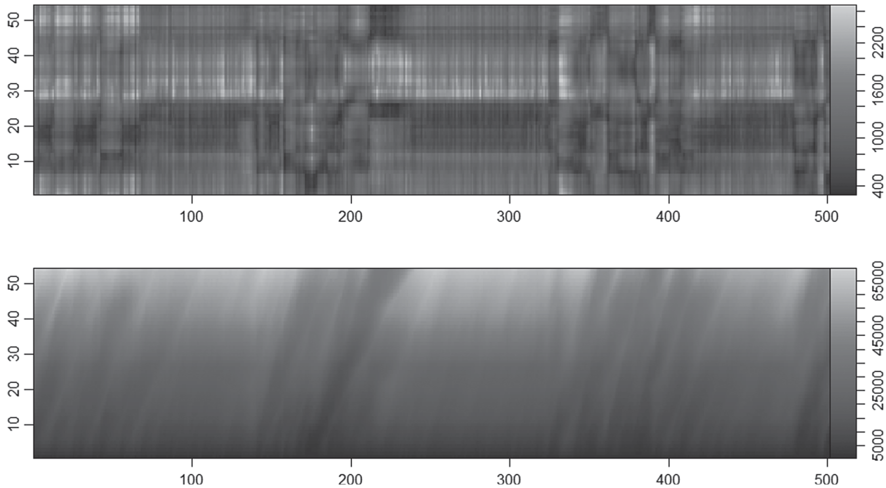



**Figure 1:** The distance matrix *D* (the upper panel) and the matching matrix *G* (the lower panel). The x-axis is the 10-ms frames of the database speech and the y-axis is the 10-ms frames of the query speech. Darker color expresses a smaller distance.

After calculating *G*, we can confirm the accumulated distances by observing $g_{i,j}$. The upper panel of Figure 2 shows the distance $g_{i,j}$. We can determine the endpoint of the query speech by picking the minima of the distance. The lower panel of Figure 2 shows the detected positions of the query and the best correspondence paths $\Phi_n$.

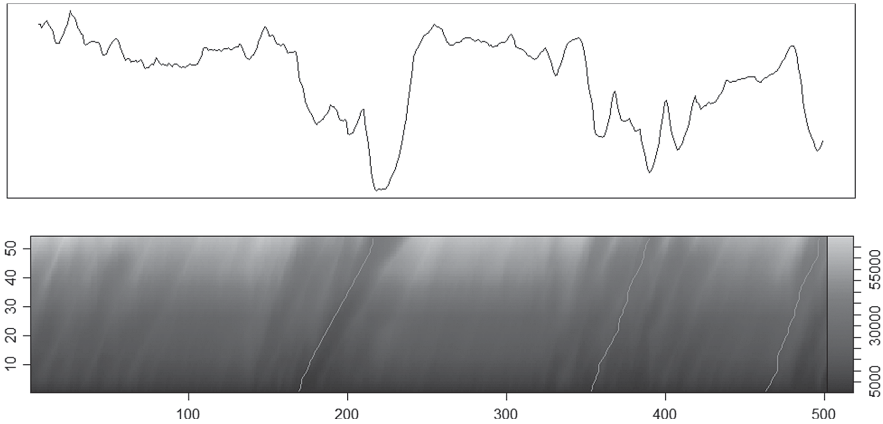



**Figure 2:** Detection of the query. The upper panel shows the accumulated distance $g_{i,j}$ and the lower panel shows the best paths $\Phi_n$ (white lines).

# 3 Speaker-independent and language-independent features

## 3.1 Speaker dependency of acoustic features

As mentioned above, conventional features like MFCC depend on pronunciation and speaker differences. Figure 3 shows an example of the effect of speaker differences. The database speech is a conversation speech of Kaqchikel, and we measured distances between two spoken queries. Both queries were the same pronunciation “matyöx” (which means “thank you”), spoken by two different speakers. One speaker was the one in the database. From Figure 3, we can see that the distance patterns of the two queries are completely different, and detection of the query by the different speakers fails.

## 3.2 The phonetic posteriorgram

It is desirable to use a feature that only depends on pronunciation differences and not on speaker differences. There have been a few studies that sought speaker-independent features for speech recognition (Malayath et al. 1997; Choi et al. 2008); however, it is difficult to find such a feature using only a signal processing tech-

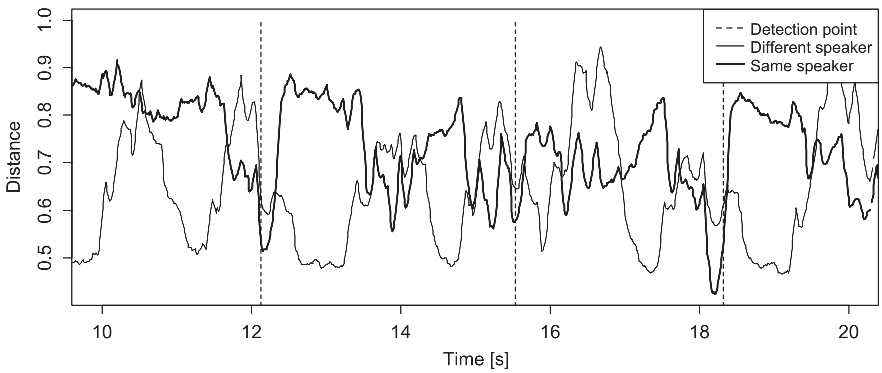



**Figure 3:** Accumulated distances of the query of the same speaker (thick line) and the different speaker (thin line).

nique because the speaker variation is sometimes larger than the feature value variation for pronunciation differences. Thus, machine learning techniques have been used to absorb only speaker variation while keeping pronunciation variation unchanged (Kato and Sugiyama 1993). The phonetic posteriorgram (PPG) (Hazen, Shen, and White 2009) is one such method that uses machine learning. A PPG is a set of posterior probabilities of phonemes, calculated frame by frame, where a frame is around 10 ms long.

When we have a feature vector $\boldsymbol{x}$ directly extracted from speech (such as MFCC), then we calculate the PPG from $\boldsymbol{x}$:

$$PPG(\boldsymbol{x}) = \left(p_1, \ldots, p_Q\right), p_k \geq 0 \tag{10}$$

$$\sum_{k=1}^{Q} p_k = 1 \tag{11}$$

where $p_i$ is a posterior probability of a specific phoneme, and $Q$ is the phoneme inventory size in the target language. The PPG is extracted using a phoneme recognizer trained by a large training set containing multiple speakers' utterances. Using a machine-learning-based phoneme recognizer, the machine learning model ignores speaker differences and focuses only on pronunciation variation. At the early stage of PPG, a Gaussian mixture model (GMM) (Zhang and Glass 2009) was used; today, a deep neural network (DNN) is used (Obara et al. 2016; Cetinkaya, Gundogdu, and Saraclar 2016; Kamiyama et al. 2017). Figure 4 shows a DNN-based PPG extractor. We first set the MFCC feature sequence, and then input multiple frames of MFCC vectors into a feed-forward neural network. The network has output units corresponding to the phonemes, and the network is trained so that only one phoneme has the output value of one while all the other phonemes are zero.

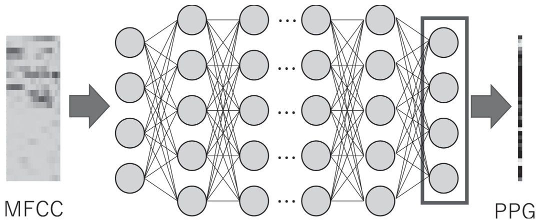



**Figure 4:** Extraction of PPG using a deep neural network.

## 3.3 Conventional language-independent features

Since phonemes are a linguistic concept, the PPG inevitably depends on the language. Therefore, we expect the performance to deteriorate when the PPG extractor for one language is applied to another language. Since the target languages of this work are zero-resource languages, we cannot develop a phoneme recognizer for the target language.

Several ideas to make speaker-independent and language-independent features have been proposed. One idea is to use discriminant features instead of phonemes (Anguera 2012; Wu, Sakti, and Nakamura 2021). Since a phoneme can be described by combining several discriminant features, we expect a discriminant feature extractor trained by one language to be effective (or, at least, more robust) when applied to another language. Another idea is to use a bottleneck feature (BNF) trained by multiple languages (Lim et al. 2017; Ram, Miculicich, and Bourlard 2019). Figure 5 shows the extraction of the BNF, which is based on a similar network to the PPG extraction. The difference is that the BNF extraction network has an extra layer in the hidden layers, which has a smaller number of units (the bottleneck layer). The network is trained in the same way as the PPG network, and the output from the bottleneck layer becomes the feature. When using the BNF, we expect useful information for discriminating phonemes to be concentrated in the BNF.

## 3.4 A language-independent feature by the combination of multiple PPGs

In addition to those conventional methods, we propose simple methods that use multiple languages. The first method, PPG_CONC, concatenates PPGs that have been cal-

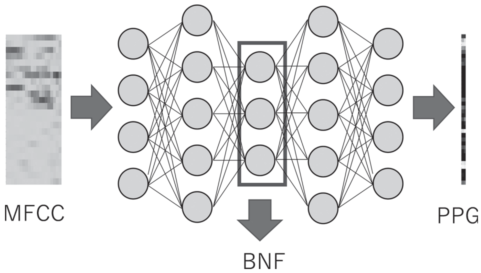



**Figure 5:** The bottleneck feature.

culated using multiple PPG extractors developed from different languages. Figure 6 shows how we extracted PPG_CONC. We used Japanese and English as resource-rich languages, so we have two PPG extractors. After extracting English and Japanese PPGs individually, we concatenated two PPGs.

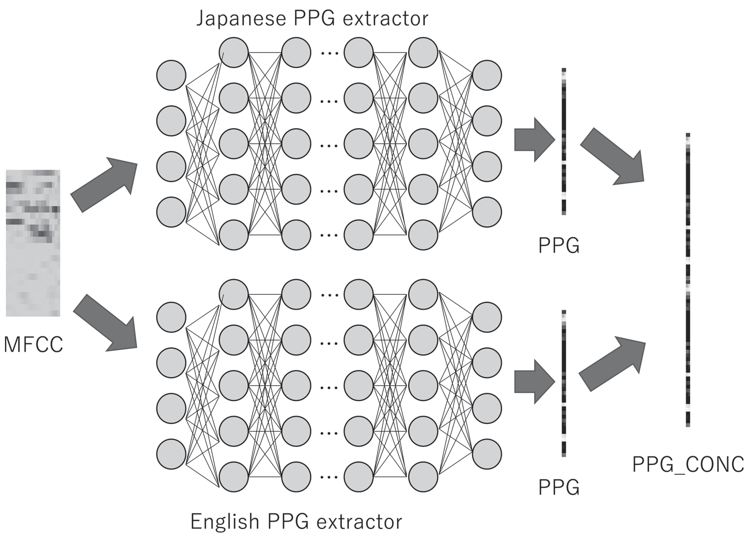



**Figure 6:** Concatenation-based multilingual PPG extraction (PPG_CONC).

Another method, PPG_MULTI, is similar to PPG_CONC. The difference is that, in PPG_ MULTI, we share the input and hidden layers, and only the output layers are prepared for Japanese and English. In this method, we expect the feature calculation in the hidden layer to become more robust because we can use more data for training. Figure 7 shows the extraction of PPG_MULTI.

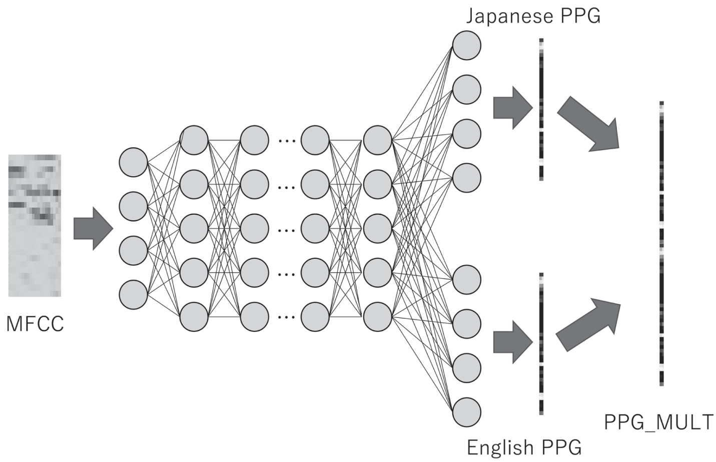



**Figure 7:** Multi-task learning-based multilingual PPG extraction (PPG_MULTI).

# 4 Experimental conditions

## 4.1 Architecture of the PPG extractor

The PPG extractor is a feed-forward neural network. The input feature was MFCC combined with its first and second derivatives (24 dimensions/frame in total). Several frames were combined and fed to the network.

Table 1 shows the conditions for neural network training. The number of hidden layers and nodes in a hidden layer are optimized using Optuna, a system for tuning hyperparameters (Akiba et al. 2019).

**Table 1:** Conditions of neural network training.

| | |
|---|---|
| No. of hidden layers | 2 to 7 |
| No. of nodes/hidden layer | 512, 1024, 2048, 4096 |
| Optimizer | Adam |
| Activation function | ReLU |
| Dropout probability | 0.5 |
| Input segment width | 1, 9, 15, 17 |
| Input nodes | 24 × Segment width |
| Output nodes | 36 (Japanese) / 46 (English) |

The bottleneck feature extractor's conditions were the same as shown in Table 1, except that the number of hidden layers was fixed to three, and the middle layer was the bottleneck layer, having 30 nodes, which was similar to the size of the network used in the previous work (Ram, Miculicich, and Bourlard 2019).

When training the PPG_MULTI network, we examined two training procedures, as shown in Figure 8. In this figure, the rightmost part (the black rectangle) shows the reference signal of the training, where the black part is 0, and the white part is 1. In the first procedure (PPG_MULTI(ALL), Figure 8 (a)), when training PPGs of one language, the network is trained to output zeros for all phonemes of the other language. Thus, the PPG_MULTI(ALL) network is trained to discriminate not only the phonemes of one language but also those of another language. On the other hand, the second procedure (PPG_MULTI(DIV), Figure 8 (b)) trains the two languages independently. Thus, when training PPGs of one language, connections to the output layer of the other language are not trained (shown by gray lines in the figure). This condition is similar to the PPG_CONC network, where two networks for two languages are trained independently. The difference is that the PPG_MULTI(DIV) network shares the input and hidden layers across the languages.

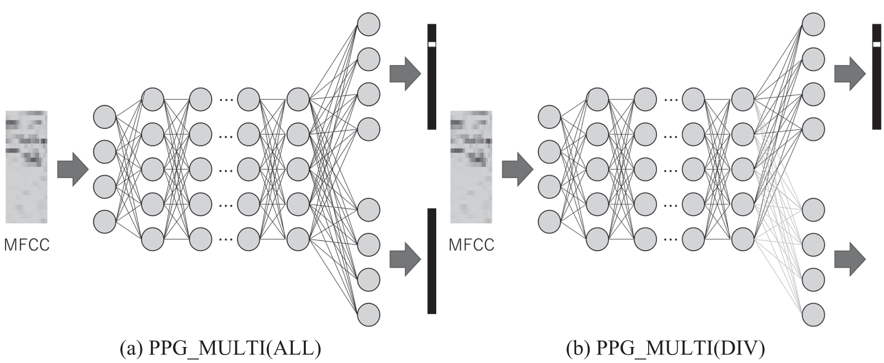



**Figure 8:** Training procedures of PPG_MULTI.

## 4.2 Datasets for experiment

We used two data sets for training the PPGs. First, we used the JNAS corpus (Itou et al. 1999) for the Japanese language, where 2794, 286, and 286 sentences were used for training, validation, and testing, respectively. Second, for the English language, we used the TIMIT corpus, where 4900, 400, and 400 sentences were used for training, validation, and testing, respectively.

Using the PPGs extracted from the trained networks, we carried out QbE-STD experiments. We prepared experimental datasets for three languages: Japanese, English, and Kaqchikel. Table 2 shows the database and query words used in the experiment. The Kaqchikel speech database is a conversational speech, parts of which are conversations in a textbook (Brown, Maxwell, and Little 2010). The test database, query words and the training data of the PPG extractor did not include the same speakers.

**Table 2:** Test data for QbE-STD experiments.

| Language | Test database | Query words |
|---|---|---|
| Japanese | 286 sentences from JNAS | *ajia, kankei, senkyo, chousa, nihon, paasento* |
| English | 400 sentences from TIMIT | dark, greasy, suit, wash, water, year |
| Kaqchikel | 347 sentences | *achike, matyöx, peraj, richin* |

The detection performance was measured by using the mean average precision (MAP) (Garofolo et al. 2000). When validating the detection results, we determined that the detection was correct when a detected term was included in the sentence, regardless of the detection position.

# 5 Experimental results

## 5.1 Phoneme recognition

First, we observed the frame-by-frame phoneme recognition accuracy by the PPG and BNF extractors. Figure 9 shows the results for each segment width. For example, EN09 shows the results for English, segment width 9. These results are of the best hyperparameters (number of hidden layers and number of hidden nodes). The results of Japanese phoneme recognition were better than those of English, possibly because English has a larger phoneme inventory than Japanese. The best results were obtained when the segment width was 15. On the other hand, the phoneme accuracy of BNF extractors was not better than the PPG extractors. The reason seems to be that the BNF extractor has a small number of hidden nodes in the bottleneck layer.

## 5.2 Comparison of MFCC, PPG, and BNF

Figure 10 shows the MAP results for detection results using different input features. From the results of QbE-STD from English and Japanese, we confirm that the matched conditions (Japanese PPG for Japanese QbE-STD and English PPG for English QbE-STD)

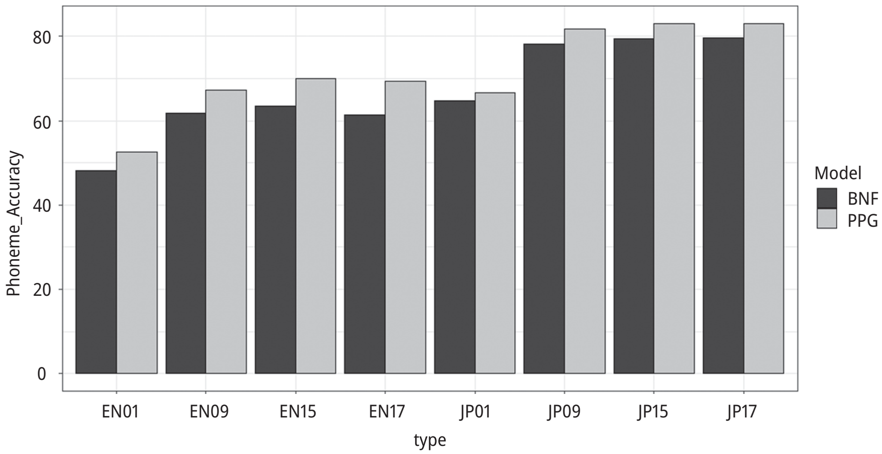



**Figure 9:** Phoneme recognition accuracies.

showed better QbE-STD performance than the unmatched conditions. For QbE-STD from English and Japanese, PPGs with both matched and unmatched conditions outperformed QbE-STD using MFCC. For the QbE-STD from Kaqchikel, on the other hand, only the Japanese PPG performed better than QbE-STD using MFCC. The BNF features did not show any better results compared with PPGs.

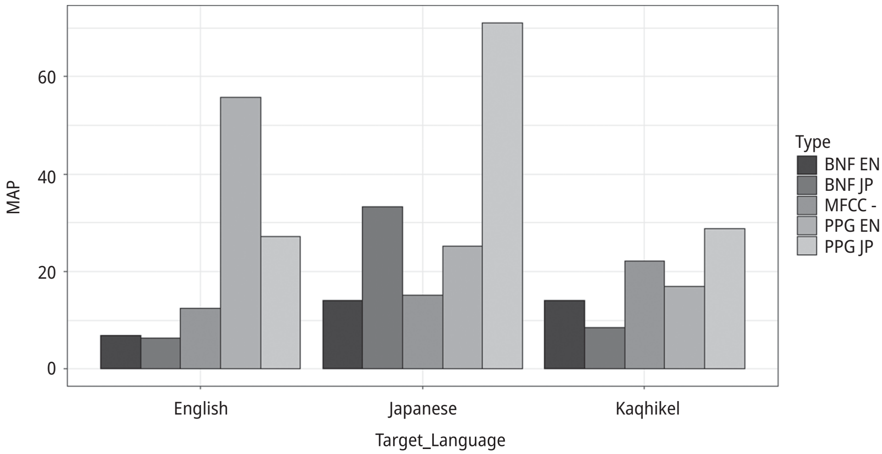



**Figure 10:** MAP results for three languages.

### 5.3 PPG with multiple language training

Finally, we compared the QbE-STD results using multiple language PPGs (PPG_CONC, PPG_MULTI(ALL), and PPG_MULTI(DIV)). In this case, we chose the best segment width according to the condition. Figure 11 shows the experimental results. From these results, the PPG_CONC feature was the best among the other features, and its performance was slightly better than that of the matched condition (PPG_EN for English and PPG_JP for Japanese). PPG_MULTI(DIV) showed almost the same performance as PPG_CONC. On the other hand, PPG_MULTI(ALL) showed lower MAP, possibly because it was difficult to discriminate similar pronunciations of different languages (such as Japanese [a] and English [ɑ]). Thus, PPGs with multiple languages, PPG_CONC and PPG_MULTI(DIV), were also effective for Kaqchikel.

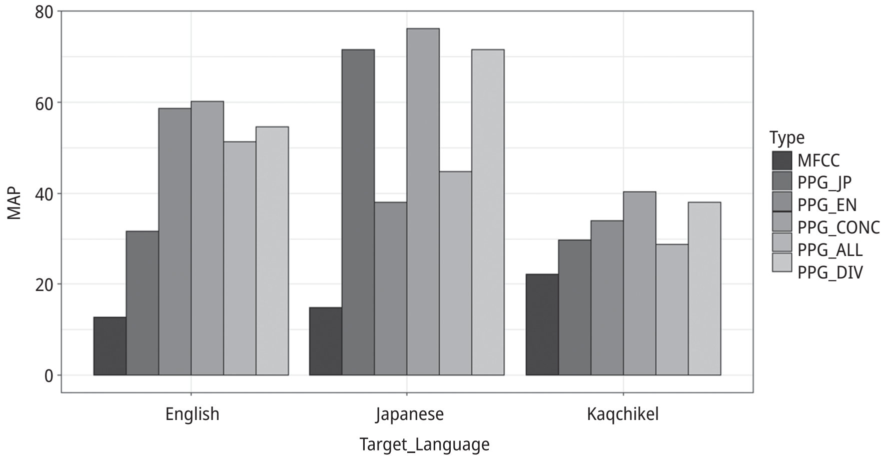



**Figure 11:** MAP results with PPGs trained with multiple languages.

## 6 Analysis of expressive power of PPGs

### 6.1 Cluster analysis by simulation

In this work, we expressed Kaqchikel speech using Japanese and English PPGs. As mentioned above, the Kaqchikel phoneme inventory includes several phonemes not included in the English and Japanese phoneme inventories, such as glottal stop (Adell 2014). Therefore, we investigated how the combined PPG expresses the phonemes of Kaqchikel. To this end, we analyzed the PPG sequence calculated for

Kaqchikel in order to determine how discriminative the PPGs are. This was done by conducting a cluster analysis for PPGs. As shown in Eq. (10), one frame of PPG is a set of posterior probabilities. If the accuracy of the phoneme classifier is good enough, a PPG is sparse; in other words, only one element of a PPG is nearly one, and all other elements are nearly zero. In this case, the Euclidean distance of two PPGs of the same phoneme is nearly zero, and between different phonemes is nearly one. Following this idea, we developed a method to measure the expressive power of PPGs.

Consider the PPG sequence $\boldsymbol{X} = \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$. If we apply a hierarchical clustering algorithm to $\boldsymbol{X}$, PPGs of the same phoneme will make clusters. Figure 12 is a simulation result of the hierarchical clustering, where the data are vectors with only one element that is nearly one and the other elements having small values (noise). From the dendrogram shown in Figure 12, we confirm that most of the same phonemes are gathered at low height, and clusters of different phonemes are gathered at a greater height.


**Figure 12:** A dendrogram generated by the hierarchical clustering (average distance method with Euclidean distance) performed on simulated data.

Next, we cut the dendrogram at an arbitrary position and count the number of clusters. When $N$ is large, the hierarchical clustering needs much computation time; therefore, we first apply the k-means clustering to the data in order to cluster the input data into 100 clusters, and then we apply the hierarchical clustering on the 100 centroids. Figure 13 shows the relationship between the relative cutting threshold (the highest position is 100) and the number of clusters with different amounts of noise. In this simulation, we set the noise factor $\alpha$ and generate the pseudo-PPG of phoneme $k$ as follows:

$$\boldsymbol{x} = (p_1, \ldots, p_Q) \tag{12}$$

$$q_i = \begin{cases} 1 + au & i = k \\ au & i \neq k \end{cases} \quad \text{where } u \sim U(0,1) \tag{13}$$

$$p_i = \frac{q_i}{\sum_{j=1}^{Q} q_j} \tag{14}$$

where, $u$ is a random number and $U$(0,1) means the uniform distribution between 0 and 1. In addition, the number of phonemes $Q$ was set to 40.

If the noise is small ($a$ = 0.1), the number of classes is almost constant. The number of classes in the constant region is the number of phonemes ($Q$ = 40). If PPGs have a large amount of noise, the region with a constant number of classes becomes narrower.

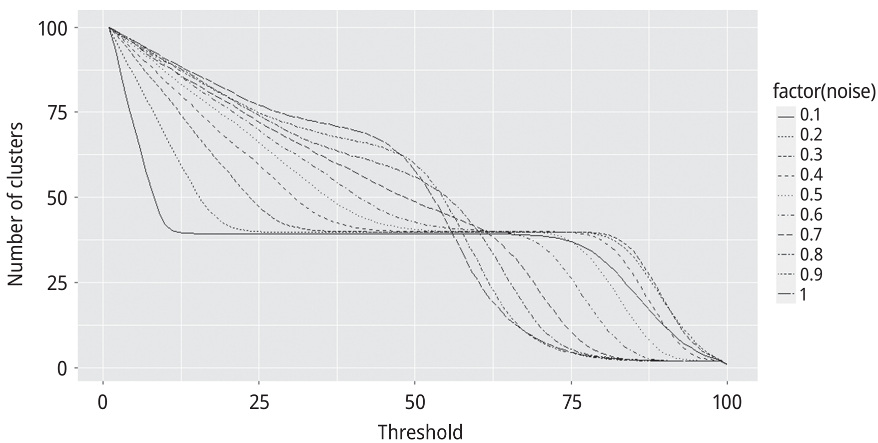



**Figure 13:** Cutting threshold and number of clusters.

Another factor is the skewness of phoneme frequency. The simulation of the above figure assumes that the frequency of different phonemes is equal. However, the actual frequencies of phonemes are different, and the frequency follows an exponential distribution. Thus, we change the skewness of frequency. When we arrange $Q$ phonemes from the most frequent to least frequent ones, the frequency of the $k$-th phoneme $F_k$ is assumed as follows:

$$F_k \propto \exp\left(-\frac{k}{S}\right) \tag{15}$$

where, $S$ is a hyperparameter to control the skewness. If $S$ is large, the frequencies become flat. In the following result, we assumed the noise factor as 0.3 and changed $S$ (shown as “decay”). Figure 14 shows the simulation result. We now confirm that different $S$ changes the height of the plateau. If the phoneme frequency is skewed, it has the same effect as if the effective number of phoneme types is small.

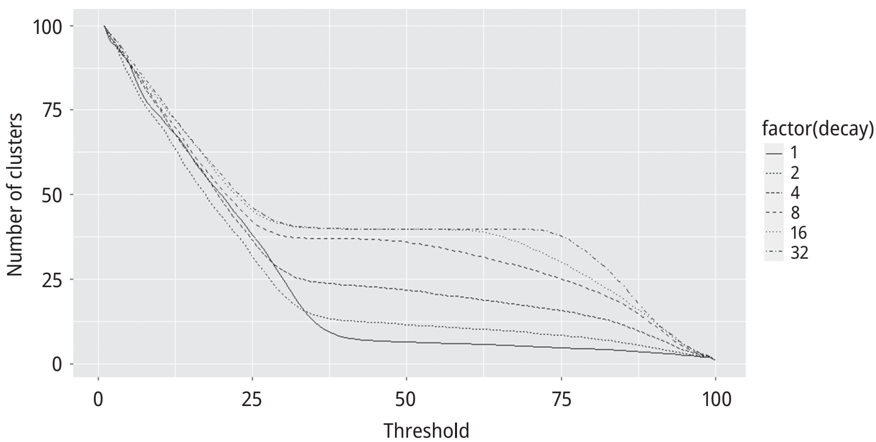



**Figure 14:** Number of clusters when the frequency skewness changes.

## 6.2 Analysis of actual PPGs

Next, we discuss the actual PPGs. The data for the analysis were the same as the database speech used in the QbE-STD experiment. We analyzed the above analysis for all sentences and took the average. Figure 15 shows the curves of actual PPGs. The black line shows the number of English and Japanese phonemes. Figure 15(a) is the result for the three languages expressed by the English PPGs, and Figure 15(b) is that by the Japanese PPGs. From Figure 15(a), we recognize that the constant region of English is longer than that of Japanese, which explains the difference between the expressive power of the English and Japanese PPGs. The curve for Kaqchikel has no constant part, which means the PPG is very noisy. As for Figure 15(b), the constant part is longer than that of English, meaning that the expressive power of the Japanese PPG is stronger than that of the English PPG. Note that the height of the constant part is lower than the number of phonemes, which implies that some of the phonemes are not detected. The Japanese phoneme inventory used in this work includes phonemes such as [$p^j$] and [$m^j$], which are seldom detected. With the

Japanese PPG, the Kaqchikel result has a constant part, consistent with the QbE-STD result shown in Figure 10.

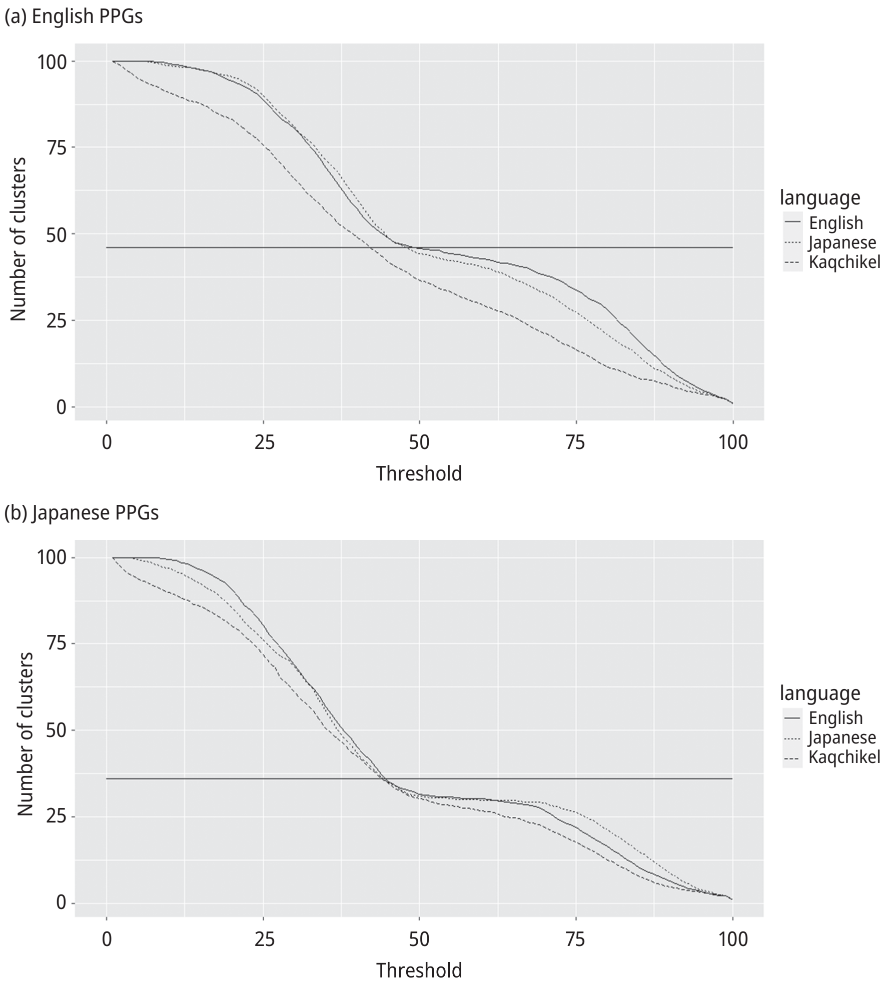



**Figure 15:** Analysis of actual PPGs.

The reason for the difference of expressive power is still not clear. Comparing Figure 15 (a) and (b), it is clear that Kaqchikel speech expressed by English PPGs is noisier than that by Japanese PPGs. One possible reason for this result is that the difference of phoneme inventories causes some kind of noise. According to the literature (Adell 2014; Bennett 2016; Bennett and Henderson 2018), Kaqchikel has

32 phonemes, in which 10 and 14 phonemes are not included in the English and Japanese phoneme inventories, respectively (in this analysis, we unified similar phonemes such as English [b] and Kaqchikel [ɓ̥], Japanese [t͡ɕ] and Kaqchikel [t͡ʃ], etc.). The phonemes that are included in only Kaqchikel are ejectives [tʃ’] [k’] [t’] [t͡s’], unvoiced uvular plosives [q] and [q’], unvoiced velar fricative [x], and glottal stop [ʔ]. Table 3 shows the number of phonemes included in PPGs but not Kaqchikel, included in both PPGs and Kaqchikel, and not included in PPGs but in Kaqchikel. Although English PPGs express more Kaqchikel phonemes than Japanese PPGs, the number of phonemes in English PPGs that are not included in Kaqchikel is larger than that of Japanese PPGs (24 and 18, respectively). These phonemes become noise when expressing Kaqchikel speech.

**Table 3:** Number of phonemes that are included in PPG and Kaqchikel phoneme inventory.

| PPG language | In PPG and not K | In both PPG and K | In only K |
|---|---:|---:|---:|
| English | 24 | 22 | 10 |
| Japanese | 18 | 18 | 14 |

# 7 Conclusions

We examined QbE-STD for zero-resource languages. In this method, we used the PPGs of resource-rich languages and used the PPGs as speaker-independent features. We examined PPGs and BNFs trained with English and Japanese and PPGs with multiple languages. From the experiment, we obtained the following findings.

- PPGs are effective compared with MFCC even if the training language is different from the target language.
- BNFs are not better than PPGs.
- Combining PPGs of multiple languages improves the performance of QbE-STD.
- The expressive power of the English and Japanese PPGs is not sufficient to describe Kaqchikel.

In a future work, we will examine more languages (such as Spanish or Chinese) to calculate PPGs. Moreover, we will examine other minority languages in addition to Kaqchikel.

# References


Adell, Eric. 2014. Descriptive account of Spanish loanword phonology in Kaqchikel. *Kansas Working Papers in Linguistics* 34. 43–59.

Anguera, Xavier. 2012. Speaker independent discriminant feature extraction for acoustic pattern-matching. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 485–488.

Akiba, Takuya, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, & Masanori Koyama. 2019. Optuna: A next-generation hyperparameter optimization framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2623–2631.

Batibo, Herman M. 2009. Language documentation as a strategy for the empowerment of the minority languages of Africa. *Selected Proceedings of the 38th Annual Conference on African Linguistics*. 193–203.

Bennett, Ryan. 2016. Mayan phonology. *Language and Linguistics Compass* 10(10). 469–514.

Bennett, Ryan & Robert Henderson. 2018. Nasal hardening and aspect allomorphy in Kaqchikel. In Ryan Bennett, Andrew Angeles, Adrian Brasoveanu, Dhyana Buckley, Nick Kalivoda, Shigeto Kawahara, Grant McGuire, & Jaye Padgett (eds.), *Hana-bana: A Festschrift for Junko Itô and Armin Mester*. Santa Cruz, CA: Department of Linguistics, University of California, Santa Cruz. https://itomestercelebration.sites.ucsc.edu/

Brown, R. McKenna, Judith M. Maxwell, & Walter E. Little. 2010. *¿La ütz awäch?: Introduction to Kaqchikel Maya Language.* Austin, TX: University of Texas Press.

Ćavar, Małgorzata E., Damir Cavar, & Hilaria Cruz. 2016. Endangered language documentation: Bootstrapping a Chatino speech corpus, forced aligner, ASR. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC)*. 4004–4011.

Cetinkaya, Gozde, Butuhan Gundogdu, & Murat Saraclar. 2016. Pre-filtered dynamic time warping for posteriorgram based keyword search. *Proceedings of IEEE Spoken Language Technology Workshop (SLT)*. 376–382.

Choi, Heeyoul, Ricardo Gutierrez-Osuna, Seungjin Choi, & Yoonsuck Choe. 2008. Kernel oriented discriminant analysis for speaker-independent phoneme spaces. *Proceedings of 19th International Conference on Pattern Recognition (ICPR)*. 1–4.

Foley, Ben, Josh Arnold, Rolando Coto-Solano, Gautier Durantin, Ellison Mark, Daan van Esch, Heath Scott, František Kratochvíl, Zara Maxwell-Smith, David Nash, Ola Olsson, Mark Richards, Nay San, Hywel Stoakes, Nick Thieberger, & Janet Wiles. 2018. Building speech recognition systems for language documentation: The CoEDL endangered language pipeline and inference system (ELPIS). *Proceedings of the 6th International Workshop on Spoken Language Technologies for Under-Resourced Languages*. 205–209.

Garofolo, John S., Cedric G. P. Auzanne, & Ellen M. Voorhees. 2000. The TREC spoken document retrieval track: A success story. *NIST SPECIAL PUBLICATION SP 500*(246). 107–130. Paris: RIAO-2000 Content-Based Multimedia Information Access.

Gerstenberger, Ciprian, Niko Partanen, Michael Rießler, & Joshua Wilbur. 2016. Utilizing language technology in the documentation of endangered Uralic languages. *Northern European Journal of Language Technology* 4. 29–47.

Gracia, Ciro, Xavier Anguera, & Xavier Binefa. 2014. Combining temporal and spectral information for query-by-example spoken term detection. *Proceedings of the 2014 22nd European Signal Processing Conference (EUSIPCO)*. 1487–1491.

Hazen, Timothy J., Wade Shen, & Christopher White. 2009. Query-by-example spoken term detection using phonetic posteriorgram templates. *Proceedings of the 2009 IEEE Workshop on Automatic Speech Recognition & Understanding*. 421–426.

Ito, Akinori & Masatoshi Koizumi. 2018. Spoken term detection of zero-resource language using machine learning. *Proceedings of the 2018 International Conference on Intelligent Information Technology*. 45–49.

Itou, Katunobu, Mikio Yamamoto, Kazuya Takeda, Toshiyuki Takezawa, Tatsuo Matsuoka, Tetsunori Kobayashi, Kiyohiro Shikano, & Shuichi Itahashi. 1999. JNAS: Japanese speech corpus for large vocabulary continuous speech recognition research. *Journal of the Acoustical Society of Japan (E)* 20(3). 199–206.

Janse, Mark. 2003. Introduction: Language death and language maintenance, problems and prospects. In Mark Janse & Sijmen Tol (eds.), *Language Death and Language Maintenance: Theoretical, Practical and Descriptive Approaches*, 9–17. Amsterdam: John Benjamins.

Jansen, Aren, Kenneth Church, & Hynek Hermansky. 2010. Towards spoken term discovery at scale with zero resources. *Proceedings of INTERSPEECH*. 1676–1679.

Kamiyama, Hosana, Atsushi Ando, Satoshi Kobashikawa, & Yushi Aono. 2017. Robust children and adults speech identification and confidence measure based on DNN posteriorgram. *Proceedings of Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. 502–505.

Kato, Yoshinaga & Masahide Sugiyama. 1993. Speaker-independent features extracted by a neural network. *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing* 1. 553–556.

Kim, Seong-do. 2001. The landscape of languages at the commencement of the 21st century. *Journal of Universal Language* 2(1). 15–23.

Krauss, Michael. 1992. The world's languages in crisis. *Language* 68(1). 4–10.

Laoire, Sioban Ni. 2008. Scottish Gaelic speech and writing: Register variation in an endangered language. *Scottish Language* 2. 115–119.

Lim, Hyungjun, Younggwan Kim, Yoonhoe Kim, & Hoirin Kim. 2017. CNN-based bottleneck feature for noise robust query-by-example spoken term detection. *Proceedings of Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. 1278–1281.

Malayath, Narendranath, Hynek Hermansky, Alexander Kain, & Rolf Carlson. 1997. Speaker-independent feature extraction by oriented principal component analysis. *Proceedings of the European Conference on Speech Communication and Technology*. 1–4.

Mantena, Gautam & Kishore Prahallad. 2013. IIIT-H SWS 2013: Gaussian posteriorgrams of bottleneck features for query-by-example spoken term detection. *Proceedings of MediaEval 2013 Workshop.*

Medress, Mark F., Timothy C. Diller, Dean R. Kloker, Larry L. Lutton, Henry N. Oredson, & Toby E. Skinner. 1978. An automatic word spotting system for conversational speech. *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* 3. 712–717.

Muscariello, Armando, Guillaume Gravier, & Frédéric Bimbot. 2011. Zero-resource audio-only spoken term detection based on a combination of template matching techniques. *Proceedings of INTERSPEECH*. 921–924.

Myers, Cory, Lawrence R. Rabiner, & Aaron E. Rosenberg. 1980. An investigation of the use of dynamic time warping for word spotting and connected speech recognition. *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* 5. 173–177.

Nakagawa, Seiichi. 1984. Connected spoken word recognition algorithms by constant time delay DP, O (n) DP and augmented continuous DP matching. *Information Sciences* 33(1–2). 63–85.

Obara, Masato, Kazunori Kojima, Kazuyo Tanaka, Shi-wook Lee, & Yoshiaki Itoh. 2016. Rescoring by combination of posteriorgram score and subword-matching score for use in query-by-example. *Proceedings of INTERSPEECH*. 1918–1922.

Palmer, Alexis, Taesun Moon, Jason Baldridge, Katrin Erk, Eric Campbell, & Telma Can. 2010. Computational strategies for reducing annotation effort in language documentation. *Linguistic Issues in Language Technology* 3(4). 1–42.

Ram, Dhananjay, Afsaneh Asaei, & Hervé Bourlard. 2018. Phonetic subspace features for improved query by example spoken term detection. *Speech Communication* 103. 27–36.

Ram, Dhananjay, Lesly Miculicich, & Hervé Bourlard. 2019. Multilingual bottleneck features for query by example spoken term detection. *Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*. 621–628.

Sakoe, Hiroaki & Seibi Chiba. 1978. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26(1). 43–49.

Simons, Gary F. & Paul M. Lewis. 2013. The world's languages in crisis: A 20-year update. In Elena Mihas, Bernard Perley, Gabriel Rei-Dovel, & Kathleen Wheatley (eds.), *Responses to Language Endangerment: In Honor of Mickey Noonan. New Directions in Language Documentation and Language Revitalization*, 3–20. Amsterdam/Philadelphia: John Benjamins.

Wu, Bin, Sakriani Sakti, & Satoshi Nakamura. 2021. Incorporating discriminative DPGMM posteriorgrams for low-resource ASR. *Proceedings of IEEE Spoken Language Technology Workshop (SLT)*. 201–208.

Yang, Men-Chien & Victoria D. Rau. 2005. An integrated framework for archiving, processing and developing learning materials for an endangered aboriginal language in Taiwan. *Proceedings of The Fifth Workshop on Asian Language Resources (ALR-05) and First Symposium on Asian Language Resources Network (ALRN)*. 32–39.

Zhang, Yaodong & James R. Glass. 2009. Unsupervised spoken keyword spotting via segmental DTW on Gaussian posteriorgrams. *Proceedings of the 2009 IEEE Workshop on Automatic Speech Recognition & Understanding*. 398–403.

Yohei Oseki

# Chapter 15
# Human language processing in comparative computational psycholinguistics

## 1 Introduction

Computational psycholinguistics of human language processing has attracted considerable attention in both experimental psycholinguistics and natural language processing (NLP), thanks to recent advances in machine learning and large corpora.[1] In computational psycholinguistics, computational models are constructed from symbolic generative models and artificial neural networks developed in NLP and, through the lens of information-theoretic complexity metrics (Hale 2016), evaluated against human behavioral and neural data collected through psycholinguistic experiments. However, the previous literature has focused almost exclusively on European languages with typologically similar characteristics (Bender 2011), so that the question whether the established conclusions in computational psycholinguistics can be generalized across languages remains to be empirically addressed. In this chapter, we advocate the comparative approach to computational psycholinguistics dubbed *comparative computational psycholinguistics*, which constructs and evaluates computational models of human language processing from comparative perspectives.

This chapter is organized as follows. Section 2 reviews the pipeline of computational psycholinguistics and, with some issues raised from comparative perspectives, proposes comparative computational psycholinguistics, building on the previous literature on comparative psycholinguistics and computational typology.

**1** Computational psycholinguistics in the broad sense includes human language *acquisition*, but here we restrict our discussions to human language *processing*.

**Acknowledgments:** This chapter originates from the talk delivered at the International Symposium on Issues in Japanese Psycholinguistics from Comparative Perspectives (IJPCP) held online on September 11–12, 2021. I would like to thank excellent collaborators such as Ryo Yoshida, Hiroshi Noji, Tatsuki Kuribayashi, Masayuki Asahara, and especially Masatoshi Koizumi for inviting me to speak at the above-mentioned symposium and contribute to this volume. I also want to thank Kohei Kajikawa and two anonymous reviewers for proofreading the manuscript and providing insightful comments. This work was supported by JST PRESTO Grant Number JPMJPR21C2 and NINJAL Collaborative Research Project "Computational Psycholinguistics of Language Processing with Large Corpora". All remaining errors are my own.


https://doi.org/10.1515/9783110778946-015

Section 3 presents the results of modeling *hierarchical syntactic structure* with Recurrent Neural Network Grammars (Dyer et al. 2016), demonstrating that hierarchical syntactic structure universally makes computational models more human-like, while optimal parsing strategies may diverge with respect to head directionality (Yoshida, Noji, and Oseki 2021). Section 4 then provides the results of modeling *cue-based memory retrieval* with Transformer architectures (Vaswani et al. 2017; Merkx and Frank 2021), suggesting that Transformer architectures are too powerful for those languages with few long-distance dependencies, which can be rendered more human-like through context limitations (Kuribayashi et al. 2021, 2022). Section 5 concludes this chapter and remarks some future directions.

# 2 Computational psycholinguistics from comparative perspectives

## 2.1 What is computational psycholinguistics?

In computational psycholinguistics, computational models are constructed from symbolic generative models and artificial neural networks developed in NLP and evaluated against human behavioral and neural data collected through psycholinguistic experiments (Crocker 1996; Lewis 2003; Hale 2017). In this sense, computational psycholinguistics is an interdisciplinary approach to human language processing at the intersection of experimental psycholinguistics and NLP. From experimental psycholinguistics, on one hand, computational psycholinguistics inherits both the scientific goal to elucidate human language processing and the human behavioral and neural data to be modeled computationally, while experimental manipulations are performed over computational models (e.g. model architecture, training data, etc.), not experimental stimuli themselves as in experimental psycholinguistics (e.g. syntactic complexity, semantic plausibility, etc.). From NLP, on the other hand, computational psycholinguistics borrows computational models such as symbolic generative models and artificial neural networks, but as computational models of human language processing with serious scientific commitments, not as pure engineering solutions as in NLP.[2]

2 Computational psycholinguistics is usually referred to as *cognitive modeling* in the NLP community. For example, the submission track for computational psycholinguistics at the Association for Computational Linguistics (*ACL) conferences is called “Linguistic Theories, Cognitive Modeling, and Psycholinguistics”, and the designated workshop for computational psycholinguistics is named “Cognitive Modeling and Computational Linguistics (CMCL)”.

Specifically, the pipeline of computational psycholinguistics generally consists of three components, as summarized in Figure 1 (Brennan and Hale 2019).

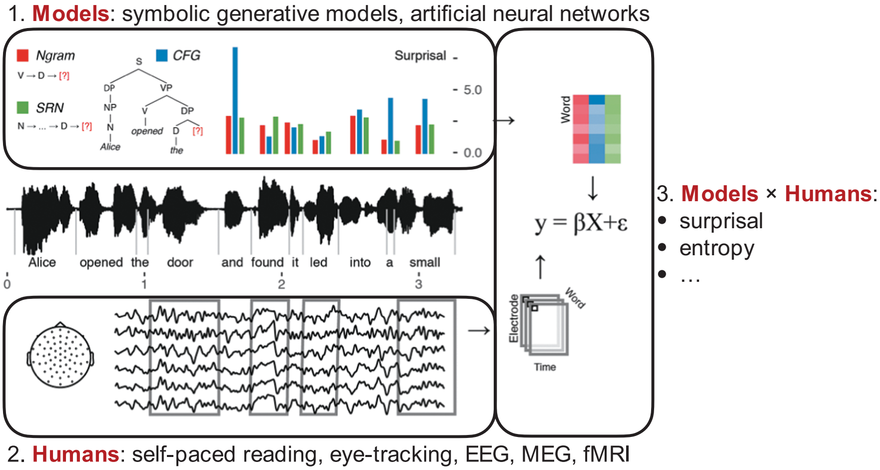



**Figure 1:** Pipeline of computational psycholinguistics (Brennan and Hale 2019).[3]

The first component is **models**: computational models are constructed from symbolic generative models and artificial neural networks developed in NLP. For example, symbolic generative models include *language models* (LMs; computational models to estimate the probabilities of the words within the sentences) such as *n*-gram models which sequentially process sentences given *n*–1 previous words and context-free grammars (CFGs) which hierarchically process sentences given their syntactic structures. In addition, artificial neural networks range from classic recurrent neural networks (RRNs; Elman 1990), through long short-term memory networks (LSTMs; Hochreiter and Schmidhuber 1997), to recent Transformer architectures (Vaswani et al. 2017).[4]

The second component is **humans**: human behavioral and neural data are collected through psycholinguistic experiments to be predicted with the computational models.[5] For instance, human behavioral data can be divided into offline

3 Figure 1 exemplifies human neural data, especially electroencephalography (EEG), but notice importantly for the purpose here that this pipeline of computational psycholinguistics can be equally applied to human behavioral data.

4 See Goldberg (2017) and Jurafsky and Martin (2022: Ch. 5–11) for the details of various architectures of artificial neural networks.

5 In practice, computational psycholinguistics proper does not collect human behavioral and neural data through psycholinguistic experiments, but rather employ publicly available language resources

measures like acceptability judgments and online measures like self-paced reading and eye-tracking. Similarly, human neural data can be classified into electrophysiological techniques like electroencephalography (EEG) and magnetoencephalography (MEG) and hemodynamic techniques like functional magnetic resonance imaging (fMRI), where the former and latter techniques exhibit higher temporal and spatial resolutions, respectively.

The third component is **models × humans**: computational models and human data are compared to evaluate which computational model most successfully predicts human behavioral and neural data. Importantly, in order to bridge the gap between probabilities estimated from computational models and processing complexities collected from human behavioral and neural data, information-theoretic complexity metrics are employed as *linking hypotheses* (Hale 2016). Specifically, there are two prominent information-theoretic complexity metrics proposed in computational psycholinguistics. The first information-theoretic complexity metric is *surprisal* (Hale 2001; Levy 2008), the negative logarithmic probability of words $w$ in context $c$ as defined in (1a) which quantifies how surprising words will be in context, hypothesizing that lower probability, hence higher surprisal, links to higher processing complexity. The second information-theoretic complexity metric is *entropy reduction* (Hale 2006), the non-negative reduction of entropy between two probability distributions $W$ over words in context $c$ as defined in (1b) which quantifies how uncertain words will be in context, hypothesizing that the higher divergence between two probability distributions, hence higher entropy reduction, links to higher processing complexity.[6]

(1) Information-theoretic complexity metrics (Hale 2016):[7]
 a. Surprisal: $I(w) = -\log_2 p(w)$
 b. Entropy: $H(W) = -\sum_{w \in W} p(w) \log_2 p(w)$

Unlike traditional complexity metrics like node and action counts (i.e. the number of nodes/actions traversed between two words; Miller and Chomsky 1963) which can be applied only to computational models with hierarchical syntactic structures, information-theoretic complexity metrics are theory-neutral with respect to rep-

naturalistically annotated through human behavioral and neural data without any experimental manipulations as “benchmarks” to evaluate the computational models (Brennan 2016).

**6** See Yun et al. (2015) and Linzen and Jaeger (2016) for the division of labor between surprisal and entropy reduction.

**7** The base 2 of logarithm can be interpreted as binary bits. For example, two bits have information to distinguish four codes: 00, 01, 10, and 11.

resentational assumptions of computational models, hence "causal bottleneck" (Levy 2008).[8]

Finally, the important logic behind this pipeline of computational psycholinguistics is that the computational model that most successfully predicts human behavioral and neural data is argued to be the most "human-like" computational model relative to baseline computational models. This logic is sometimes called the *constructive* approach or *abductive* inference in related fields such as cognitive robotics (Taniguchi et al. 2019).

## 2.2 Issues with computational psycholinguistics

Despite the remarkable success thanks to recent advances in machine learning and large corpora, there exist several issues with computational psycholinguistics.[9] One of the most urgent issues with the NLP community in general is that the previous literature has focused almost exclusively on European languages with typologically similar characteristics, especially Germanic languages like English. For example, Bender (2011) reported that, among the single-language studies published in the Association for Computational Linguistics (ACL 2008) and the European chapter of the Association for Computational Linguistics (EACL 2009), English accounted for 63% in ACL 2008 and 55% in EACL 2009, Germanic languages 71% in both ACL 2008 and EACL 2009, and surprisingly the European languages even 85% in ACL 2008 and 91% in EACL 2009.[10]

Unfortunately, computational psycholinguistics is not an exception: language resources naturalistically annotated with human behavioral and neural data such as Dundee Corpus (Kennedy and Pynte 2005) are mostly available in European languages, especially English. Language resources naturalistically annotated through human behavioral and neural data are summarized in Table 1 (Oseki and Asahara 2020). Accordingly, model evaluations of computational psycholinguistics have largely been limited to European languages, so that the question whether the established conclusions in computational psycholinguistics can be generalized

**8** Those information-theoretic complexity metrics might be collectively called the *Information(al) Theory of Complexity* (ITC), in contrast with the traditional Derivational Theory of Complexity (DTC; Miller and Chomsky 1963; Fodor, Bever, and Garrett 1974).

**9** Other issues with computational psycholinguistics include, but are not limited to: the plausibility of artificial neural networks as computational models of human language processing, the adequacy of information-theoretic complexity metrics as linking hypotheses between computational models and human data.

**10** As Bender (2011: fn.18) herself correctly pointed out, given the recent trend in multilingual models and low-resource languages in the NLP community, the situation might have improved over the past years.

across languages remains to be empirically addressed. Therefore, comparative perspectives need to be brought into the field of computational psycholinguistics.

**Table 1:** Language resources naturalistically annotated through human behavioral and neural data (Oseki and Asahara 2020).

| Language Resource | Language | Self-Paced | Eye-Track | EEG | fMRI | Reference |
|---|---|---|---|---|---|---|
| Dundee Corpus | English/<br>French | | ✓(10)<br>✓(10) | | | Kennedy and Pynte (2005) |
| Potsdam Sentence Corpus | German | | ✓(144) | | | Kliegl et al. (2006) |
| Natural Stories Corpus | English | ✓(19) | | | | Futrell et al. (2018) |
| | | | | | ✓(78) | Shain et al. (2019) |
| Ghent Eye-Tracking Corpus (GECO) | English/<br>Dutch | | ✓(14)<br>✓(19) | | | Cop et al. (2017) |
| UCL Corpus | English | ✓(117) | ✓(43) | | | Frank et al. (2013) |
| | | | | ✓(24) | | Frank et al. (2015) |
| Alice Corpus | English | | | ✓(52) | | Brennan and Hale (2019) |
| | | | | | ✓(29) | Brennan et al. (2016) |
| Zurich Cognitive Language Processing Corpus (ZuCo) | English | | ✓(12) | ✓(12) | | Hollenstein et al. (2018) |
| BCCWJ-Eye Track | Japanese | ✓(24) | ✓(24) | | | Asahara et al. (2016) |
| BCCWJ-EEG | Japanese | | | ✓(40) | | Oseki and Asahara (2020) |

## 2.3 Comparative computational psycholinguistics

In order to address the question raised in the subsection above, we advocate the comparative approach to computational psycholinguistics dubbed *comparative computational psycholinguistics*. In comparative computational psycholinguistics, computational models are constructed from symbolic generative models and artificial neural networks developed in NLP and, through the lens of information-theoretic complexity metrics (Hale 2016), evaluated against human behavioral and neural data collected through psycholinguistic experiments just like computational psycholinguistics, but crucially from *comparative* perspectives.

In order to clarify the essence of comparative computational psycholinguistics, here we will review several related approaches proposed in the previous literature, which share some (but not all) of the key features with comparative computational psycholinguistics: (i) computational psycholinguistics, (ii) comparative psycholinguistics, and (iii) computational typology. First of all, as already pointed out in the subsection above, *computational psycholinguistics* constructs and evaluates computational models of human language processing against human behavioral and neural data (Crocker 1996; Lewis 2003; Hale 2017), but lacks the comparative perspectives, hence the strong bias towards the European languages. Second, *comparative psycholinguistics* elucidates human language processing from comparative perspectives (Grillo and Costa 2014; Chacón et al. 2016), but through psycholinguistic experiments with experimental manipulations performed over experimental stimuli, not computational models. Finally, the computational approach to linguistic typology called *computational typology* has recently emerged in the NLP community which employs both computational models and massively comparative perspectives (Ackerman and Malouf 2013; Futrell, Levy, and Gibson 2020), but investigates linguistic universals, not human language processing. These related approaches taken together, comparative computational psycholinguistics can be regarded as the new interdisciplinary approach to human language processing from both computational and comparative perspectives. Related approaches with comparative computational psycholinguistics are summarized in Table 2.

**Table 2:** Related approaches with comparative computational psycholinguistics.

| | **Psycholinguistic** | **Computational** | **Comparative** |
|---|---|---|---|
| Computational psycholinguistics | ✓ | ✓ | |
| Comparative psycholinguistics | ✓ | | ✓ |
| Computational typology | | ✓ | ✓ |
| Comparative computational psycholinguistics | ✓ | ✓ | ✓ |

To recapitulate, this section reviewed the method and the problem with computational psycholinguistics and then proposed comparative computational psycholinguistics, in systematic comparisons with related approaches proposed in the previous literature. In the next two sections, we will present the results of modeling *hierarchical syntactic structure* (Section 3; Yoshida, Noji, and Oseki 2021) and *cue-based memory retrieval* (Section 4; Kuribayashi et al. 2021), where computational models are constructed and evaluated against human behavioral data from comparative perspectives, with particular emphasis on typologically different languages such as English and Japanese. Specifically, hierarchical syntactic structure and cue-based memory retrieval will be modeled with Recurrent Neural Network Grammars (Dyer et al. 2016) and Transformer architectures (Vaswani et al. 2017), respectively.

## 3 Modeling hierarchical syntactic structure

Linguistic theories assume that the grammar represents sentences as *hierarchical syntactic structure*, not just linear word sequence (Chomsky 1957; Everaert et al. 2015). Accordingly, given the Competence Hypothesis that the relationship between the grammar and the parser is maximally transparent (Chomsky 1965), psycholinguistic theories also hypothesize that the parser processes sentences hierarchically (i.e. building hierarchical structures), not just sequentially (i.e. tracking word sequences).[11]

In sharp contrast with those linguistic and psycholinguistic theories, the NLP community has been dominated by recurrent neural networks (RNNs; Elman 1990) with the *recurrence* mechanism which propagates information through time and, despite the lack of explicit hierarchical structures, successfully processes sentences. One of the reasonable hypotheses for this success is that RNNs inductively learn hierarchical representations and *implicitly* represent sentences as hierarchical structures, as evidenced by acceptability judgment experiments where RNNs can capture long-distance dependencies like subject-verb agreement (Linzen, Dupoux, and Goldberg 2016; Warstadt, Singh, and Bowman 2019).

Nevertheless, the previous literature has also implemented the RNN architectures that *explicitly* represent sentences as hierarchical structures and, interestingly, demonstrated that those RNN architectures with syntactic supervision outperform RNNs in explaining long-distance dependencies (Kuncoro et al. 2018; Wilcox et al. 2019) and even human neural responses (Hale et al. 2018). The representative RNN architecture with syntactic supervision is *Recurrent Neural Network Grammars* (RNNGs; Dyer et al. 2016), a deep generative model which models not only sentences but also their hierarchical structures. Specifically, RNNGs adopt the stack LSTM (Dyer et al. 2015), an augmentation of Long Short-Term Memory networks (LSTMs; Hochreiter and Schmidhuber 1997) originally developed for dependency parsing, and estimate probability distributions over three parsing actions as defined in (2), where the composition function of REDUCE is bidirectional LSTMs. The architecture of RNNGs is summarized in Figure 2.

---

**11** Since the failure of the Derivational Theory of Complexity (DTC) that processing complexity is a function of the number of derivational steps to generate the sentences in question (Miller and Chomsky 1963; Fodor, Bever, and Garrett 1974), psycholinguistic theories have ramified into those with and without the Competence Hypothesis, the latter of which claim that human language processing is insensitive to hierarchical structures (Frank and Bod 2011; Frank, Bod, and Christiansen 2012).

(2) Parsing actions of Recurrent Neural Network Grammars (Dyer et al. 2016):
   a. NT: introduce nonterminal symbols (e.g. NP, VP)
   b. GEN: generate terminal symbols (e.g. *the*, *hungry*, *cat*)
   c. REDUCE: compose symbols into phrases via composition function

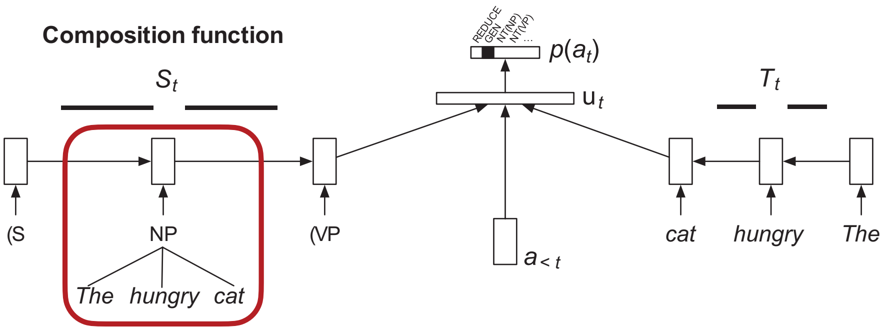



**Figure 2:** Architecture of Recurrent Neural Network Grammars (Dyer et al. 2016).[12]

However, RNNGs have been evaluated only in English, so that whether hierarchical syntactic structure universally makes computational models more human-like remains to be empirically verified. Moreover, the vanilla RNNG implemented by Dyer et al. (2016) adopts the top-down parsing strategy, but given the consensus in the parsing literature (Abney and Johnson 1991; Resnik 1992) that the top-down parsing strategy works effectively for right-branching structures instantiated by head-initial languages like English, the performance of RNNGs might have been overestimated by the accidental match between top-down parsing and head directionality. Therefore, in order to assess the robustness of RNNGs across languages, we should evaluate RNNGs with both top-down and left-corner parsing strategies against head-final languages like Japanese.

**12** The vanilla RNNG implemented in DyNet by Dyer et al. (2016) estimates probability distributions over parsing actions collectively through stack ($S_t$), output buffer ($T_t$), and history of actions ($a_{<t}$), but the recent batched RNNG implemented in PyTorch by Noji and Oseki (2021) used in this chapter adopts the stack-only RNNG (Kuncoro et al. 2017) which estimates probability distributions over parsing actions based solely on stack ($S_t$).

## 3.1 Methods

The experiments follow the pipeline of computational psycholinguistics in Figure 1. First, three computational models were constructed from symbolic generative models and artificial neural networks developed in NLP and trained on the NINJAL Parsed Corpus of Modern Japanese (NPCMJ): Long Short-Term Memory (LSTM; Hochreiter and Schmidhuber 1997) and Recurrent Neural Network Grammars (RNNGs; Dyer et al. 2016) with top-down (Top-down RNNG) and left-corner (Left-corner RNNG) parsing strategies. Second, human behavioral data were collected through psycholinguistic experiments to be predicted with those computational models: BCCWJ-EyeTrack (Asahara, Ono, and Miyamoto 2016). Finally, three computational models and human behavioral data were compared to evaluate which computational model most successfully predicts human behavioral data, through information-theoretic complexity metrics like surprisal (Hale 2001; Levy 2008) and linear mixed-effects regression models (Baayen, Davidson, and Bates 2008). Following Goodkind and Bicknell (2018), perplexity (PPL; how successfully computational models predict next words) and psychometric predictive power (PPP; how successfully computational models predict human data relative to the baseline model with control variables like length and frequency) were adopted as the evaluation metrics.

## 3.2 Results

The results of modeling hierarchical syntactic structure are summarized in Figure 3 (Yoshida, Noji, and Oseki 2021). There are three important observations. First, RNNGs with both top-down and left-corner parsing strategies generally outperform LSTMs, demonstrating that hierarchical syntactic structure universally makes computational models more human-like (Kuncoro et al. 2018; Wilcox et al. 2019). Second, among those RNNGs, left-corner parsing strategies outperform top-down parsing strategies, indicating that optimal parsing strategies may diverge with respect to head directionality (Abney and Johnson 1991; Resnik 1992). Finally, there seems to be a linear correlation between perplexity and psychometric predictive power for RNNGs, whereas this linear correlation does not hold for LSTMs, contradicting the established conclusion (Goodkind and Bicknell 2018).

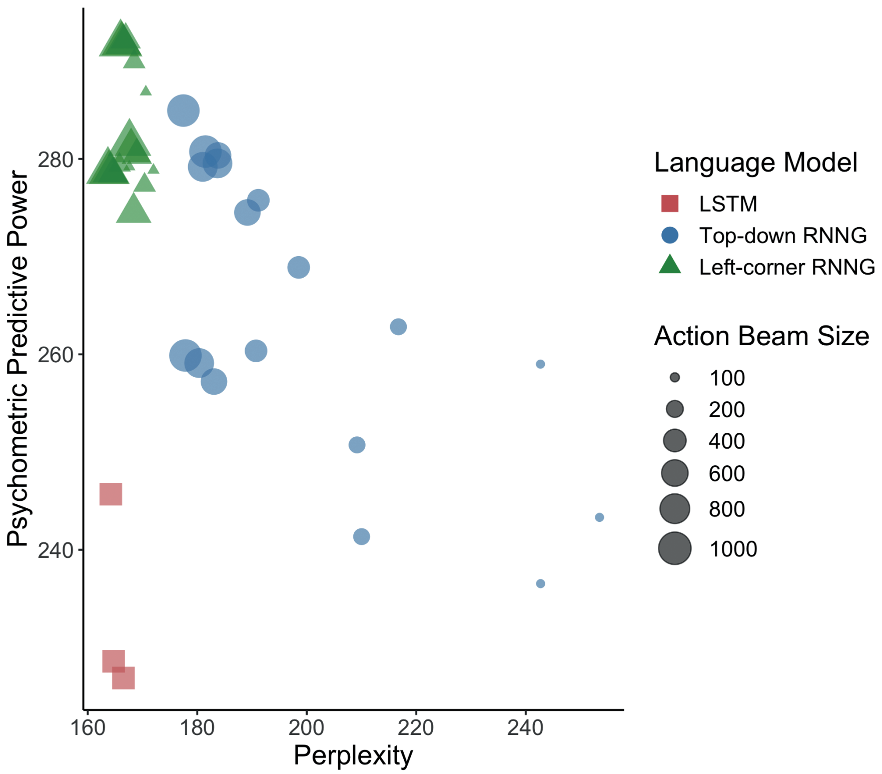



**Figure 3:** Results of modeling hierarchical syntactic structure (Yoshida, Noji, and Oseki 2021).[13]

## 3.3 Summary and discussion

In summary, this section presented the results of modeling *hierarchical syntactic structure* with Recurrent Neural Network Grammars (Dyer et al. 2016), demonstrating that hierarchical syntactic structure universally makes computational models more human-like, while optimal parsing strategies may vary with respect to head directionality (Yoshida, Noji, and Oseki 2021). The main results are summarized below.

– Hierarchical syntactic structure universally makes computational models more human-like (Kuncoro et al. 2018; Wilcox et al. 2019).

13 See Yoshida, Noji and Oseki (2021) for further experimental manipulations on action beam size and parsing accuracy, which we omitted due to space limitations.

- Optimal parsing strategies may vary with respect to head directionality (Abney and Johnson 1991; Resnik 1992).
- Perplexity and psychometric predictive power are linearly correlated for RNNGs, but not for LSTMs (Goodkind and Bicknell 2018).

In the next section, assuming that human language processing is not only *expectation*-based but also *memory*-based, we provide the results of modeling *cue-based memory retrieval* with Transformer architectures (Vaswani et al. 2017; Merkx and Frank 2021).

## 4 Modeling cue-based memory retrieval

In addition to hierarchical syntactic structure discussed in the previous section, the memory mechanism called *cue-based memory retrieval* has been proposed in the psycholinguistic literature (Lewis and Vasishth 2005; Lewis, Vasishth, and Van Dyke 2006).[14] For example, in long-distance dependencies like subject-verb agreement, subjects should be stored in memory and selectively retrieved at verbs through such retrieval cues as number and person features of the subjects in order to correctly inflect the verbs.

In the same vein, the memory mechanism has also been implemented into artificial neural networks in the NLP literature. Specifically, RNNs involve the *recurrence* mechanism (Elman 1990) which propagates information through time but cannot "remember" information for long time, namely the vanishing gradient problem, while LSTMs employ the *gate* mechanism which not only "remembers" but also effectively "forgets" information through time, capturing long-distance dependencies like subject-verb agreement (Linzen, Dupoux, and Goldberg 2016). More recently, building on various insights from machine translation, Transformer architectures have dominated the NLP community (Vaswani et al. 2017) and achieved the state-of-the-art performance on various downstream tasks. The key innovation of Transformer architectures is the *attention* mechanism which dispenses with the recurrence and gate mechanisms and selectively attends previous information, and importantly has been cognitively interpreted as a computational model of human cue-based memory retrieval (Merkx and Frank 2021). The architecture of Transformers is summarized in Figure 4, in comparison with RNNs.

**14** Cue-based memory retrieval has been computationally implemented within the framework of the cognitive architecture called *Adaptive Control of Thought–Rational* (ACT–R; Anderson 1983, 2007).

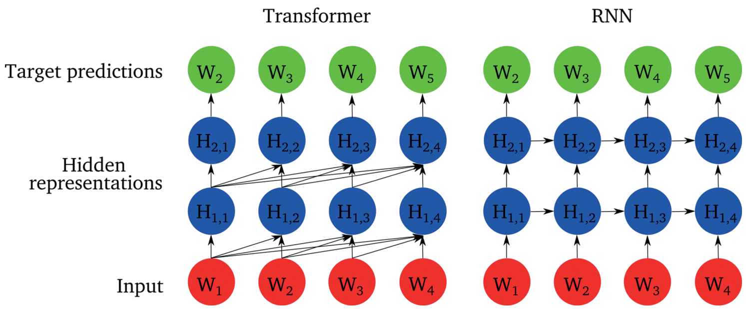



**Figure 4:** Architecture of Transformers (Vaswani et al. 2007; Merkx and Frank 2021).

However, while Transformer architectures seem to be cognitively plausible for European languages like English with various long-distance dependencies (e.g. subject-verb agreement, *wh*-movement) and the so-called *locality* effect (i.e. local dependencies are less costly), whether this established conclusion can be transported to typologically different languages is not self-evident. That is, Transformer architectures might be too powerful for Asian languages like Japanese with few long-distance dependencies (e.g. no subject-verb agreement, no *wh*-movement) and the opposite *anti-locality* effects (i.e. non-local dependencies are less costly). Therefore, in order to assess the cognitive plausibility of Transformer architectures as a computational model of cue-based memory retrieval, we should evaluate Transformer architectures against those languages like Japanese.

## 4.1 Methods

The experiments here also follow the same pipeline of computational psycholinguistics in Figure 1. First, four computational models were constructed from symbolic generative models and artificial neural networks developed in NLP and trained on Wikipedia articles in both English and Japanese: *n*-gram models (`N-gram`, where $n = \{3, 4, 5\}$), Long Short-Term Memory (`LSTM`; Hochreiter and Schmidhuber 1997), and Transformer architectures (Vaswani et al. 2017) with large (`Trans-lg`) and small (`Trans-sm`) numbers of hyperparameters. Second, just like the previous section, human behavioral data were collected through psycholinguistic experiments to be predicted with those computational models: Dundee Corpus for English (Kennedy and Pynte 2005) and BCCWJ-EyeTrack for Japanese (Asahara, Ono, and Miyamoto 2016). Finally, four computational models and human behavioral data were compared to evaluate which computational model most successfully predicts human behavioral data, through information-theoretic complexity metrics like surprisal

(Hale 2001; Levy 2008) and linear mixed-effects regression models (Baayen, Davidson, and Bates 2008). The evaluation metrics were the same as the previous section.

## 4.2 Results

The results of modeling cue-based memory retrieval are summarized in Figure 5 (Kuribayashi et al. 2021). For the Dundee Corpus in English, the correlation between perplexity and psychometric predictive power was negative, corroborating the established conclusion that Transformer architectures are cognitively plausible for European languages like English with various long-distance dependencies. In contrast, for the BCCWJ-EyeTrack in Japanese, while the correlation between perplexity and psychometric predictive power was negative with perplexity > 400, the correlation became positive with perplexity < 400, suggesting that Transformer architectures might be too powerful for Asian languages like Japanese with few long-distance dependencies.

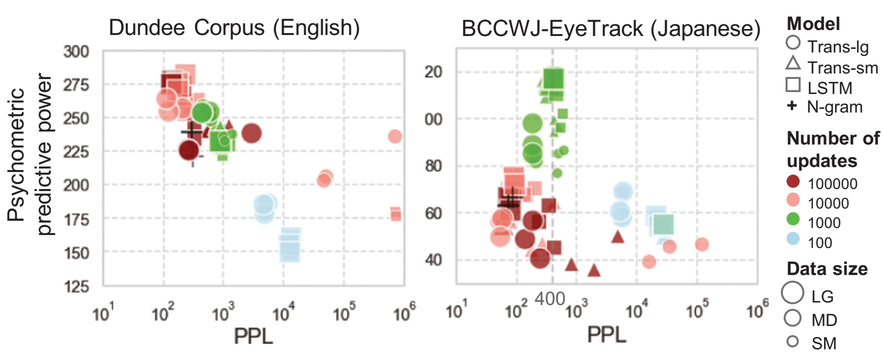



**Figure 5:** Results of modeling cue-based memory retrieval (Kuribayashi et al. 2021).[15]

If the attention mechanism of Transformer architectures is too powerful for those languages with few long-distance dependencies which do not require such “skilled” cue-based memory retrieval, the prediction is that context limitations make Transformer architectures only accessible to local information and thus more humanlike in those languages with few long-distance dependencies. The results of context limitations are summarized in Figure 6 (Kuribayashi et al. 2022), where surprisal

15 See Kuribayashi et al. (2021) for further experimental manipulations on number of updates and data size, which we omitted due to space limitations.

of Transformer architectures is computed given $n$–1 previous words as in $n$-gram models. Interestingly, when the same architecture is compared (e.g., `GPT2-xs-Wiki`), while psychometric predictive power did not change significantly through context limitations in English, context limitations can render Transformer architectures more human-like in Japanese.

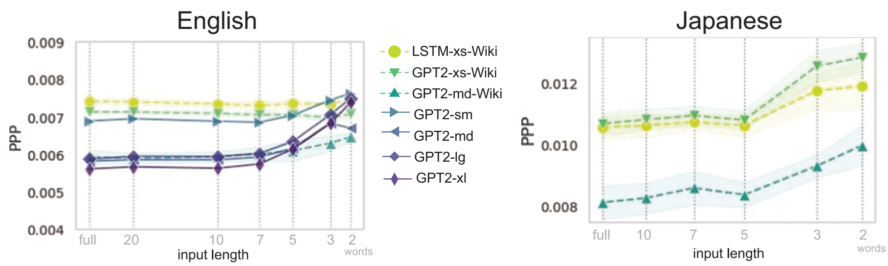



**Figure 6:** Results of context limitations (Kuribayashi et al. 2022).[16]

## 4.3 Summary and discussion

In summary, this section provided the results of modeling *cue-based memory retrieval* with Transformer architectures (Vaswani et al. 2017; Merkx and Frank 2021), suggesting that Transformer architectures are too powerful for those languages with few long-distance dependencies, which can be rendered more human-like through context limitations (Kuribayashi et al. 2021, 2022). The main results are summarized below:

- Transformer architectures are cognitively plausible for European languages like English with various long-distance dependencies (Merkx and Frank 2021).
- In contrast, Transformer architectures are too powerful for Asian languages like Japanese with few long-distance dependencies (Kuribayashi et al. 2021).
- Context limitations can render Transformer architectures more human-like in Japanese (Kuribayashi et al. 2022).

Now several theoretical implications will be discussed in light of the results above. First, as pointed out above, while human language processing has traditionally been assumed to be both expectation-based ("look-ahead" prediction of next words)

16 Note that the $y$-axes in Figures 3, 5, 6 all represent psychometric predictive power of the computational models, but their scales are not directly comparable due to various differences in training data, test data, computational models, among others.

and memory-based ("look-behind" retrieval of previous words), hence the radical dichotomy between expectation-based and memory-based theories in psycholinguistics, those two theories should not be mutually exclusive, merely reflecting different aspects of human language processing (Demberg and Keller 2008; Futrell, Gibson, and Levy 2020). Second, the Transformer architectures with context limitations can be regarded as a hybrid computational model of expectation-based and memory-based theories, suggesting the possibility that cue-based memory retrieval itself is universal, while what counts as "cue" is parametrized across languages.

# 5 Conclusion

To summarize, this chapter advocated the comparative approach to computational psycholinguistics dubbed *comparative computational psycholinguistics*, which constructs and evaluates computational models of human language processing from comparative perspectives. Specifically, we presented the results of modeling *hierarchical syntactic structure* with Recurrent Neural Network Grammars (Dyer et al. 2016), demonstrating that hierarchical syntactic structure universally makes computational models more human-like, though optimal parsing strategies may vary with respect to head directionality (Yoshida, Noji, and Oseki 2021). Then, we provided the results of modeling *cue-based memory retrieval* with Transformer architectures (Vaswani et al. 2017; Merkx and Frank 2021), suggesting that Transformer architectures are too powerful for those languages with few long-distance dependencies, which can be rendered more human-like through context limitations (Kuribayashi et al. 2021, 2022). For future directions, this comparative approach to computational psycholinguistics should be extended to (i) typologically more diverse languages like Kaqchikel Maya and Tongan (Koizumi et al. 2014) and (ii) human neural data like EEG/MEG and fMRI (Hale et al. 2022).

In conclusion, we believe that comparative computational psycholinguistics will be a promising approach to human language processing from both computational and comparative perspectives, towards machines that process natural languages like humans.

# References

Abney, Steven & Mark Johnson. 1991. Memory requirements and local ambiguities of parsing strategies. *Journal of Psycholinguistic Research* 20. 233–250.
Ackerman, Farrell & Robert Malouf. 2013. Morphological organization: The low conditional entropy conjecture. *Language* 89. 429–464.
Anderson, John. 1983. *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.
Anderson, John. 2007. *How Can the Human Mind Occur in the Physical Universe?* New York: Oxford University Press.
Asahara, Masayuki, Hajime Ono, & Edson Miyamoto. 2016. Reading-time annotations for "Balanced Corpus of Contemporary Written Japanese". In Yuji Matsumoto & Rashmi Prasad (eds.), *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics*, 684–694. Osaka: The COLING 2016 Organizing Committee.
Baayen, Harald, Douglas Davidson, & Douglas Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59. 390–412.
Bender, Emily. 2011. On achieving and evaluating language-independence in NLP. *Linguistic Issues in Language Technology* 6. 1–26.
Brennan, Jonathan. 2016. Naturalistic sentence comprehension in the brain. *Language and Linguistics Compass* 10. 299–313.
Brennan, Jonathan & John Hale. 2019. Hierarchical structure guides rapid linguistic predictions during naturalistic listening. *PLOS ONE* 14. e0207741.
Chacón Dustin, Mashrur Imtiaz, Shirsho Dasgupta, Sikder Murshed, Mina Dan, & Colin Phillips. 2016. Locality and word order in active dependency formation in Bangla. *Frontiers in Psychology* 7. 1235.
Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.
Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
Crocker, Matthew. 1996. *Computational Psycholinguistics: An Interdisciplinary Approach to the Study of Language*. Dordrecht: Kluwer Academic Publishers.
Demberg, Vera & Frank Keller. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition* 109. 193–210.
Dyer, Chris, Miguel Ballesteros, Wang Ling, Austin Matthews, & Noah Smith. 2015. Transition-based dependency parsing with stack long short-term memory. In Chengqing Zong & Michael Strube (eds.), *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, 334–343. Pittsburgh, PA: Association for Computational Linguistics.
Dyer, Chris, Adhiguna Kuncoro, Miguel Ballesteros, & Noah Smith. 2016. Recurrent neural network grammars. In Kevin Knight, Ani Nenkova, & Owen Rambow (eds.), *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 199–209. Pittsburgh, PA: Association for Computational Linguistics.
Elman, Jeffrey. 1990. Finding structure in time. *Cognitive Science* 14. 179–211.
Everaert, Martin, Marinus Huybregts, Noam Chomsky, Robert Berwick, & Johan Bolhuis. 2015. Structures, Not Strings: Linguistics as Part of the Cognitive Sciences. *Trends in Cognitive Sciences* 19. 729–743.
Fodor, Jerry, Thomas Bever, & Merrill Garrett. 1974. *The Psychology of Language*. New York: McGraw-Hill.
Frank, Stefan & Rens Bod. 2011. Insensitivity of the human sentence-processing system to hierarchical structure. *Psychological Science* 22. 829–834.

Frank, Stefan, Rens Bod, & Morten Christiansen. 2012. How hierarchical is language use? *Proceedings of the Royal Society B: Biological Sciences* 279. 4522–4531.

Futrell, Richard, Edward Gibson, & Roger Levy. 2020. Lossy-context surprisal: An information-theoretic model of memory effects in sentence processing. *Cognitive Science* 44. e12814.

Futrell, Richard, Roger Levy, & Edward Gibson. 2020. Dependency locality as an explanatory principle for word order. *Language* 96. 371–413.

Goldberg, Yoav. 2017. *Neural Network Methods for Natural Language Processing*. Vermont: Morgan & Claypool Publishers.

Goodkind, Adam & Klinton Bicknell. 2018. Predictive power of word surprisal for reading times is a linear function of language model quality. In Asad Sayeed, Cassandra Jacobs, Tal Linzen, & Marten van Schijndel (eds.), *Proceedings of the 8th Workshop on Cognitive Modeling and Computational Linguistics*, 10–18. Salt Lake City: Association for Computational Linguistics.

Grillo, Nino & João Costa. 2014. A novel argument for the universality of parsing principles. *Cognition* 133. 156–187.

Hale, John. 2001. A probabilistic early parser as a psycholinguistic model. *Proceedings of the Second Conference of the North American Chapter of the Association for Computational Linguistics*, 159–166. Pittsburgh, PA: Association for Computational Linguistics.

Hale, John. 2006. Uncertainty about the rest of the sentence. *Cognitive Science* 30. 609–642.

Hale, John. 2016. Information-theoretical complexity metrics. *Language and Linguistics Compass* 10. 397–412.

Hale, John. 2017. Models of human sentence comprehension in computational psycholinguistics. In Mark Aronoff et al. (eds.), *Oxford Research Encyclopedia of Linguistics*. Retrieved 23 Dec. 2022, from https://oxfordre.com/linguistics/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-377.

Hale, John, Luca Campanelli, Jixing Li, Shohini Bhattasali, Christophe Pallier, & Jonathan Brennan. 2022. Neurocomputational models of language processing. *Annual Review of Linguistics* 8. 427–446.

Hale, John, Chris Dyer, Adhiguna Kuncoro, & Jonathan Brennan. 2018. Finding syntax in human encephalography with beam search. In Iryna Gurevych & Yusuke Miyao (eds.), *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, 2727–2736. Pittsburgh, PA: Association for Computational Linguistics.

Hochreiter, Sepp & Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9. 1735–1780.

Jurafsky, Daniel & James Martin. 2022. *Speech and Language Processing*, 3rd edn. https://web.stanford.edu/~jurafsky/slp3/.

Kennedy, Alan & Joël Pynte. 2005. Parafoveal-on-foveal effects in normal reading. *Vision Research* 45. 153–168.

Koizumi, Masatoshi, Yoshiho Yasugi, Katsuo Tamaoka, Sachiko Kiyama, Jungho Kim, Juan Esteban Ajsivinac Sian, & Lolmay Pedro Oscar García Mátzar. 2014. On the (non)universality of the preference for subject-object word order in sentence comprehension: A sentence-processing study in Kaqchikel Maya. *Language* 90. 722–736.

Kuncoro, Adhiguna, Miguel Ballesteros, Lingpeng Kong, Chris Dyer, Graham Neubig, & Noah Smith. 2017. What do recurrent neural network grammars learn about syntax? In Mirella Lapata, Phil Blunsom, & Alexander Koller (eds.), *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, 1249–1258. Valencia: Association for Computational Linguistics.

Kuncoro, Adhiguna, Chris Dyer, John Hale, Dani Yogatama, Stephen Clark, & Phil Blunsom. 2018. LSTMs can learn syntax-sensitive dependencies well, but modeling structure makes them better. In Iryna Gurevych & Yusuke Miyao (eds.), *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, 1426–1436. Pittsburgh, PA: Association for Computational Linguistics.

Kuribayashi, Tatsuki, Yohei Oseki, Ana Brassard, & Kentaro Inui. 2022. Context limitations make neural language models more human-like. https://arxiv.org/abs/2205.11463.

Kuribayashi, Tatsuki, Yohei Oseki, Takumi Ito, Ryo Yoshida, Masayuki Asahara, & Kentaro Inui. 2021. Lower perplexity is not always human-like. In Chengqing Zong, Fei Xia, Wenjie Li, & Roberto Navigli (eds.), *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, 5203–5217. Pittsburgh, PA: Association for Computational Linguistics.

Lewis, Richard. 2003. Computational psycholinguistics. In Lynn Nadel (ed.), *Encyclopedia of Cognitive Science*, 787–794. London: Macmillon.

Lewis, Richard & Shravan Vasishth. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* 29. 375–419.

Lewis, Richard, Shravan Vasishth, & Julie Van Dyke. 2006. Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences* 10. 447–454.

Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106. 1126–1177.

Linzen, Tal, Emmanuel Dupoux, & Yoav Goldberg. 2016. Assessing the ability of LSTMs to learn syntax-sensitive dependencies. *Transactions of the Association for Computational Linguistics* 4. 521–535.

Linzen, Tal & Florian Jaeger. 2016. Uncertainty and expectation in sentence processing: Evidence from subcategorization distributions. *Cognitive Science* 40. 1382–1411.

Merkx, Danny & Stefan Frank. 2021. Human sentence processing: Recurrence or attention? In Emmanuele Chersoni, Nora Hollenstein, Cassandra Jacobs, Yohei Oseki, Laurent Prévot, & Enrico Santus (eds.), *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, 12–22. Pittsburgh, PA: Association for Computational Linguistics.

Miller, George & Noam Chomsky. 1963. Finitary models of language users. In R. Duncan Luce (ed.), *Handbook of Mathematical Psychology*, 419–491. New York: Wiley.

Noji, Hiroshi & Yohei Oseki. 2021. Effective batching for recurrent neural network grammars. In Chengqing Zong, Fei Xia, Wenjie Li, & Roberto Navigli (eds.), *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, 4340–4352. Pittsburgh, PA: Association for Computational Linguistics.

Oseki, Yohei & Masayuki Asahara. 2020. Design of BCCWJ-EEG: Balanced corpus with human electroencephalography. In Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, & Stelios Piperidis (eds.), *Proceedings of the 12th Language Resources and Evaluation Conference*, 189–194. Marseille: European Language Resources Association.

Resnik, Philip. 1992. Left-corner parsing and psychological plausibility. *Proceedings of the 14th Conference on Computational Linguistics*, 191–197. Stroudsburg, PA: Association for Computational Linguistics.

Taniguchi, Tadahiro, Daichi Mochihashi, Takayuki Nagai, Satoru Uchida, Naoya Inoue, Ichiro Kobayashi, Tomoaki Nakamura, Yoshinobu Hagiwara, Naoto Iwahashi, & Tetsunari Inamura. 2019. Survey on frontiers of language and robotics. *Advanced Robotics* 33. 700–730.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan Gomez, Łukasz Kaiser, & Illia Polosukhin. 2017. Attention is all you need. In Ulrike von Luxburg, Isabelle Guyon, Samy Bengio, Hanna Wallach, & Rob Fergus (eds.), *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 5998–6008. Red Hook, NY: Curran Associates Inc.

Warstadt, Alex, Amanpreet Singh, & Samuel Bowman. 2019. Neural network acceptability judgments. *Transactions of the Association for Computational Linguistics* 7. 625–641.

Wilcox, Ethan, Peng Qian, Richard Futrell, Miguel Ballesteros, & Roger Levy. 2019. Structural supervision improves learning of non-local grammatical dependencies. In Jill Burstein, Christy Doran, & Thamar Solorio (eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 3302–3312. Pittsburgh, PA: Association for Computational Linguistics.

Yoshida, Ryo, Hiroshi Noji, & Yohei Oseki. 2021. Modeling human sentence processing with left-corner recurrent neural network grammars. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, & Scott Wen-tau Yih (eds.), *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 2964–2973. Punta Cana: Association for Computational Linguistics.

Yun, Jiwon, Zhong Chen, Tim Hunter, John Whitman, & John Hale. 2015. Uncertainty in processing relative clauses across East Asian languages. *Journal of East Asian Linguistics* 24. 113–148.

# Index




https://doi.org/10.1515/9783110778946-016