AMLaP Asia 2024

Book of Abstracts

Dec. 5th - 7th

National University of Singapore

https://blog.nus.edu.sg/amlapasia2024/

Program at a glance

Dec. 5th, Thursday	Dec. 6th, Friday	Dec. 7th, Saturday
08:30 **Registration**	09:00 **Keynote 3** Debra Titone *McGill University*	09:00 **Keynote 5** Ming Xiang *University of Chicago*
09:00 **Welcome** 09:15 **Keynote 1** Bob McMurray *University of Iowa*	10:00 **Tea Break & Poster session 3**	10:00 **Tea Break & Poster session 5**
10:15 **Tea Break & Poster session 1**	10:45 **Oral Session 3-1** *Sentence Comprehension and Language Cognition* Session Chair: Zheng Shen	10:45 **Oral Session 4-1** *Sentence Processing* Session Chair: Zheng Shen
11:00 **Oral Session 1** *Speech Recognition and Production* Session Chair: Aine Ito	12:25 **Lunch Break**	12:25 **Lunch Break**
12:20 **Lunch Break**	13:30 **Oral Session 3-2** *Sentence Comprehension and Language Cognition* Session Chair: Nick Huang	13:30 **Oral Session 4-2** *Sentence Processing* Session Chair: Nick Huang
13:30 **Oral Session 2** *Semantic Processing* Session Chair: Cynthia Siew	14:30 **Tea Break & Poster session 4**	14:30 **Tea Break & Poster session 6**
14:10 **Tea Break & Poster session 2**	15:15 **Keynote 4** Suzy Styles *Nanyang Technological University*	15:15 **Oral Session 4-3** *Sentence Processing* Session Chair: Aine Ito
15:35 **Keynote 2** Sachiko Kinoshita *Macquarie University*	17:00 **Conference Dinner**	16:15 **Closing**

Index

Welcome

Weclome to AMLaP Asia 2024!

Architectures and Mechanisms for Language Processing Asia was created as the Asian venue for presenting and discussing research on interdisciplinary psycholinguistic research. It aims to bring together experimental, computational, and theoretical perspectives on the cognitive architectures and mechanisms that underlie any aspect of human language processing.

The first AMLaP Asia conference was held in 2018 at the University of Hyderabad. The second AMLaP Asia conference was held in 2023 at the Chinese University of Hong Kong.

This year's AMLaP Asia is held in person, on December 5th – 7th, 2024, organised by the Faculty of Arts and Social Sciences, National University of Singapore.

Singapore is a remarkable fusion of diverse cultures, economic prosperity, and a myriad of attractions. The country reflects its multicultural essence with significant Chinese, Malay, Indian, and other ethnic communities.

We hope that you will have a pleasant and memorable stay in Singapore!

AMLaP Asia 2024 Organizing Committee
FACULTY: Aine Ito, Nick Huang, Zheng Shen, Cynthia Siew
STUDENTS: Daiwen Gong, Danning Sun

Student Volunteers
Beth Chan Hui Fang, Hui Chi Kin Ethan, Jiang Zehua, Li Yumeng,
Lim Wee Ping, Qian Zongyu, Suri Gurnoor Kaur, Yang Mingyuan

Reviewers Acknowledgement

We would like to extend our sincere gratitude to all reviewers for their valuable feedback and contributions, which have greatly enhanced the quality of AMLaP Asia 2024.

Agnieszka Konopka	Chao Sun	Emmanuel Chemla	Lijuan Liang	Robert Hartsuiker	Tommi Leung
Aiana Hanulikova	Charles Clifton	Esther Janse	Margreet Vogelzang	Robert Kluender	Tongquan Zhou
Alan Garnham	Charles Lin	Evan Kidd	Maria Lobo	Ruth Corps	Umesh Patil
Alexander Göbel	Charlotte Hauser	Francois Rigalleau	Martin Corley	Saveria Colonna	Vadim Kimmelman
Alice Foucart	Chi Zhang	Fuyun Wu	Martin Pickering	Savithry Namboodiripad	Wenting Xue
Anna Banaszkiewicz	Chiara Gambi	Hannah Rohde	Masako Hirotani	Sendy Caffarra	Wing Yee Chow
Anna Laurinavichyute	Chin-lung Yang	Helene Kreysa	Mikel Santesteban	Serene Siow	Xiaohu Yang
Anuenue Kukona	Chris Cummins	Ian Cunnings	Ming Xiang	Shelia Kennison	Xuemei Chen
Aya Meltzer-Asscher	Christoph Scheepers	Julien Mayor	Mohammad Momenian	Shinri Ohta	Yang Yang
Barbara Hemforth	Clare Patterson	Kate Messenger	Monsieur Edward Gibson	Shota Momma	Yao Yao
Barbara Mertins	Daniel Parker	Kate Stone	Nan Jiang	Shravan Vasishth	Youngah Do
Benjamin Clarke	Dario Paape	Katy Carlson	Natalia Kartushina	Spencer Caplan	Yuki Kamide
Bo Yao	Dave Kush	Kevin McManus	Paaic Monaghan	Stephen Politzer-Ahles	Zenzi Griffin
Cai Rendong	David Haslett	Kiel Christianson	Patrick Sturt	Suki Yiu	Zhenguang Cai
Camilo Roiguez Ronderos	Doug Roland	Laurel Brehm	Pia Knoeferle	Tessa Warren	Zhu Jingtao
Céline Pozniak	Elsi Kaiser	Leigh Fernandez	Qingqing Qu	Titus von der Malsburg	

General Information

Conference venue

Ngee Ann Kongsi Auditorium

8 College Ave West, Level 2 Education Resource Centre, Singapore 138608

Lunch at NUS

There are a wide variety of food options available near the conference venue, where you can find cuisines from many different cultures. For more details on operating hours and specific food stalls, please refer to https://uci.nus.edu.sg/oca/retail-dining/food-and-beverage-utown/

Wi-Fi Information

- To access NUS guest Wi-Fi users had to fill in the SG phone number, captcha, and submit the form by click on I agree to the term of use and proceed button.
- After submitted the form, users will be redirected to OTP page.
- Users need to fill in the received OTP and click on confirm button.
- After OTP is submitted successfully, users will be redirected to NUS Home Page.

Code of Conduct

At AMLaP Asia 2024, the well-being and safety of all attendees, speakers, and volunteers are paramount. As such, we require everyone to adhere to the code of conduct (https://blog.nus.edu.sg/amlapasia2024/code-of-conduct/), which will be strictly enforced throughout the conference.

Photo Permission

By participating in AMLaP Asia 2024, attendees grant permission to the organizers of the conference to capture photographs during the event. If attendees have concerns about their likeness being used in promotional materials, they may notify the organizing committee in writing (ellbox16@nus.edu.sg) prior to the event, and reasonable efforts will be made to accommodate their request.

Conference Dinner

Dinner will take place at Permata Restaurant in the Gedung Kuning building. The restaurant serves Malay-Nusantara cuisine, and is incidentally nestled in the heart of the Kampung Glam district, a culturally and historically significant area with plenty to do around. We hope that everyone gets the opportunity to explore, and to sample some good local cuisine.

Keynote Speeches

Keynote Speech 1

Towards a Psycholinguistics of All People:
Studies of word recognition across the lifespan, and across differences in language and hearing ability reveal the fundamental dimensions of language processing

Bob McMurray
Dept. of Psychological and Brain Sciences, Dept. of Communication Sciences and Disorders
Dept. of Linguistics, Dept. of Otolaryngology
University of Iowa

Cognitive science seeks to uncover universal language mechanisms that apply to all people. This research emphasizes the “modal” language user: normal hearing, neurotypical, monolingual, young adults. In this framing, differences from the modal listener are explained away as deriving (for example) from a developmentally immature system or a disorder. However, the modal language user is rare: multilingualism is the norm world-wide; at some point, all people are developmentally immature or older; and language, hearing and reading disorders are shockingly common. Consequently, this framing explains away the kinds of language function that may be most common.

This talk argues for an alternative: that there may be no universal solution to basic problems of language (e.g., recognizing words, categorizing speech sounds). Instead, theories need the inherent flexibility to account for variation in how listeners solve these problems. Then, by looking afresh at so-called “populations” with the theoretical and methodological tools of cognitive science we can characterize how different people process and learn language and identify fundamental dimensions along which these mechanisms differ.

I start with a grounding in recent views in the philosophy of medicine. I then illustrate this approach with work on the real-time mechanisms of spoken word recognition. Cognitive science on modal adults has converged on a clear mechanistic description in which word recognition is a product of real-time dynamic competition. It has developed methods like the eye-tracking in the Visual World Paradigm that can precisely characterize this competition as it unfolds over time. I apply this approach to a diverse set of listeners to understand the lawful ways in which this processing changes across the developmental trajectory, and how it is configured (and reconfigured) based on a person’s abilities and listening needs. In doing so, I present new results from large studies of developing children (including those with so-called language and reading disorders), older adults across the lifespan, and recent work on people who use cochlear implants. Along the way I introduce new statistical and methodological tools that can help reveal the fundamental dimensions of processing long which people vary.

As a whole, this work argues that cognitive theories must account for not just the modal language user, but the diversity among people. Even if there is no one way to solve a problem like recognizing words, the goal of cognitive science should be to reveal the identify the fundamental degrees of freedom that allow cognitive models to fully account for the range of language users.

Keynote Speech 2

Morphological processing of compound words in Chinese and English: Insights from masked priming

Sachiko Kinoshita
School of Psychological Sciences, Macquarie University

Two important and welcome shifts in recent research on visual word recognition are the focus away from simplex words to morphologically complex words, and away from the anglo- and Euro-centricity to more diverse languages. Masked priming paradigm is a popular method used to study the earliest stage of visual word recognition, and in recent decades, it has been used fruitfully in studying morphologically complex words. In this talk, I present our recent work using the various masked priming manipulations (e.g., compound priming of constituent targets, transposed-constituent priming) examining the processing of compound words written in English and Chinese, two languages with very different writing systems. I will discuss some language-universal processes, as well as cross-language differences stemming from the unique properties of alphabetic and logographic writing systems.

Keynote Speech 3

Language(s) in Context: A Systems View of Multilingualism

Debra Titone
Department of Psychology, Montreal Bilingualism Initiative (MoB - https://www.mcgill.ca/mobi/), Centre for Research on Brain, Language, & Music (CRBLM - https://crblm.ca/), McGill University

Most people around the globe communicate daily using more than one language. Moreover, global patterns of language use are driven by a variety of historic and current social forces that can operate at an individual level. This has led some to theorize that multilingualism possesses the hallmarks of a complex cognitive system (e.g., Atkinson et al., 2016; de Bot, Lowie, and Verspoor, 2007; Titone & Tiv, 2022; Tiv et al., 2022), which is amenable to a socioecological theoretical frame. In this talk, I review empirical work from my laboratory that works to quantify social patterns of language use to assess their impact on both language processing and cognition more generally. This work is guided by a Systems View of Bilingualism (e.g., Titone & Tiv, 2022) that encourages an inclusive understanding of multilingual experience that bridges individual, interpersonal, and socioecological levels.

Keynote Speech 4

Into the language jungle: Translanguaging, conversational microstructures and measures of complexity in parent-child speech from multilingual Singapore

Suzy Styles
Psychology, School of Social Sciences, Nanyang Technological University

Research on language development typically centres the experience of children growing up in majority monolingual contexts – much less is known about how families with multilingual resources share their languages with their children, and how these variations influence emerging language skills and systems. Multilingual Singapore is a nation with four official languages, where bilingualism is considered a 'cornerstone' of the education system, and 85% of the population report being bi- or multilingual. In this context, each family has their own unique pattern of linguistic resources, and unique ways of coordinating those resources with children under the age of school entry. This talk will provide an overview of our investigations into multilingual language use in-the-wild, and the development of measurements and metrics designed to capture complexity in early language exposure in a multilingual context.

The Talk Together Study, is a large-scale micro-longitudinal study designed to create a corpus of naturalistic parent-child talk in Singapore. Over 140 families with children ranging from 8 months and 4 years of age participated in a micro-longitudinal study including a task where parents narrated a wordless storybook to their children. The wordless picturebook 'What a Scary Storm!' was designed to allow parents to use any language resources in any combination. Our large team of transcribers created rich, linguistically-informed transcriptions of over 400 audio recordings, providing a unique lens into patterns of language use in the multilingual community. In our first wave of analysis, we find that translanguaging is a normal linguistic behaviour in this context – with almost all parents blending resources from multiple languages within a single utterance at least some of the time. We also observed frequent use of local vocabulary that does not easily classify into the 'named languages' of the region. We present novel metrics and tools for characterising conversational microstructures that reflect turn taking, grammatical complexity, and specific vocabulary uses. The talk will also discuss the challenges of creating such metrics and developing tools for use with organically diverse multilingual samples, including structural biases introduced by AI tools automating some of our metrics.

Just as biodiversity is higher a jungle than in a cultivated garden, these findings reinforce the need for studies which collect organically diverse language samples, and to produce high-quality linguistically-informed human annotations for such datasets. Artificially 'tidy' datasets not only constrain our understanding of language use, and language development, but they constrain the tools we develop from them – thereby increasing biases against underrepresented populations.

Keynote Speech 5

Structuring discourse in sentence comprehension

Ming Xiang
Department of Linguistics, University of Chicago

There is broad consensus that language comprehension operates in a highly context-dependent manner. However, the precise characterization of context and the mechanisms through which context exerts its effects remain open questions. The Question under Discussion (QUD) framework provides an influential approach for modeling the general structure of discourse context. According to this framework, discourse context progresses as a goal-driven process, with interlocutors engaging in cooperative exchanges to resolve relevant issues and, in essence, respond to pertinent QUDs. The relevance of an utterance to a communicative context is determined by its relation to the QUD.

This talk examines the effect of QUD on language processing. The first part of the talk focuses on the variability of pragmatic inferences. Using scalar implicature variability as a case study, I will show how variability arises because pragmatic inferences are generated relative to a specific QUD. Beyond offline pragmatic reasoning, the second part of the talk demonstrates how online language processing is also sensitive to QUD-based discourse structure. In particular, working memory encoding and retrieval of linguistic information during online processing are closely tuned to the incremental updating of discourse representations. While arguing that the QUD approach is highly productive in modeling the effects of discourse structure on language processing, I will also identify potential challenges and future directions for this approach.

Oral Presentations

OS.1.01

Recognition of words with segmental–prosodic incongruity in Japanese: Evidence from repetition priming

Terumichi Ariga (The University of Tokyo)
ariga.terumichi@gmail.com

Lexical prosody plays a crucial role in spoken word recognition (e.g., Cutler et al., 1997; McQueen & Dilley, 2021). In Japanese, lexically specified pitch accent information helps narrowing down the potential candidates in lexical access (Sekiguchi & Nakajima, 1999). This in turn implies that lexically inappropriate pitch accent inhibits the process of recognition (Minematsu & Hirose, 1995). Then, how do listeners recognize spoken words when there is segmental–prosodic incongruity in the word? For example, a production *ta'nsu* HLL is ambiguous as to whether it is a prosodic mispronunciation of *tansu* LHH "wardrobe" or a segmental mispronunciation of *da'nsu* HLL "dance" (H or L indicates high or low tone assigned to morae). When the initial consonant is acoustically equivocal in terms of voicing (whether /t/ or /d/), listeners perceive the segment so that the spoken word becomes a meaningful word based on lexical prosody (Ariga & Matsubara, 2023). However, it is still unknown that how listeners interpret the word if segments and lexical prosody blatantly contradict in terms of lexical status. To address the question, we conducted a cross-modal repetition priming experiment.

We used 12 pairs of words that contrast both segmentally and prosodically, such as *tansu* LHH "wardrobe" and *da'nsu* HLL "dance" for experimental materials. In each trial, an auditory prime was presented, and a visual target word for a lexical decision was presented at the offset of the auditory prime (ISI = 0 ms). The design of conditions is shown in Table 1. The auditory prime was *congruent* with the target, included *prosodic* violation (mispronounced pitch accent), included *segmental* violation (mispronounced initial consonant), or was unrelated to the target (*baseline*). Thirty-two native Tokyo Japanese speakers participated in the experiment.

The mean reaction time (RT) for each condition is shown in Figure 1. An LME analysis (Table 2) suggested that the RT in the congruent condition was significantly shorter than that in the baseline condition, suggesting that there was a facilitative repetition priming effect. In contrast, the RT in the prosodic condition was significantly longer than that in the baseline condition, and the RTs did not differ between in the segmental condition and in the baseline condition. These RTs demonstrated that both segmental and prosodic mispronunciation did not facilitate the access to the intended word. However, an inhibitory priming effect occurred only when the type of mispronunciation is prosodic.

These results suggested that even though both segmental and prosodic mispronunciation affects appropriate lexical access to the intended words, mispronunciation of lexical prosody is more serious inhibitor than mispronunciation of segments.

Table 1 Example of experimental materials.

Prime Pattern	Auditory Prime	Visual Target
congruent	*tansu* LHH	箪笥
prosodic	*ta'nsu* HLL	*tansu* LHH
segmental	*dansu* LHH	"wardrobe"
baseline	*sabaku* LHH "desert"	

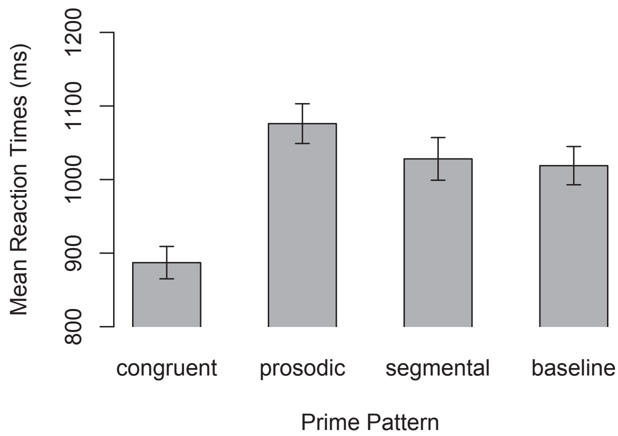


Figure 1 Mean reaction time (ms) and standard error for each condition.

Table 2 Summary of the LME analysis (referent: the baseline condition).

	Estimate	*SE*	*df*	*t*	*p*	
(intercept)	1016.184	36.041	30.166	28.195	< .001	***
congruent	–83.512	29.402	404.759	2.840	.005	**
prosodic	72.214	28.767	399.796	2.510	.012	*
segmental	48.856	30.429	403.237	1.606	.109	

Significance codes: *** $p < .001$, ** $p < .01$, * $p < .05$

References

Ariga, T., & Matsubara, R. (2023). Top-down effects of lexical pitch accent on phonetic categorization in Japanese. In: R. Skarnitzl, & Jan Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 210–214). Guarant International.

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, *40*(2), 141–201.

McQueen, J. M., & Dilley, L. C. (2021). Prosody and spoken-word recognition. In C. Gussenhoven & A. Chen (Eds.), *The Oxford handbook of language prosody* (pp. 509–521). Oxford University Press.

Minematsu, N., & Hirose, K. (1995). Role of prosodic features in the human process of perceiving spoken words and sentences in Japanese. *Journal of the Acoustical Society of Japan (E)*, *16*(5), 311–320.

Sekiguchi, T., & Nakajima, Y. (1999). The use of lexical prosody for lexical access of the Japanese language. *Journal of Psycholinguistic Research*, *28*(4), 439–454.

OS.1.02

The attention control of younger and older adults in Hong Kong (HK): Cantonese tone processing in the forced attention dichotic listening task

Yuqi WANG (ywangqi@connect.ust.hk), Zhen QIN
The Hong Kong University of Science and Technology

Attention control facilitates the processing of the attended targets despite distractions, which can be measured in forced-attention dichotic listening (FADL) task. In Hugdahl et al. [1], participants heard two stimuli contrasting in consonants (e.g., /ba/ vs. /ta/) in each ear and identified a clearer sound in the non-forced (NF) condition. A right ear advantage (REA) was shown as the baseline due to the dominance of the left hemisphere [1]. In forced-left (FL) and forced-right (FR) conditions, the attention had an impact on shifting the baseline to the instructed ear advantage patterns by requiring a report of the instructed left ear or right ear sound. This ear preference change has been found in younger adults, but less pronounced in older adults due to the age-related decline in attention [2]. Different from previous FADL studies, lexical tone (i.e., Cantonese tone) is selected as the stimuli to examine the role of attention. Instead of using consonants prevalent in many languages, using tones that are present in our participants' L1-Cantonese (not in L2-English) aims to eliminate the influence of linguistic knowledge from other acquired languages. Therefore, the study investigates 1) the effects of attention in the FADL task of Cantonese tones by HK younger adults and 2) whether the attention control is different for older adults and younger adults.

In Experiment 1 (Exp 1), young Cantonese speakers aged between 18 and 25 were recruited. As shown in Fig.1, they took a Cantonese tone training to familiarize with the mapping of tones and labels (T1-T6). In the FADL task, participants heard dichotic stimuli of Cantonese tone pairs (Tab 1). They were required to identify the tones from either ear in three conditions (NF, FL, and FR) by pressing keys (1-6). A mixed-effects regression model on accuracy showed an interaction of ear and condition ($F = 582.05$, $p <.001$). Young adults increased the degree of baseline LEA to a larger LEA in the FL condition and shifted the baseline to the opposite REA in the FR condition (Fig 2). In Experiment 2 (Exp 2), old Cantonese speakers aged 60 to 70, with average hearing level of 20dB in 125-2000Hz, were recruited. With the same procedure of Exp 1, the pattern in each condition was also the same ($F = 181.26$, $p <.001$; Fig 2). Importantly, the degree of ear preference change from the NF to FL was larger for younger than older adults ($F = 6.11$, $p =.014$).

The findings support the attention effect on auditory processing [1], which regulated the LEA baseline pattern of Cantonese tones by attentional instructions. Also, it supports an age-related decline in attention control ability in the auditory domain. Notably, this decline was observed only in one forced condition [2], which might suggest that attention control for tone processing can be resistant to aging, possibly because older adults can attend to temporal cues lasting for a longer time [4]. Visual attention tasks will be added to test the domain-general decline in the future.

Figures

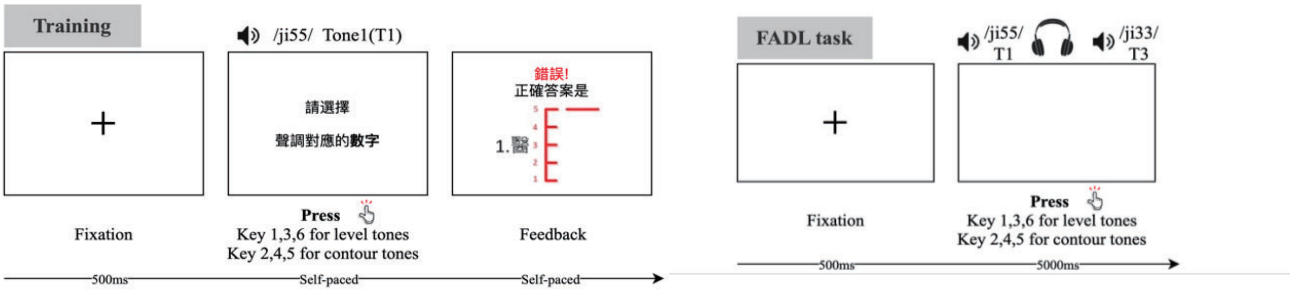


Fig. 1: The procedure of Cantonese tone traning and FADL task

Tab. 1: Examples of stimuli. In the training, stimuli were isolatedly presented. In the FADL, stimuli with the same syllable (i.e., /ji/, /se/, /fu/) and tone type were paired to present dichoticly, avoiding the experiment being too long and challenging (especially for the old group). There were 36 tone pairs in total.

Level tone pairs (left-right)	T1 /55/	T3 /33/	T6 /22/	*Contour tone pairs*	T2 /25/	T4 /21/	T5 /23/
T1 (/ji55/ 'doctor')	T1-T1	T1-T3	T1-T6	T2 (/ji25/ 'chair')	T2-T2	T2-T4	T2-T5
T3 (/ji33/ 'meaning')	T3-T1	T3-T3	T3-T6	T4 (/ji21/ 'son')	T4-T2	T4-T4	T4-T5
T6 (/ji22/ 'two')	T6-T1	T6-T3	T6-T6	T5 (/ji23/ 'ear')	T5-T2	T5-T4	T5-T5

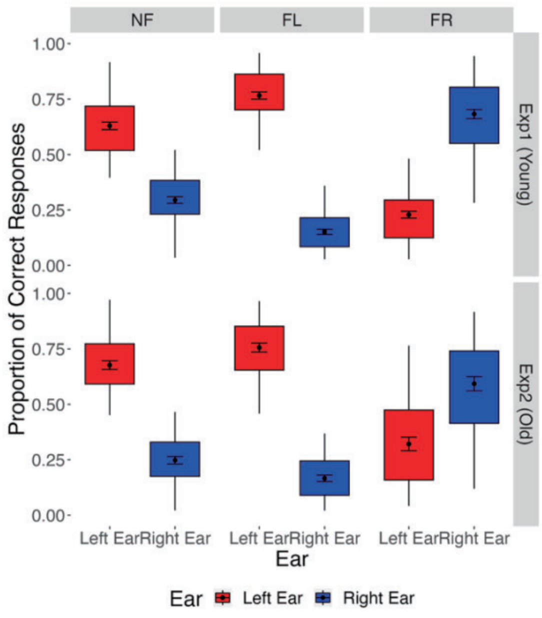


Fig. 2: Results of Exp 1 (Young adults, top) and of Exp 2 (old adults, bottom): Left ear (red) and right ear (blue) proportion of correct responses across NF (left), FL (middle), FR (right) conditions. The response for each trial was coded as a left ear correct response, a right ear correct response, or an error; left or right ear proportion of correct responses equals to the number of the correct responses in the left or right ear divided by the sum of the number of correct responses in both ear and the errors. For the statistical analysis, the mixed-effects models for each age group included the interaction of ear and condition as the fixed factors and the random factor of participant with the slope of ear and condition. No trial-level data was included in the random structure as the accuracy was not binary coded and it was calculated for each ear before putting into the model.

References

[1] Hugdahl, K., Westerhausen, R., Alho, K., Medvedev, S., Laine, M., & Hämäläinen, H. (2009). Attention and cognitive control: unfolding the dichotic listening story. *Scandinavian journal of psychology, 50*(1), 11–22.

[2] Hugdahl, K., Carlsson, G., & Eichele, T. (2001). Age effects in dichotic listening to consonant-vowel syllables: interactions with attention. *Developmental neuropsychology, 20*(1), 445–457.

[3] Jia, S., Tsang, Y. K., Huang, J., & Chen, H. C. (2013). Right hemisphere advantage in processing Cantonese level and contour tones: evidence from dichotic listening. *Neuroscience letters, 556*, 135-139.

[4] Snell, K. (1997). Age-related changes in temporal gap detection. *Journal of the Acoustical Society of America,101*, 2214–2220.

OS.1.03

Antecedent retrieval in strong and weak crossover configurations in L2 English

Jun Lyu (Peking University)
junlyu@pku.edu.cn

Introduction Recent L2 processing studies increasingly focus on how L2 learners make skilled use of retrieval cues (e.g., semantic and syntactic heuristics) to establish long-distance dependencies (e.g., Cunnings'17; Clahsen & Felser'18). To contribute to this line of research, the present study focuses on real-time retrieval of antecedents in the crossover configurations. In a crossover configuration, a *wh*-phrase (*which man*) moves across a co-indexed pronoun, like *he/his* in (1a,2a), which violates a slew of structural constraints (e.g., Reinhart'83; Lasnik & Stowell'91). We refer to these constraints collectively as the crossover constraint. Bound pronoun readings in STRONG crossover (SCO) are less acceptable than in WEAK crossover (WCO) as SCO additionally violates Condition C (Chomsky'81). By probing the L2 processing patterns, we examine (i) whether L2 learners can make efficient use of syntactic constraints to rule out ungrammatical pronoun interpretations and (ii) how the increased weighting of structural cues in SCO (crossover + Principle C) relative to WCO (crossover) impacts pronoun resolution at different processing stages in L2 English.

Methods Advanced Chinese-speaking L2 English speakers (N = 47 in Exp.1 on SCO; N = 47 in Exp.2 on WCO) read sentences (24 targets, 48 fillers) incrementally (self-paced reading) and made acceptability judgment on a 0-6 Likert scale after each trial. The gender congruency of the TARGET (match/mismatch) and the DISTRACTOR (match/mismatch) were crossed in a 2x2 factorial design. See ex.(1-2). The target refers to the structurally licit co-referent (e.g., *Peter*), while the distractor refers to the structurally illicit co-referent (e.g., *which man*). Following prior work (e.g., Sturt'03; Kush et al.'17), we take the **gender mismatch effect**, characterized by reading slowdowns when a pronoun and its antecedent have different genders, as a diagnostic of whether readers attempt to establish coreferences.

Results The acceptability scores are displayed in **Fig.1** (SCO) and **Fig.2** (WCO), and reading times (RTs) are shown in **Fig.3** (SCO) and **Fig.4** (WCO). We found that L2 learners can use syntactic constraints to rule out illicit pronoun readings in the SCO configuration – no gender mismatch effect from the distractor – but may treat the illicit *wh*-phrase as the binder in the WCO configuration – a *reversed* gender mismatch effect from the distractor at the critical pronoun (i.e., *his/her*) region, suggesting interference nonetheless. These L2 patterns contrast with our (unreported) L1 control study which shows that L1 English speakers do not consider illicit distractors at the pronoun region in either SCO or WCO configurations.

Conclusion Together with our L1 control study, this L2 study shows that while L2 speakers do not engage in shallow parsing – otherwise, they should show gender mismatch effects in SCO as well – they nonetheless assign less weight to the WCO constraint, which aligns with a cue-weight-based approach to L2 processing (e.g., Cunnings, 2017).

Example target stimuli in Exp.1 – SCO (slashes mark different regions in self-paced reading):

(1a) TARGET MATCH/DISTRACTOR MATCH
Peter/ told/ us/ which man/ in/ the office/ he/ had/ promoted__/ recently.
(1b) TARGET MATCH/DISTRACTOR MISMATCH
Sarah/ told/ us/ which man/ in/ the office/ she/ had/ promoted__/ recently.
(1c) TARGET MISMATCH/DISTRACTOR MATCH
Sarah/ told/ us/ which man/ in/ the office/ he/ had/ promoted__/ recently.
(1d) TARGET MISMATCH/DISTRACTOR MISMATCH
Peter/ told/ us/ which man/ in/ the office/ she/ had/ promoted__/ recently.

Example target stimuli in Exp.2 – WCO
(2a) TARGET MATCH/DISTRACTOR MATCH
Peter/ wondered/ which man/ in/ the office/ his/ supervisor/ had/ promoted__/ recently.
(2b) TARGET MATCH/DISTRACTOR MISMATCH
Sarah/ wondered/ which man/ in/ the office/ her/ supervisor/ had/ promoted__/ recently.
(2c) TARGET MISMATCH/DISTRACTOR MATCH
Sarah/ wondered/ which man/ in/ the office/ his/ supervisor/ had/ promoted__/ recently.
(2d) TARGET MISMATCH/DISTRACTOR MISMATCH
Peter/ wondered/ which man/ in/ the office/ her/ supervisor/ had/ promoted__/ recently.

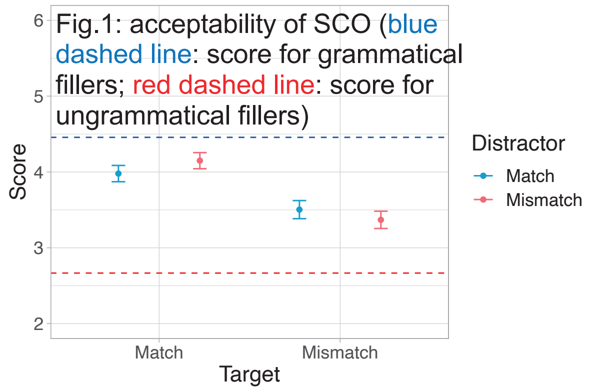


Fig.1: acceptability of SCO (blue dashed line: score for grammatical fillers; red dashed line: score for ungrammatical fillers)

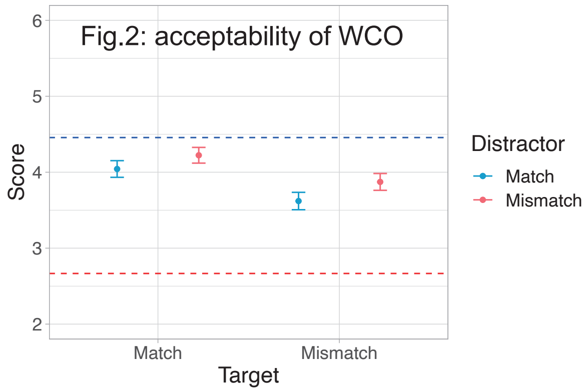


Fig.2: acceptability of WCO

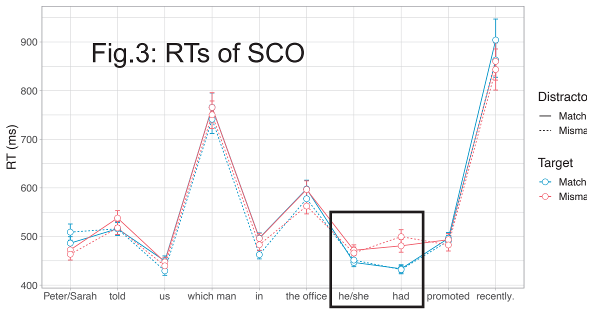


Fig.3: RTs of SCO

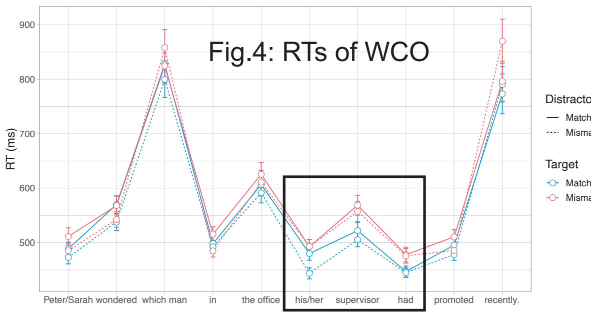


Fig.4: RTs of WCO

Selected references

Kush, Dave, Lidz, Jeffrey & Phillips, Colin. 2017. Looking forward and backwards: The real-time processing of strong and weak crossover. • Lasnik, Howard & Stowell, Tim. 1991. Weakest crossover. • Reinhart, Tanya. 1983. *Anaphora and Semantic Interpretation*. • Sturt, Patrick. 2003. The time-course of the application of binding constraints in reference resolution.

OS.1.04

Competitive or non-competitive selection? Lexical coactivation in action naming

Zenzi Griffin, griffinz@utexas.edu, University of Texas at Austin, USA; Gary Oppenheim, Bangor University, UK

The more likely a speaker is to produce an object's dominant (i.e., modal) name, the faster their response; this is the classic name agreement effect, an exceptionally robust finding that is usually attributed to the idea that strong alternatives actively slow the selection of a dominant name via competition (e.g., coactivation of 'sofa' hinders the selection of 'couch'). The theoretical importance of this interpretation cannot be overstated: the assumption of competitive selection is the foundation for the dominant theoretical model of language production (viz. Levelt, Roelofs, & Meyers, 1999). However, recent work on object naming has shown that, after controlling for dominant name agreement, pictures with stronger 'competitor' names are named faster, not slower, than those with weaker competitors. This secondary name agreement effect holds for speakers who demonstrably use both names, and is predicted under a simple race model as a result of statistical facilitation (Oppenheim, submitted; cf. Raab, 1962).

In this project, we used action naming (implemented as a web-based timed typing task with 480 images) to replicate and extend the secondary name agreement effect. Action words (~verbs) are acquired later than object words (~nouns) and are thought to differ in their cognitive and neural processing, offering a test case for assessing the effect's generality. We also used a novel, pre-registered, multi-site, multi-dialect design, testing 50 UK and 51 US participants in parallel to allow treating lexical coactivation as a within-items predictor (e.g., both sets of participants tended to describe a basketball scene as 'bouncing', but the US participants were much more likely to also use the term 'dribbling'), and further characterized semantic and formal relations between names to constrain alternative explanations for the effect. Multilevel analyses of keystroke RTs confirmed the presence of both the classic name agreement effect and the secondary name agreement effect (see Figure 1). Both were robust to the addition of standard between-items covariates (e.g., lexical frequency, age of acquisition), and neither effect was limited to particular classes of dominant-secondary relations (e.g., synonyms, near-synonyms, category coordinates, or responses that shared initial phonemes or keystrokes).

Thus, strong alternatives do not detectably hinder lexical selection in either object or action naming, challenging most accounts of interference effects in word production. The results instead support "good enough" horse-race-like models (see Goldberg & Ferreira, 2022; Oppenheim, Dell, & Schwartz, 2010) in which the first candidate to reach a threshold is selected, while others provide backup names with no online cost.

References

Goldberg, A. E., & Ferreira, F. (2022). Good-enough language production. Trends in Cognitive Sciences, 26(4), p300-311. https://doi.org/10.1016/j.tics.2022.01.005

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. Behavioral and Brain Sciences, 22(1), 1–38. https://doi.org/10.1017/s0140525x99001776

Oppenheim, G. M. (submitted). Lexical selection in language production. Draft available at https://doi.org/10.31234/osf.io/k2s3p

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. Cognition, 114(2), 227–252. https://doi.org/10.1016/j.cognition.2009.09.007

Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, *24*(5 Series II), 574–590. https://doi.org/10.1111/j.2164-0947.1962.tb01433.x

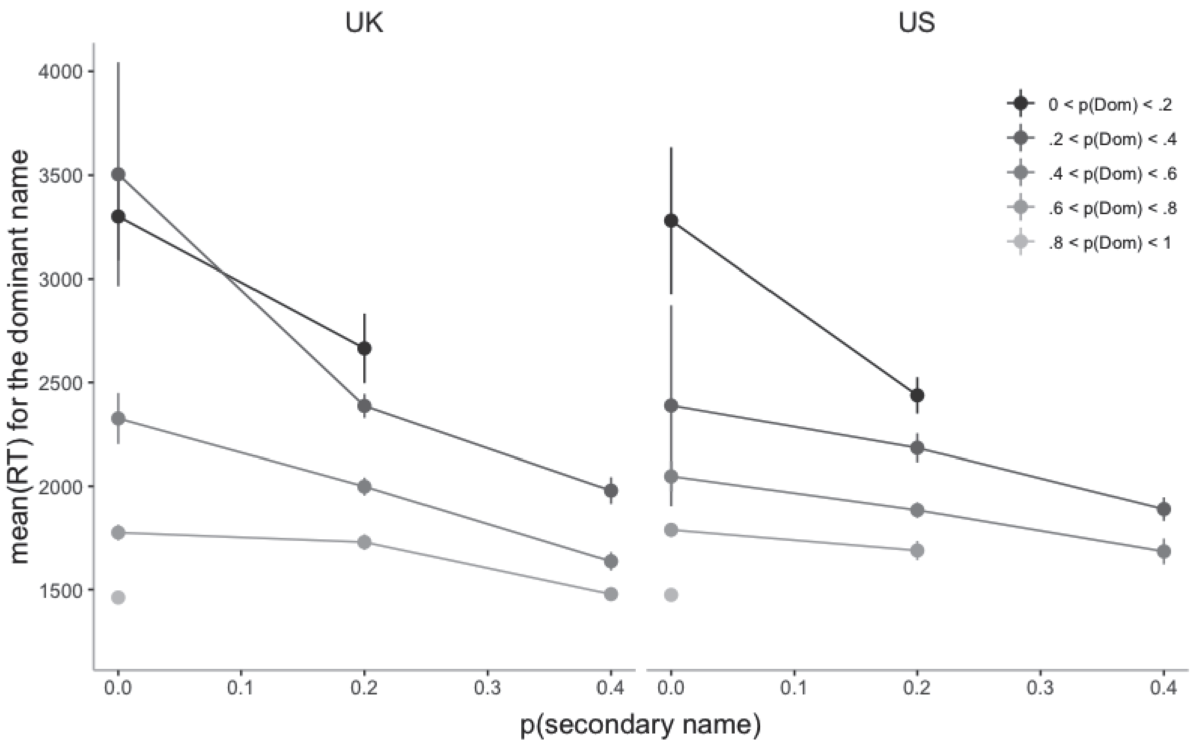


Figure 1. After controlling for dominant name agreement, stronger secondary name agreement facilitates rather than hinders dominant name production. That is, RTs for the dominant name of a certain probability decrease as the probability of its secondary name increases. This figure shows that the relation holds for each dialect independently as a between-items effect, by calculating mean response times for discrete bins of dominant and secondary name agreement. By integrating the trial-level data from the two dialects, our main statistical analyses show that the same relation also holds as a within-items effect.

OS.2.01

Word or Character? Investigating Minimal Unit in Chinese Distributional Semantics

Zehua Jiang, Cynthia S. Q. Siew

Humans learn from environmental statistical regularities, underpinning language acquisition from infancy. Distributional semantics constructs semantic representations by quantifying words and their contexts, assuming similar distributions share meanings. For instance, 'kitten' and 'cat' co-occur with words like 'pet' and 'animal,' unlike 'linguistic.' These patterns enable quantitative word representations. Research on various models (Kumar, 2021) has enriched our understanding of human language performance. A key question in these models is the minimal unit of distributional semantics—words or smaller units like characters. This is crucial for Chinese, where continuous characters require effective visual and auditory word segmentation.

To investigate whether Chinese distributional semantics operate at the word or character level, we focused on the multi-character words, and using Continuous Bag of Words (CBOW) across different window sizes on the SUBTLEX-CH corpus (Cai & Brysbaert, 2010) to construct word embeddings on word and character levels. We examined the impact of these models on word recognition using the lexical decision task (Tsang et al., 2018) and word naming task(Zhang et al., 2024). By constructing distributional networks from the embeddings of these models, we quantified their effects based on network structure properties at both micro and macro levels, focusing on the R-squared increase and Akaike Information Criterion (AIC) drops in linear models (reaction time) and generalised linear model(error rate).

Our results indicate that distributional semantic models outperform the baseline model in terms of R-squared values in both tasks (See Figure 1). For the lexical decision task, character-level embeddings outperformed word-level embeddings in reaction time($-\triangle AIC_{character\text{-}zRT}$: 93~173, $-\triangle AIC_{word\text{-}zRT}$: 18~48) across window sizes. Similar patterns were observed in the word naming task, where character-level embeddings exhibited a better fit to the data compared to word-level models ($-\triangle AIC_{character\text{-}zRT}$: -4~37,$-\triangle AIC_{word\text{-}zRT}$: -7~-5).

Our findings demonstrate that character might be the minimal unit of the Chinese distributional semantic, as the character-level embeddings outperform the word-level embeddings in both word recognition tasks. These findings corroborate existing literature on semantic composition in Chinese language processing, reinforcing the critical role of character-level representations in cognitive linguistic mechanisms and providing empirical support for the current understanding of semantic representation in Chinese comprehension.

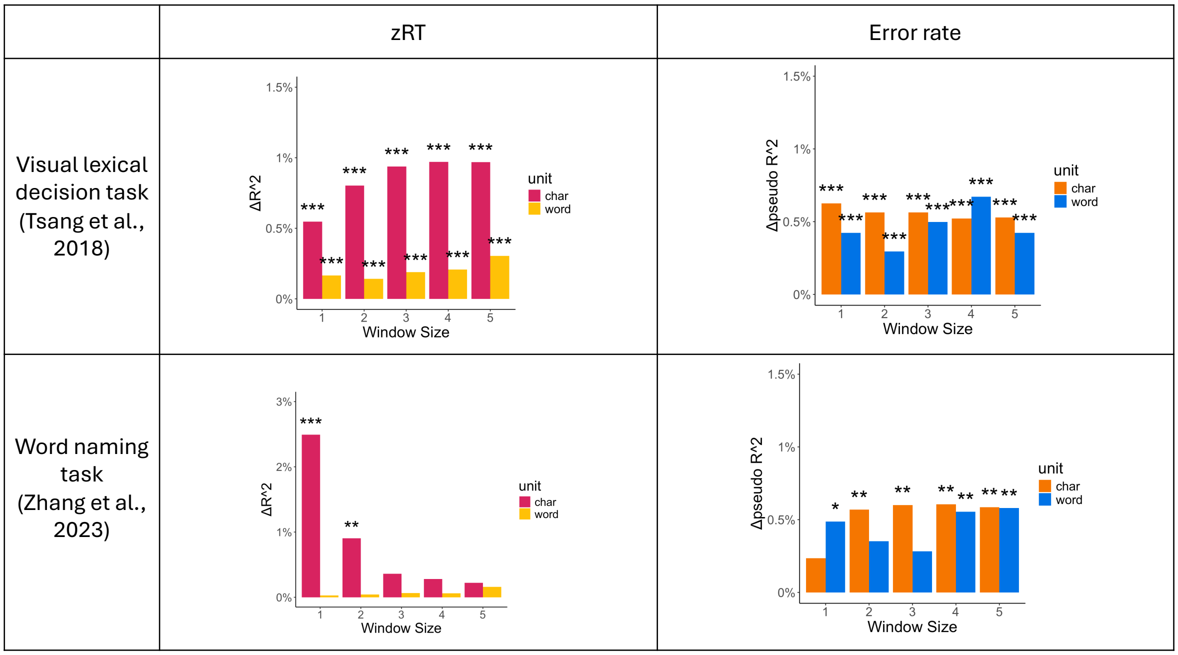


Figure 1 The comparative R^2 analysis of Continuous-Bag-of-Word (CBOW) Models

References

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese Word and Character Frequencies Based on Film Subtitles. *PLOS ONE*, *5*(6), e10729. https://doi.org/10.1371/journal.pone.0010729

Kumar, A. A. (2021). Semantic memory: A review of methods, models, and current challenges. *Psychonomic Bulletin & Review*, *28*(1), 40–80. https://doi.org/10.3758/s13423-020-01792-x

Tsang, Y.-K., Huang, J., Lui, M., Xue, M., Chan, Y.-W. F., Wang, S., & Chen, H.-C. (2018). MELD-SCH: A megastudy of lexical decision in simplified Chinese. *Behavior Research Methods*, *50*(5), 1763–1777. https://doi.org/10.3758/s13428-017-0944-0

Zhang, M., Liu, Z., Botezatu, M. R., Dang, Q., Yuan, Q., Han, J., Liu, L., & Guo, T. (2024). A large-scale database of Chinese characters and words collected from elementary school textbooks. *Behavior Research Methods*, *56*(5), 4732–4757. https://doi.org/10.3758/s13428-023-02214-1

OS.2.02

Semantic structure and verbal analogical reasoning are enhanced in curious individuals

Adel Chaouch-Orozco, City University of Hong Kong; achaouch@um.cityu.edu.hk

Curiosity, our drive to seek new information, plays a pivotal role in fostering verbal creative thinking. Recombinant knowledge search (Schilling and Green, 2011), which involves connecting unrelated concepts and analogical reasoning (i.e., adapting solutions to problems from unrelated situations), may be more prevalent in curious individuals who possess a wide range of interests. This process could result in more interconnected semantic systems in curious individuals, which would support their verbal analogical reasoning abilities.

To examine this question, we are testing monolingual English speakers (target $N = 200$) using an analogy decision task, where participants judge whether two pairs of words (e.g., wire : copper :: knife : steel) formed valid analogies. Subjects also complete the 5-Dimensional Curiosity Scale (Kashdan et al., 2018) to obtain their curiosity profiles and divide them into high- and low-curiosity groups with a Gaussian Mixture Model. Additionally, participants perform two semantic fluency tasks (animals, fruits and vegetables) to derive semantic networks (Christensen & Kenett, 2021). Following Kenett et al. (2014), we hypothesize that more curious individuals will exhibit more cohesive and interconnected semantic networks, characterized by larger clustering coefficient (CC; i.e., interconnectedness), lower average shortest-path length (ASPL; i.e., navigability), and lower modularity (Q; i.e., community structure) values. Enhanced semantic organization is expected to allow high-curiosity individuals to outperform their counterparts in analogical reasoning by facilitating the retrieval and application of diverse knowledge structures.

Preliminary analyses with 80 subjects show that, as predicted, more curious individuals display significantly better-organized semantic networks (Fig. 1, 2). Importantly, a generalized linear mixed-effects model indicates that they are also significantly more accurate in judging the validity of analogies ($N = 120$; $\beta = 1.11$, $z = 2.15$, $p < .05$; Fig. 3). Further, the effect of curiosity on analogical reasoning is not modulated by attention nor by vocabulary size, indicating that constant exposure to diverse concepts and their integration into novel solutions leads to more efficient semantic structure. This enhanced organization, in turn, facilitates the identification and integration of functional relationships between concepts, improving analogical reasoning.

These preliminary findings underscore the importance of curiosity in cognitive processes, particularly in enhancing lexical semantic processing and integration. By demonstrating the association between curiosity, semantic organization, and verbal analogical reasoning, our study highlights the value of fostering curiosity to improve verbal problem-solving skills. This understanding can inform educational strategies, promoting environments that stimulate curiosity.

References:

- Khare, V., Verma, A., Kar, B., Srinivasan, N., & Brysbaert, M. (2013). Bilingualism and the increased attentional blink effect: Evidence that the difference between bilinguals and monolinguals generalizes to different levels of second language proficiency. *Psychological Research*, *77*(6), 728–737. https://doi.org/10.1007/s00426-012-0466-4
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0
- Revelle, W., & Condon, D. M. (2018). Reliability. In P. Irwing, T. Booth, & D. J. Hughes (Eds.), *The Wiley Handbook of Psychometric Testing* (1st ed., pp. 709–749). Wiley. https://doi.org/10.1002/9781118489772.ch23
- Rizopoulos, D. (2006). **ltm**: An *R* Package for Latent Variable Modeling and Item Response Theory Analyses. *Journal of Statistical Software*, *17*(5). https://doi.org/10.18637/jss.v017.i05

Figures:

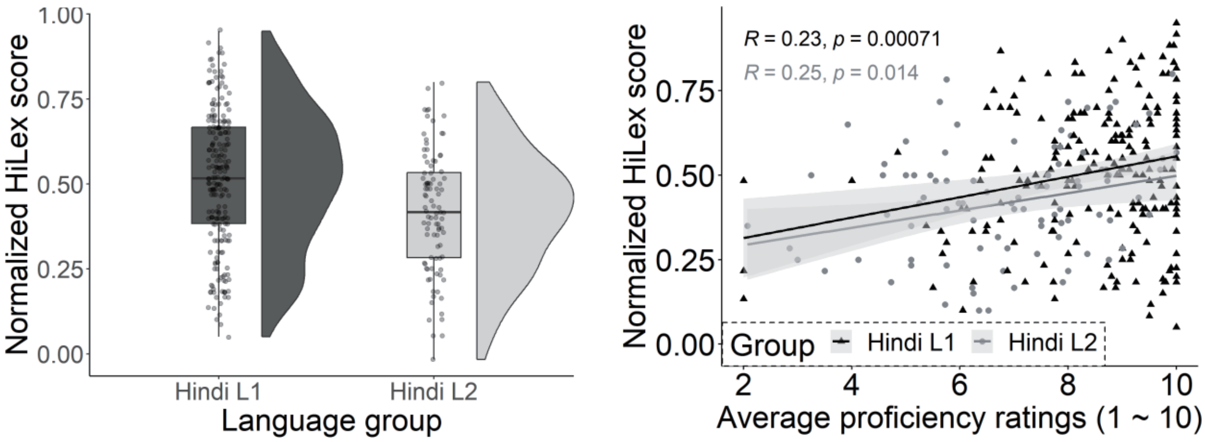


Fig. 1: (a) Distribution of HiLex scores for Hindi L1 and L2 participants, (b) Pearson's correlation coefficient of self-ratings (avg) with HiLex scores for L1 and L2 groups

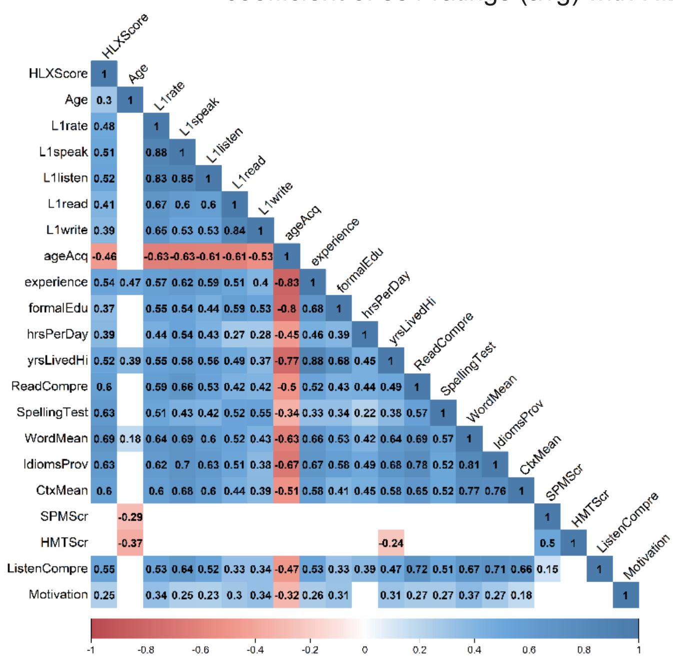


Fig. 2: Overall correlation matrix of validation study for both L1 and L2 participants combined. Missing values represent non-significant correlation coefficient.

Appendix 1 (Scoring):

a) Original LexTALE (English) score, Averaged accuracy (range 0 to 1)

$$= \frac{(N_{correct-words}/60) + (N_{correct-nonwords}/30)}{2} = \frac{(N_{correct-words} + 2N_{correct-nonwords})}{120}$$

b) Ghent score, HiLex score (range -60 to +60)

$$= \left(N_{yes-to-words} - 2N_{yes-to-nonwords}\right) = hits - 2\ (false\ alarms)$$

c) Normalized Ghent score, Normalized HiLex score (range -1 to +1)

$$= \frac{\left(N_{yes-to-words} - 2N_{yes-to-nonwords}\right)}{60} = \frac{hits - 2\ (false\ alarms)}{N_{words}}$$

Appendix 2 (Hindi specific tasks used in validation study):

Task	**Material**
Demographics + LBQ	Demographics and modified LEAP-Q *(Marian et al., 2007)*
HiLex	Lexical Decision Task on 60 words; 30 pseudo-words
Reading comprehension	Fiji country informative easy 215 Press & Media historical easy 210 Antibodies scientific hard 242 Akbar's harem social historical hard 207
Spelling	Yes/ No responses on 30 words (15 with correct spelling)
Word meaning	20 words, each with 4 response options
Idioms and proverbs	15 idioms, each with 4 response options
Contextual meaning	20 statements, each with 4 response options
Fluid intelligence test	Raven's SPM-9 (shortened) – *(Bilker et al., 2012)* Hagens Matrices Test (HMT-6) *(Timo Heydasch et al., 2013)*
Verbal fluency tests	Phonemic fluency - प (pa), and स (sa) Semantic fluency – animal, and food/drinks
Reading efficiency	Sighted Word Reading: 90 words Phonemic Decoding: 60 pseudowords
Listening comprehension	Bonded labour informative easy 1:20 Advertisement satire easy 1:49 Birds factual hard 1:50 Life lecture hard 1:08

OS.2.03

New social media-based corpora for Hindi with better frequency estimates

Pawan Kumar[1], Anurag Khare[2], Niket Agrawal[3], Vivek Singh Sikarwar[4] and Ark Verma[5]
[1,2,3,4,5]IIT Kanpur; [1]LNMIIT Jaipur; [4]Manipal University Jaipur
[2]akhare@iitk.ac.in

Introduction: Words with specific properties are needed for controlled exp which are often selected from different corpora (see references). The language register from which the corpora are derived also impacts the predictive power of the measures obtained from the corpora such as frequency etc. E.g.: Wordlex_blog's freq. est. outperformed that of newspaper-based Shabd's; and Shabd outperformed other corpora (Verma et al., 2022). Here we are presenting new social-media based corpora: i) Twitter_Hi (1M unique words), ii) Youtube_Hi sampled from YT comments (390k unique words), and iii) Merged corpus having 1.2M unique word types.

Exp 1 validated freq. estimates of our corpora. Words were chosen with freq. (high/ low) and length (long/ short), 125 in each condition. 36 participants took part in LDT. The small/large conditions of length and frequency were coded as −1 and +1, respectively. LME model showed significant effects of freq. (estimate=0.1286, t=21.86, p<0.001), word length (estimate=0.0644, t=9.16, p<0.001) and their interaction (estimate=0.0129, t=2.66, p<0.008) with log (RT) dependent variable. Similar analysis was performed on accuracy which revealed main effects of frequency and word-length (both p<0.001) but no interaction.

Exp 2a, b, c, d were performed to compare divergent frequency estimates of current corpora (Youtube and Twitter) with previous corpora (Shabd1 and Wordlex_Blog). Data was collected from 24 x 4 participants. Exp-2a compared Twitter_Hi with Shabd, 2b compared Twitter_Hi with Wordlex_Blog, 2c compared Youtube_Hi with Shabd1 and 2d compared Youtube_Hi with Wordlex_Blog. Stimuli selection, procedure and design were similar for all experiments, where linear regression was performed on corpus A with freq. of corpus B as predictors to obtain three groups of words (divergent estimates), 150 each (i) higher freq. in corpA than corpB (ii) higher freq. in corpB than corpA (iii) no significant difference in freq. between two corpora. Vuong test for non-nested models (Vuong, 1989) show that YouTube's freq. estimates outperformed that of Wordlex_blog's (RT: z=5.297, p<0.001, accuracy: z=6.809, p<0.001). Wordlex_blog's freq. estimates had outperformed that of Twitter's (RT: z= -5.593, p<0.001, accuracy: z= -4.187, p<0.001). These three corpora outperformed Shabd corpus.

Discussion: As expected, our social-media derived corpora outperform the existing newspaper-based corpora for Hindi, i.e., Shabd (Verma et al., 2021), owing probably to a more naturalized language register. Interestingly, Wordlex_blog still outperforms Twitter but not YouTube, which would be due to the more formal nature of language available on twitter, but almost uncensored comments on YouTube, again showing the importance of an ecologically valid language register for predicting LD times (Brysbaert & New, 2009).

Sample stimuli: words**:** दीन, कर्मचारियों, मोहतरमा; non-words**:** उमाना, आत्मनिर्धर, अर्गदर्शन

References:

- Baker, P., Hardie, A., McEnery, T., Cunningham, H., & Gaizauskas, R. J. (2002, May). EMILLE, A 67-Million Word Corpus of Indic Languages: Data Collection, Mark-up and Harmonisation. In Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02).
- Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. Behavior Research Methods, 41(4), 977-990.
- Gimenes, M., & New, B. (2015). Worldlex: Twitter and blog word frequencies for 66 languages. Behavior Research Methods, 48(3), 963–972.
- Kunchukuttan, A., Mehta, P., & Bhattacharyya, P. (2017). The IIT - Bombay English-Hindi Parallel Corpus. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018).
- Saxena, A., Madhyasta, P. S., & Nivre, J. (2008). Building the Uppsala Hindi Corpus. SLTC 2008, 11.
- Verma, A., Sikarwar, V., Yadav, H., Jaganathan, R., & Kumar, P. (2022). Shabd: A psycholinguistic database for Hindi. Behavior Research Methods, 54(2), 830–844.
- Vuong, Q. H. (1989). Likelihood ratio tests for model selection and nonnested hypotheses. Econometrica: Journal of the Econometric Society, 307-333.

Table 1: Mean RT and accuracy for words of different conditions in Experiment 1

Condition	RT (ms); Mean (SD)	Accuracy: Mean (SD)
High_long	926.21 (284.82)	0.95 (0.22)
High_short	829.10 (227.86)	0.92 (0.28)
Low_long	1185.70 (336.62)	0.62 (0.49)
Low_short	1029.11 (306.65)	0.50 (0.50)

Tables 2, 3, 4 and 5: Summary of statistics of acc. and RT for Experiment 2A, B, C, D

	Table 2: Experiment 2A		Table 3: Experiment 2B	
Condition	Acc: mean (SD)	RT (ms): Mean (SD)	Acc: mean (SD)	RT (ms): Mean (SD)
Words with low freq. in Twitter	0.70 (0.46)	1007.57 (345.16)	0.87 (0.33)	838.45 (269.96)
Randomly sampled words	0.71 (0.46)	996.82 (333.75)	0.77 (0.42)	938.40 (307.79)
Words with high freq. in Twitter	0.93 (0.25)	840.01 (264.58)	0.80 (0.40)	957.63 (313.38)
	Table 4: Experiment 2C		Table 5: Experiment 2C	
Condition	Acc: mean (SD)	RT (ms): Mean (SD)	Acc: mean (SD)	RT (ms): Mean (SD)
Words with low freq. in YouTube	0.62 (0.49)	1094.70 (366.50)	0.59 (0.49)	1063.48 (338.08)
Randomly sampled words	0.73 (0.44)	951.91 (324.57)	0.76 (0.43)	956.49 (310.65)
Words with high freq. in YouTube	0.96 (0.20)	758.80 (224.82)	0.87 (0.33)	907.19 (295.76)

Additional information about Hindi Language:

Hindi differs from other languages in terms of how syllables and vowels are defined in the language. Consonants in the language have hidden vowels at their end, also called as schwa, for example, क /ka/ is just akshara in Hindi but can be considered as a whole syllable in some cases. However, in most cases, diacritic markers i.e. vowels are added to such aksharas or consonants to add to its phonology. Such markers are called matras and can be added to the aksharas in both linear and non-linear fashion, example (क+ा = का /kaa/ and क+ि = कि /ki/). Such markers (matras) can also be written to top or bottom of the consonant (akshara), each changing the pronunciation in different ways. Thus, we defined sub syllabic unit in Hindi as a combination of akshara (consonant) + matra (diacritic marker or vowel). (if matra is present), or just akshara (consonant) (if matra is absent).

OS.2.04

The semantics of silence: Time-to-respond as a cue to conversational meaning

Yingjia Wan; Craig Chambers (University of Toronto; Email: yingjia.wan@utoronto.ca)

Conversational turn-taking normally reflects predictable patterns of timing. However, silences between contributions of dialogue partners ('inter-turn-intervals': ITIs) can become pragmatically meaningful, e.g., ITI duration can color listeners' perception of speakers' credibility [1] and confidence [2]. This study examines how ITI influences the interpretation of a response to a "socially-loaded" question ("Does this dress make me look fat?" […] "No."). Intuitively, long delays can lead a "no" response to be interpreted more like a "yes", and vice-versa. Our goals were (i) to explore this intuition using systematically-manipulated differences in ITI duration, and (ii) to explore if sensitivity depends on a perceiver's own processing speed. For example, "slow" listeners might find a delayed response to be more sincere than other listeners. We tested three listener groups thought to vary in processing speed/efficiency: (i) young adult native speakers (control), (ii) young adult non-native speakers, who show delays in L2 listening (e.g., [3]), and (iii) older adults, who exhibit declines in processing speed (e.g., [4]). Participants listened to recorded dialogues. Critical trials contained a socially-loaded yes/no question and a second person's response (see example). Voices were varied across dialogues and were computer-generated to maintain uniform intonation/speech rate. ITI duration was varied across four 500 ms steps (600-2100 ms). Across participants, each dialogue was tested with all ITIs, yet a given participant heard a specific dialogue in only one ITI. After each dialogue, participants rated the truthfulness of a statement (providing a measure of the sincerity of the response), using a slider (0 = very unlikely; 100 = very likely). Results (analyzed with LME) showed replies were rated as **more insincere as ITI duration increased** in all three groups (young adults: $t = 5.53$, $p < .001$; older adults: $t = 3.48$, $p < .001$; non-native speakers: $t = 2.58$, $p = .010$). Intriguingly, non-native young adults and older (native) adults rated replies as **more sincere** ($p < .001$) than young native speakers, consistent with the idea that slower overall processing might entail more literal/semantic interpretation of the response (although the gradient sensitivity to the increasing ITI was still evident). Together, the methodology and results provide promising directions for future studies of conversational pacing.

Example of Critical Trial

Audio: (Speaker X): *Does this dress make me look fat?*
(Speaker Y): [**after 600/1100/1600/2100 ms pause**] *No.*

On Screen: Based on what you just heard, how likely do you think the following statement is true:

Y thinks the dress makes X look fat.

(Participants knew that *X* always meant the first speaker and *Y* the second. Participants' ratings for trials where Speaker *Y* answered "yes" were inverse-coded.)

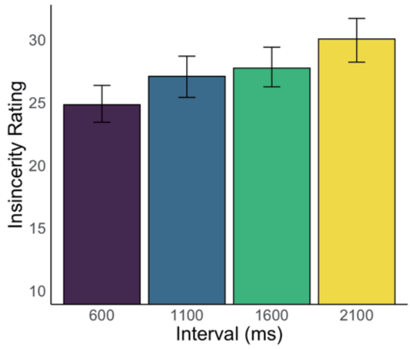


Figure 1. Perceived insincerity of response by ITI, collapsed across groups

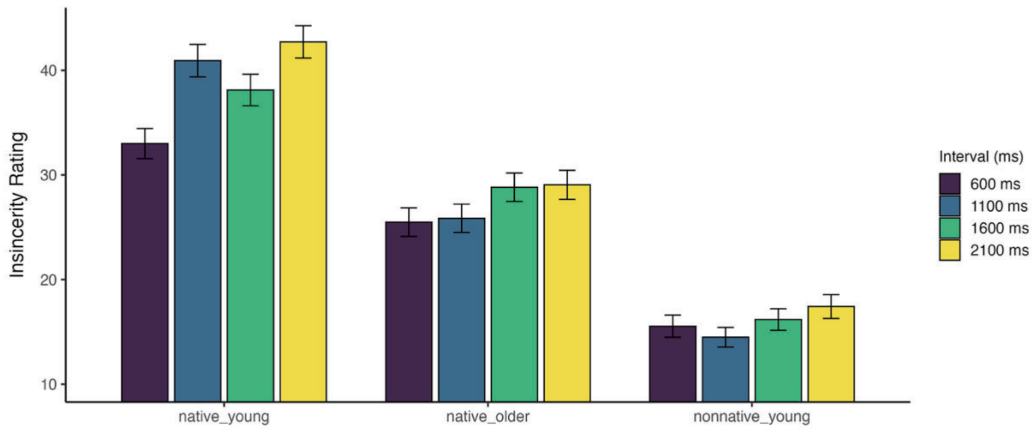


Figure 2. Perceived insincerity of response by ITI, separated by participant group

Native-speaking younger adults (n = 84, mean age = 28.8, range: 18 - 38 years)
Native-speaking older adults (n = 83, mean age = 68.5, range: 65 - 85 years)
Non-native younger adults (n = 82, mean age = 28.0, range: 18 - 38 years)
(All participant groups recruited via Prolific)

References

1. Ozuru, Y. & Hirst, W. (2006). *Memory and Cognition, 34*, 1512-1526.
2. Smith V. L, & Clark H. H. (1993). *Journal of Memory and Language*, *32*, 25–38.
3. Weber, A., & Cutler, A. (2004). *Journal of Memory and Language*, *50*, 1-25.
4. Salthouse, T. A. (1996). *Psychological Review*, *103*, 403.
5. Stine-Morrow & Radvanksy (2017). Routledge Handbook of Discourse Processes.

Silent reading, accent phrases and delta brain waves

Philippe Martin, LLF, UFRL, Université Paris Cité, Paris, France. philippe.martin@utoronto.ca

Silent reading performance is usually measured in terms of words read per minute (Dunning 2010; Ciuffo et al. 2017; Brysbaert 2019). However, it is easy to see that we don't read a written text word by word, either silently or orally. Instead, it can be shown that we read accent phrase by accent phrase (Blanche-Benveniste 2003), i.e. groups of words containing one and only one obligatory stressed syllable (thus excluding optional emphatic stress).

Interestingly, lexically stressed languages, such as English or Italian, and rhythmically stressed languages, such as French and Korean, behave differently in terms of segmenting speech into accent phrases. Indeed, in English and Italian, a mandatory syllabic stress (the "word stress") is usually located on some lexically defined syllable of a content word (verb, adjective, adverb or noun), which, combined with some grammatical word(s) (conjunction, article, preposition...), form predictable accent phrases. In French and Korean, on the other hand, the accent is placed on the final syllable of any word, whatever its grammatical category. It follows that the composition of accent phrases depends only on the reading rate (silent or oral) adopted by the reader, with an average of 4-5 syllables per second, the limit being about 8 to 9 syllables, the last one being stressed. (Martin 2014).

Despite these differences in the composition of accent phrases in lexically and rhythmically stressed languages, experimental results show that the faster reading rate of silent speech is limited to about 4 accent phrases per second (250 ms per accent phrase). This rate is independent of the number of words in the accent phrases. However, the 70-80 ms duration of eye movement saccades during reading (Quercia 1010; Rayner et al. 2010; Reichle et al. 1998) should allow a much faster reading rate, reaching around 12 accent phrases per second. These results suggest that silent reading may be slowed down by brain delta waves oscillating in the range of 0.8 Hz to 4 Hz, i.e. in the range of 250 ms to 1250 ms (Friederici, 2010; Giraud and Poeppel, 2012).

On the other side of the time scale, French data show that successive stressed syllables cannot be separated by more than about 1250 ms in continuous speech, leading to the inevitable realization of a secondary stress in long words, even during silent reading (Martin 2014). This may suggest that delta oscillations, with a maximum period of 1250 ms, induce the perception of a secondary stress that is obviously acoustically absent in silent reading (Fig. 1).

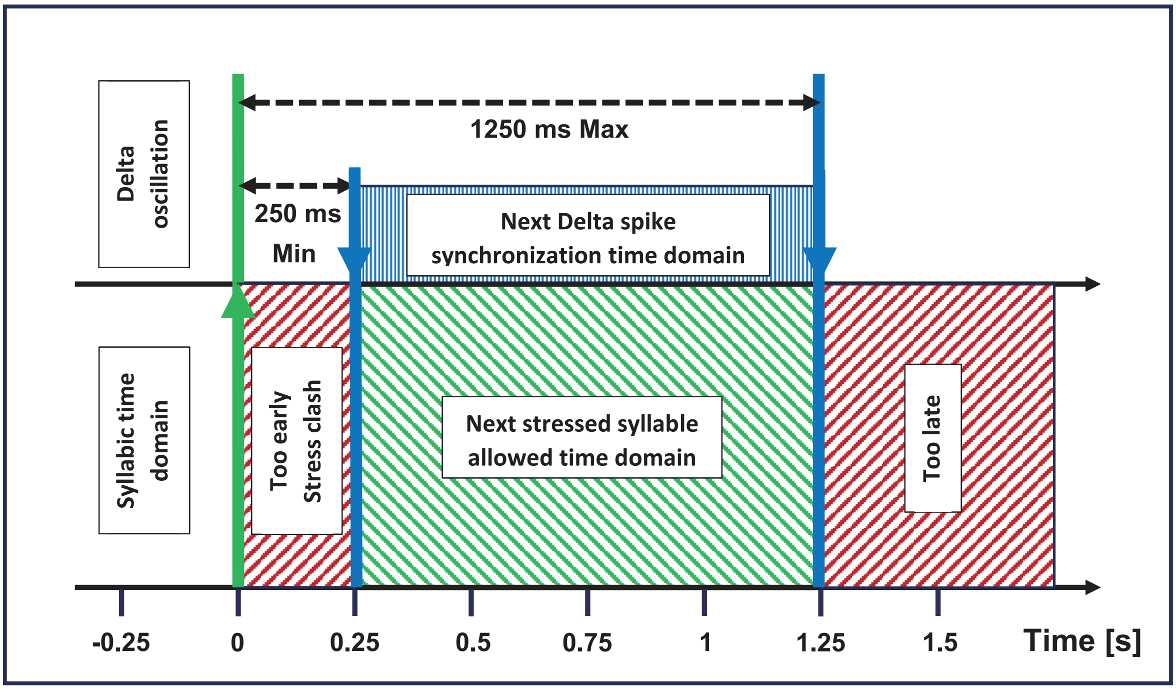


Figure 1: The last stressed syllable occurs at time 0. It synchronizes the delta free oscillation (green arrow up). Subject to phonological constraints (e.g. last syllable of the word in French), the next stressed syllable must occur within the period of free delta oscillation (blue arrows downwards). In this time range (in green, i.e. between about 250 ms and 1250 ms after time 0) a delta spike is synchronized by the occurrence of a stressed syllable. Before 250 ms, a stress clash condition occurs. After 1250 ms (e.g., in the case of long words), any syllable that falls within the green allowed time range and meets some phonological rule of the language (e.g., an internal morphological boundary) will be perceived as stressed, even if it is not acoustically stressed by a longer duration or a larger pitch change.

References

Blanche-Benveniste C. 2003. La naissance des syntagmes dans les hésitations et répétitions du parler, in Araoui J.L. ed. *Le sens et la mesure. Hommages à Benoît de Cornulier*, Honoré Champion, Paris, 40-55.

Brysbaert M. 2019. OSF | Reading Rate, 2019

Ciuffo, M., Myers, J., Ingrassia, M. *et al*. 2017. How fast can we read in the mind? Developmental trajectories of silent reading fluency. *Read Writ* **30**, 1667–1686 (2017). https://doi.org/10.1007/s11145-017-9744-2.

Dunning, B. 2010. Speed Reading, Skeptoid Podcast. Skeptoid Media, Inc., 26 Oct 2010. Web. 12 Dec 2013. http://skeptoid.com/episodes/4229

Friederici A. D., Wartenburger I. 2010. Language and brain, *Cognitive Science*, vol. 10, 150-159, 2010.

Gagl B, Gregorova K, Golch J, et al. 2022. Eye movements during text reading align with the rate of speech production. *Nat Hum Behav*. 2022; 6(3):429-442. doi:10.1038/s41562-021-01215-4

Giraud A-L., Poeppel D. 2012. Cortical oscillations and speech processing: emerging computational principles, *Nature neuroscience E-pub*, doi: 10.1038/nn.3063, 2012.

Martin Ph. 2014. Spontaneous speech corpus data validates prosodic constraints, *Proceedings of the 6th conference on speech prosody*, Campbell, Gibbon, and Hirst (eds.), 525-529.

Quercia P. 2010. Ocular movements and reading: a review, *J. Fr. Ophtalmologie*, 33 (6): 416-423.

Rayner K., Slattery T., Bélanger N. 2010. Eye movements, the perceptual span, and reading speed, *Psychon Bull Rev.*, vol. 17, n° 6, 834-839, déc. 2010.

Reichle E. D., Pollatsek A. et al. 1998. Toward a model of eye movement control in reading, *Psychological Review*, vol. 105, n° 1, 125-157, 1998.

OS.3.02

Predicting Phonological and Orthographic Representations during Chinese Comprehension: A Visual-World Eye-Tracking Study

Mingyuan YANG (mingyuan.y@u.nus.edu), Aine ITO National University of Singapore

People regularly predict the upcoming word during comprehension. While evidence for semantic prediction seems robust and consistent [1,2,3,4], evidence for phonological/orthographic word form prediction is scarce and mixed [5,6,7]. Moreover, although some studies utilized logographic languages such as Chinese and dissociated a phonological prediction effect from an orthographic effect [8], whether people pre-activate orthographic representations during listening comprehension is still underspecified in prediction models [9]. Hence, in the current study, we used Chinese and the visual-world eye-tracking paradigm to investigate prediction of orthographic and phonological representations.

Forty native Chinese speakers listened to sentences with a highly predictable target word (20 items). A scene with four objects appeared 1000ms before target word onset. The scene contained the target and three distractors in Target Condition, and an orthographic competitor, a phonological competitor, and two distractors in Orthographic-Phonological Condition (see Figure 1 for examples). We analyzed fixation proportion on the critical objects from -800ms to 200ms relative to the target word onset to capture eye movements before the target word could be processed. We computed fixation proportion log-ratio (a) between the target and the averaged distractors, (b) between the two competitors combined and the averaged distractors, and (c) between the orthographic and phonological competitors for every 20ms time bins. We used the growth curve analysis (GCA) with first-order and second-order orthogonal polynomials to compare looks to each object and how the difference changed over time.

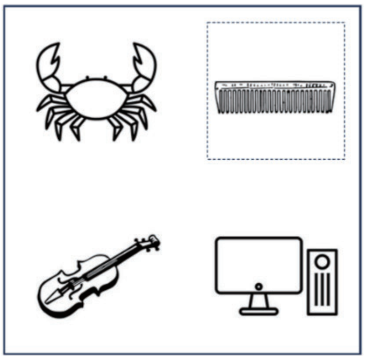
We found that the target attracted more fixations than the distractors in Target Condition. Moreover, in Orthographic-Phonological Condition, we found a significant effect on the quadratic term, indicating different fixation curvatures for distractors and combined competitors. Critically, we also found a significant effect on the linear term, indicating a steeper fixation increase for competitors vs. distractors. These results suggest that people pre-activated both orthographic and phonological representations of a highly predictable word during Chinese comprehension. Furthermore, the time course of fixation proportion (Figure 2) shows that the orthographic competitor effect peaked earlier than the phonological competitor effect, suggesting that orthographic representations were pre-activated earlier than phonological representations. These findings suggest that native Chinese speakers can predict orthographic forms earlier than phonological forms during listening comprehension, which could be because activation spreads faster from meaning to orthographic form than from meaning to phonological form.

Example of critical sentence and experiment design

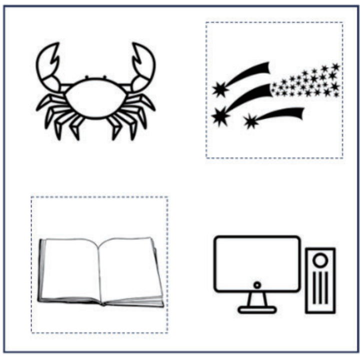
小丽坐在镜子前打理着头发，手里拿着一把梳子。

Gloss translation: 小丽 Xiao Li/ 坐 sits/ 在化妆镜前 in front of the mirror/ 打理着 tidying/ 头发 the hair/ 手里 in hand/ 拿着一把梳子 with a comb

"Xiao Li sits in front of the mirror, tidying the hair with a comb in hand."



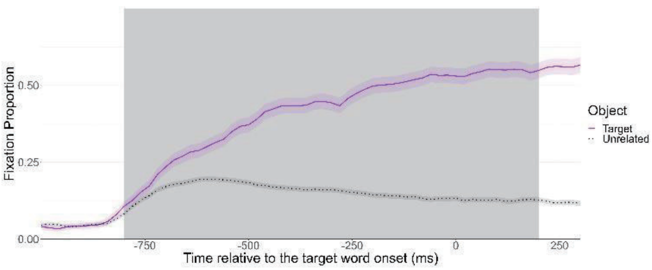
Target Condition



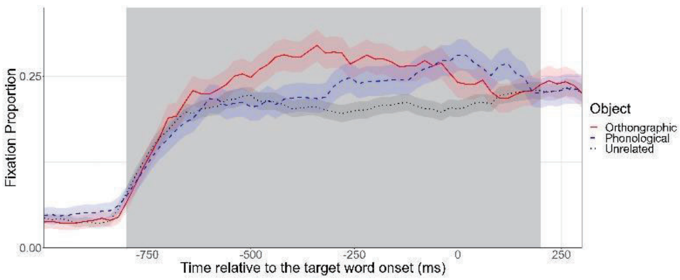
Orthographic-Phonological Condition

Figure 1 | An example of the visual stimuli for two conditions of the critical sentence above. The items in the dotted box indicate the critical objects: 梳子/shu1zi0/(comb) for the target in Target Condition; 流星/liu2xing1/(meteor) for the orthographic competitor and 书本/shu1ben3/(book) for the phonological competitor in Orthographic-Phonological Condition. The distractors had no semantic, orthographic and phonological connection with the target word: 螃蟹/pang2xie4/(crab), 电脑/dian4nao3/(computer), 提琴/ti2qin2/(violin).

Results



Target Condition



Orthographic-Phonological Condition

Figure 2 | Mean fixation proportion with ±1 standard error (shaded area around each line) for each object in Target Condition (left) and Orthographic-Phonological Condition (right). The grey shaded area indicates the analyzed time window. Time 0 indicates the onset of the target word.

Reference

[1] Altmann and Kamide, 1999; [2] Kamide et al., 2003; [3] Kukona et al., 2011; [4] Ito et al., 2018a; [5] DeLong et al., 2005; [6] Ito & Husband, 2017; [7] Nieuwland et al., 2018; [8] Li et al, 2022; [9] Pickering and Garrod, 2013

OS.3.03

Multimodal Prediction: Human and Language Model Performance Using the Visual World Paradigm

Shuqi Wang[1], Xufeng Duan[1], Zhenguang Cai[1,2]
[1] Department of Linguistics and Modern Languages, The Chinese University of Hong Kong
[2] Brain and Mind Institute, The Chinese University of Hong Kong
Email: ShuqiWang@link.cuhk.edu.hk

Background. Using the visual world paradigm (VWP), psycholinguistic studies have shown that humans make anticipatory eye movements to objects based on linguistic cues like verbs (Altmann & Kamide, 1999) and gender stereotypes (Kamide et al., 2003). This led use to question whether multimodal large language models (MLLMs) exhibit similar predictive behaviors, though current evidence is limited (Kewenig et al., 2024).

Method. Following the design of Corps et al. (2022), the experiment had a 2 (**verb**: match vs. mismatch) × 2 (**gender**: match vs. mismatch) design with 28 sentence pairs, each featuring male or female characters ("James/Kate would like to wear the nice tie/dress") and accompanied by a visual display of four objects. The objects varied in their match to the verb and gender stereotype of the sentence subject (Fig.1). The stimuli were presented to LLAVA 1.5 (Liu et al., 2023), a transformer-based MLLM, to compute its **attention proportion** to the four objects at different sentence segments: name ("James"), verb ("James will wear"), before target ("James will wear the nice"), and target ("James will wear the nice tie").

Results. A linear mixed-effect model analysis showed that LLAVA significantly preferred verb-matching objects (e.g., tie and dress) over verb-mismatching objects (e.g., drill and hairdryer) upon encountering the verb (e.g., "wear"), indicating its ability to use verb semantics to allocate attention to objects in the scene ($\beta = 0.01$, $SE = 0.00$, $t = 4.17$, $p < .001$), like humans. Interestingly, unlike humans, LLAVA did not show any preference for gender-matching objects at this stage or any subsequent stage of the sentence, as evidenced by the lack of a main effect of gender and the absence of an interaction between gender and verb (**Fig. 2**). A layer-wise analysis of the model revealed that middle layers of LLAVA were primarily responsible for the verb-based prediction effect (**Fig. 3**).

Conclusion. The results reveal that LLAVA, one of the MLLMs, can predictively attend to verb-relevant objects similar to humans but fails to demonstrate gender-based anticipatory attention. This study is pioneering in using psycholinguistic paradigms to compare multimodal predictive attention, revealing both similarities and differences between human and model behaviors.

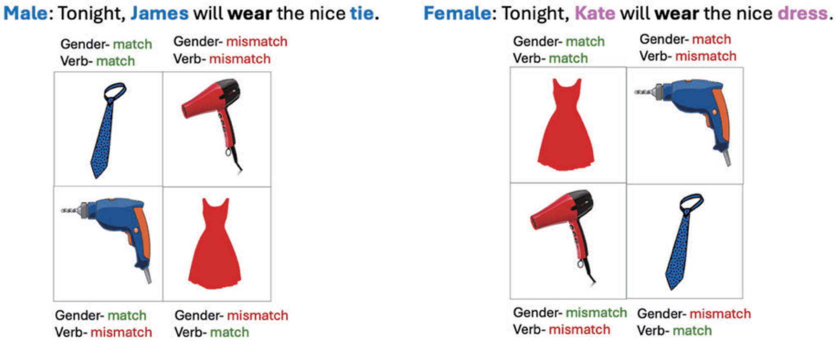


Figure 1. Stimuli design.

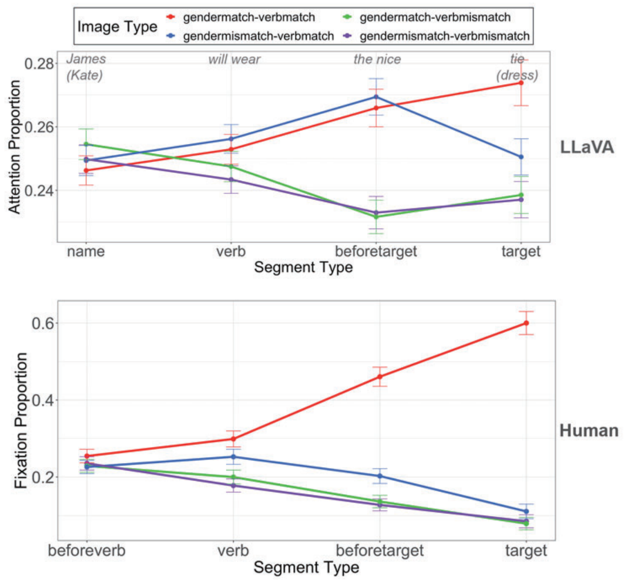


Figure 2. Compare attention proportion of LLAVA (top panel) and fixation proportion of human (bottom panel; data from Corps et al., 2022).

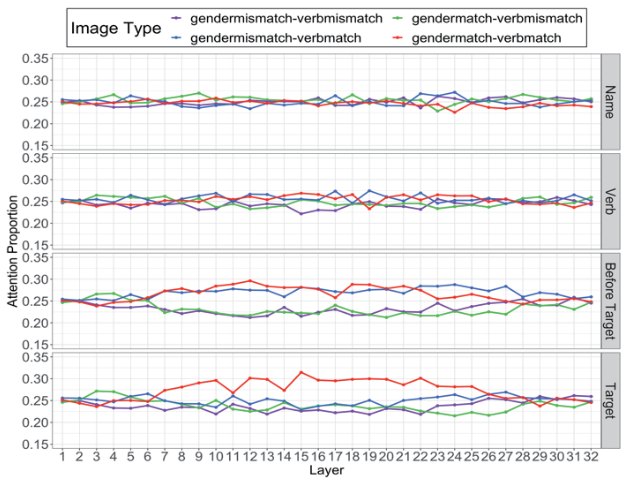


Figure 3. Attention proportion by layers.

References

Altmann, G. T., & Kamide, Y. (1999). *Cognition*, *73*(3), 247–264.

Corps, R. E., Brooke, C., & Pickering, M. J. (2022).. *Journal of Memory and Language*, *122*, 104298.

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003).. *Journal of Memory and Language*, *49*(1), 133–156.

Kewenig, V., Lampinen, A., Nastase, S. A., Edwards, C., DEstalenx, Q. L., Rechardt, A., Skipper, J. I., & Vigliocco, G. (2024). arXiv:2308.06035.

Liu, H., Li, C., Li, Y., & Lee, Y. J. (2023). arXiv:2310.03744.

OS.3.04

When A Man Says He Is Pregnant: ERP Evidence for A Rational Account of Speaker-contextualized Language Comprehension

Hanlin Wu & Zhenguang G. Cai The Chinese University of Hong Kong
hanlin.wu@link.cuhk.edu.hk

Spoken language is often, if not always, understood in a context that includes the identities of speakers. For instance, we can easily make sense of an utterance such as “I’m going to have a *manicure* this weekend” or “The first time I got *pregnant* I had a hard time” when the utterance is spoken by a woman, but it would be harder to understand when it is spoken by a man. Previous event-related potential (ERP) studies have shown mixed results regarding the neurophysiological responses to such speaker-mismatched utterances, with some reporting an N400 effect and others a P600 effect. In an experiment involving 64 participants, we showed that these different ERP effects reflect distinct cognitive processes employed to resolve the speaker-message mismatch. When possible, the message is integrated with the speaker context to arrive at an interpretation, as in the case of violations of social stereotypes (e.g., men getting a manicure), resulting in an N400 effect. However, when such integration is impossible due to violations of biological knowledge (e.g., men getting pregnant), listeners engage in an error correction process to revise either the perceived utterance or the speaker context, resulting in a P600 effect. Additionally, we found that the social N400 effect decreased as a function of the listener’s personality trait of openness, while the biological P600 effect remained robust. Our findings help to reconcile the empirical inconsistencies in the literature and provide a rational account of speaker-contextualized language comprehension.

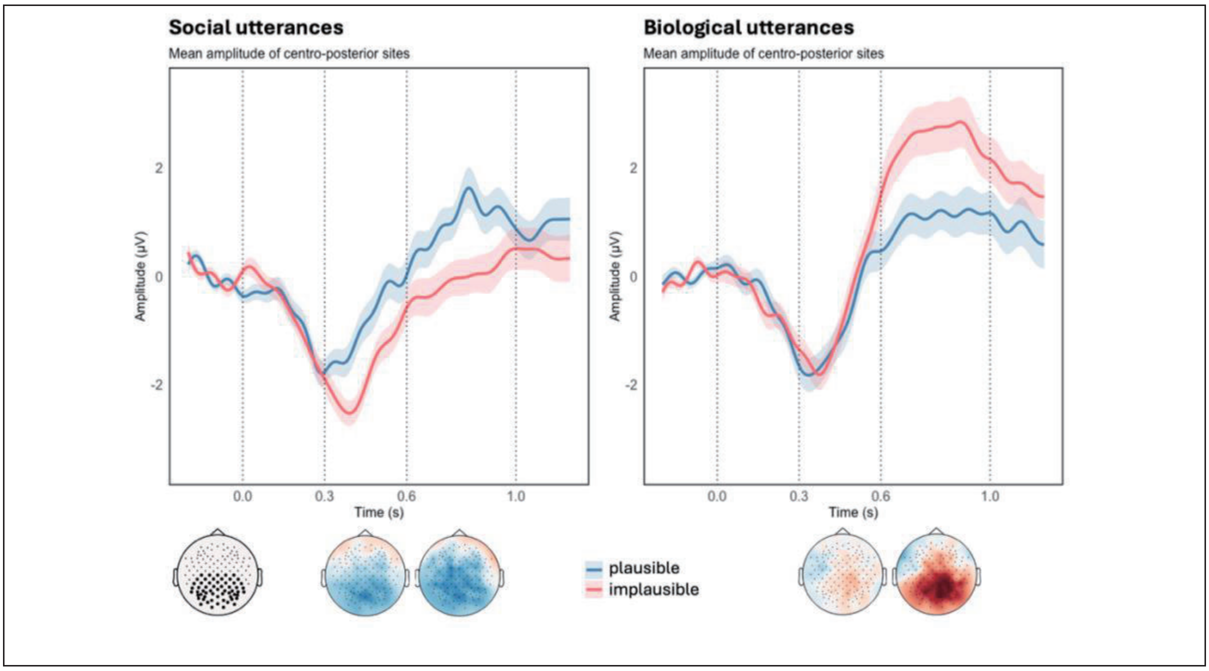


Fig 1. Brain potentials elicited by social and biological utterances during 300-600 ms and 600-1000 ms after the critical word onset; shaded areas represent SEs.

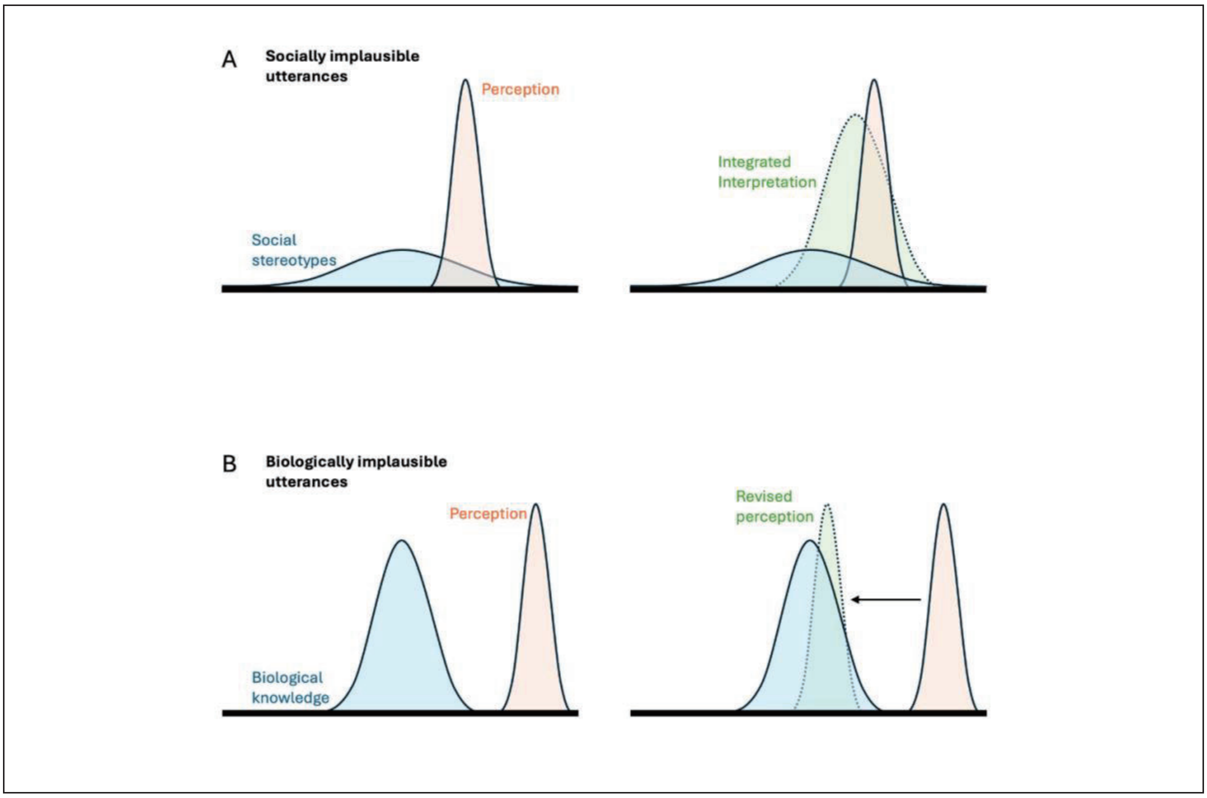


Figure 2. Schematic representation of speaker-contextualized language comprehension. (A) Upon hearing a socially implausible utterance, listeners integrate their perception of the utterance with their social stereotypes from the speaker context to arrive at an interpretation; (B) Upon hearing a biologically implausible utterance, listeners tend to revise their perception of the speech content or the speaker characteristics.

OS.3.05

Large language models but not humans are sensitive to the implied dialectic background of their interlocutor in word meaning access

Max S. Dunn[1] & Zhenguang G. Cai[2]
Hong Kong Polytechnic University[1], Chinese University of Hong Kong[2]
max-shaw.dunniii@polyu.edu.hk

Both humans and large language models (LLMs) are sensitive to the dialectic background of their interlocutor in language comprehension (e.g., Cai et al., 2017, 2023). For example, when hearing/reading cross-dialectally ambiguous words (e.g., *flat*) produced by an American English (AE) or British English (BE) interlocutor, both people and LLMs tend to access the dialect-specific meaning (i.e., *level* vs. *an apartment* for AE and BE interlocutors respectively). However, the dialectic background of an interlocutor can be inferred not only from their accent but also from their lexical use, spelling or cultural references. In this study, we test whether humans and LLMs are also sensitive to non-spoken dialectic cues in word meaning access.

In Experiment 1 (with 36 items), 42 human participants played a dialogue game with an interlocutor (in reality pre-scripted responses), where the interlocutor typed in a word according to a given definition and the participants typed the first word they thought of upon reading their partner's word. In target trials, the interlocutors typed a word that was a cross-dialectally ambiguous word (e.g., *flat*) and we were interested whether participants would access a meaning that was appropriate to the interlocutor's supposed dialect by examining the participant's associate (e.g., *level* vs. *housing* for AE and BE respectively). The interlocutor's dialectic background (AE vs. BE) was either explicitly mentioned to the participant (e.g., *I'm from America/England*; used as a baseline), implied via culture-specific references in the self-introduction (e.g., *I like to watch Saturday Night Live/East Enders*), implied via lexical use in the typed words (e.g., *vacation/holiday*), or implied via spellings in the typed words (e.g., *theater/theatre*). In none of the conditions did we observe a tendency for participants to access word meanings appropriate to the interlocutor's explicitly mentioned or implied dialectic background (see Fig. 1). These results suggest that human participants were not sensitive to these non-spoken dialectic cues in word meaning access.

Experiment 2 was similar to Experiment 1, except we tested ChatGPT as the participant. A Python script simulated a human interlocutor doing the experimental task with ChatGPT for 1,000 runs (i.e., equivalent to 1,000 participants), with an introduction given to ChatGPT that varied either the cultural references, lexical items, or spelling of AE or BE interlocutors. The results indicate that ChatGPT is sensitive to the dialectal background of their interlocutors in meaning access (see Fig. 2), with more AE associates when interacting with AE (50.7%) compare to BE interlocutors (34.8%), and with this effect being present in all conditions (p's $< .001$).

Overall, the results of Experiments 1 and 2 indicate that humans and LLMs are not fully equivalent linguistic comprehenders. LLMs are more sensitive to and make use of subtle text cues on implied dialectal background in word meaning access. Therefore, while LLMs have lesser linguistic abilities than humans in certain areas (e.g., Dentella et al., 2023), LLMs seem to be better than humans in modelling certain characteristics of interlocutors and/or applying these interlocutor models in language comprehension.

Figure 1: The results of Experiment 1.

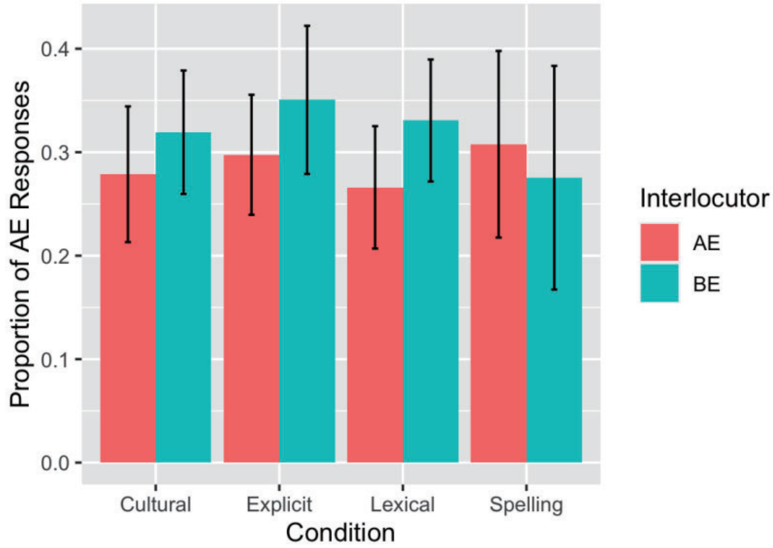
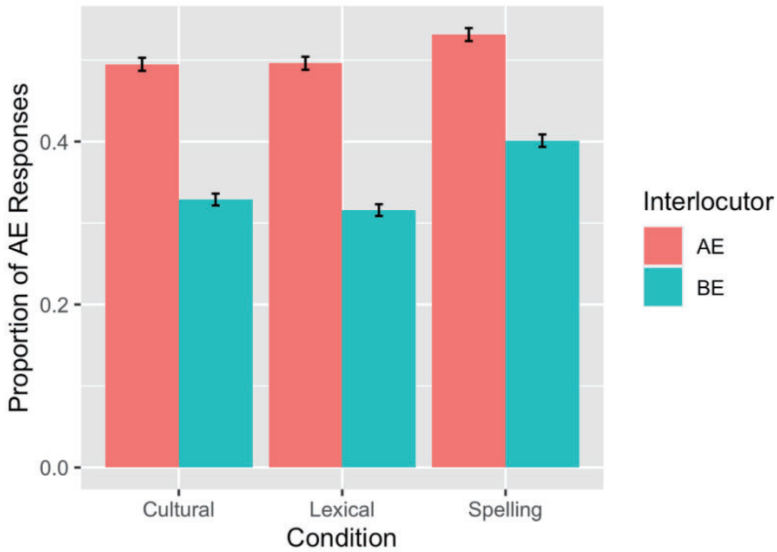


Figure 2: The results of Experiment 2.



Cai, Z. G. (2022). Interlocutor modelling in comprehending speech from interleaved interlocutors of different dialectic backgrounds. *Psychonomic Bulletin & Review*, *29*(3), 1026-1034.

Cai, Z. G., Gilbert, R. A., Davis, M. H., Gaskell, M. G., Farrar, L., Adler, S., & Rodd, J. M. (2017). Accent modulates access to word meaning: Evidence for a speaker-model account of spoken word recognition. *Cognitive Psychology*, *98*, 73-101.

Dentella, V., Günther, F., & Leivada, E. (2023). Systematic testing of three Language Models reveals low language accuracy, absence of response stability, and a yes-response bias. *Proceedings of the National Academy of Sciences*, *120*(51), e2309583120.

OS.3.06

An R Package for Behavioural Experimentation on Large Language Models

Xufeng Duan, Shixuan Li, Zhenguang G. Cai (The Chinese University of Hong Kong)

xufeng.duan@link.cuhk.edu.hk

There is growing interest in studying the behaviour of large language models (LLMs) through psychological experiments. In response to this, we have developed an R package that aims to interact with over 100 language models (e.g., GPT family, Llama family and various open-weight models) within a single framework, streamlining the process of experiment design, stimuli presentation, model manipulation and response/token probability logging.

We validated the package using three experiments with GPT-3.5 (proprietary model), Llama-2 7B (open-weight base model), and Vicuna-1.5 13B (open-weight fine-tuned model). These experiments aimed to replicate the findings of sound-gender association in LLMs, a phenomenon where models infer gender from novel personal names based on phonological cues (Cassidy et al., 1999, Cai et al. 2023). Our results consistently demonstrated that these models exhibit human-like tendencies to infer gender. The first experiment employed a multiple-trials-per-run design, presenting several experimental trials within a single conversational context. This method allowed us to observe cumulative effects and contextual influences on the models' responses. The second experiment utilized a one-trial-per-run design, isolating each trial to minimize contextual influence from preceding trials, thereby providing a clearer measure of the models' behavior in response to individual stimuli. The third experiment focused on analyzing token probabilities, examining the likelihood of gender-specific pronouns as the first token following a sentence fragment. This approach showed the models' initial response tendencies without the influence of subsequent context.

Our analyses revealed a significant sound-gender association across all models and experiments. Names ending with open syllables were more frequently associated with feminine pronouns, while those ending with closed syllables were more likely to be associated with masculine pronouns. These findings not only validate the functionality of this package but also underscore its potential for facilitating diverse and complex machine behavior studies.

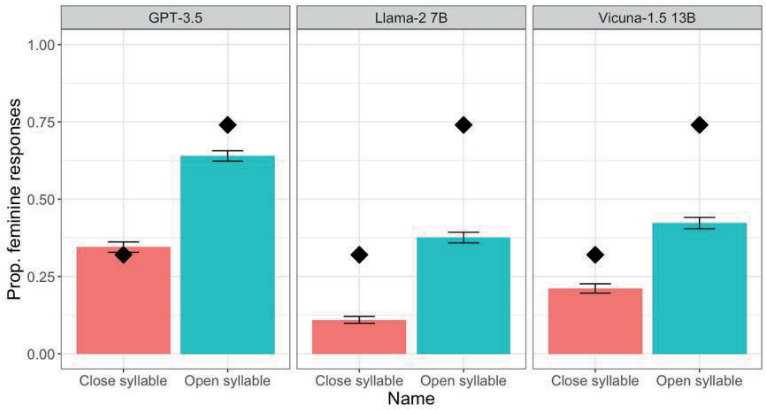


Figure 1: Proportion of feminine pronouns for novel names across different LLMs in Experiment 1. Error bars indicate 95% confidence intervals. The diamonds refer to conditional means in Cassidy et al., 1999.

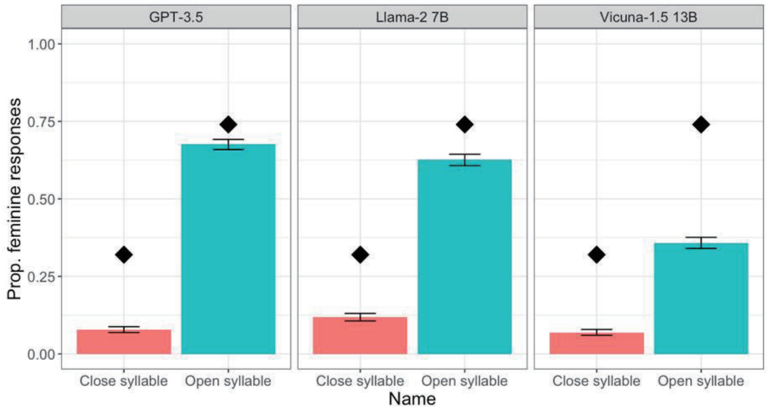


Figure 2: Proportion of feminine pronouns for novel names in Experiment 2.

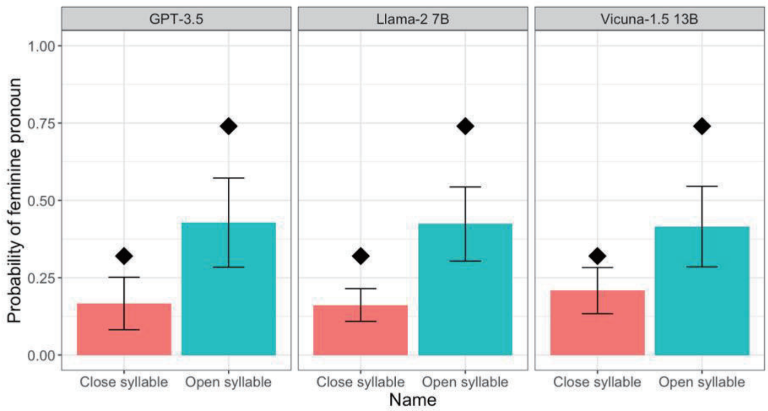


Figure 3: Probability results of all models in Experiment 3.

Reference

Cai, Z., Duan, X., Haslett, D., Wang, S., & Pickering, M. (2024). Do large language models resemble humans in language use?. In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics* (pp. 37-56).

Cassidy, K. W., Kelly, M. H., & Sharoni, L. A. J. (1999). Inferring gender from name phonology. *Journal of Experimental Psychology: General, 128*(3), 362.

OS.3.07

Mapping Language to Mind: Examining Linguistic Encoding of Space & Cognitive Effects

Pratik Nanda (Centre for Linguistic Science and Technology, Indian Institute of Technology Guwahati) Email: n.pratik@iitg.ac.in
Prof. Bidisha Som(Department of Humanities and Social Sciences, , Indian Institute of Technology Guwahati)
Dr. Cota Navin Gupta (Department of Biosciences and Biotechnology, Indian Institute of Technology Guwahati)

The concept of 'space' is among the fundamental notions in human thought and language. At this level, it is universal; however, remarkable relativity is also seen among the languages of the world, in terms of how this is 'broken down' into 'spatial categories'. This brings us to the question: What are the aspects of the visuo-spatial scene that play an important role in the linguistic and perceptual representation? Mental representation of spatial relations is claimed to be constructed upon a foundational set of "natural" concepts like adjacency, axial position, contact, and so on. Depending on language, one or more of these properties are considered while assigning spatial adpositions denoting location of objects. These aspects also affect the human understanding of spatial information and linguistic processing while engaging in mental processes like attention and memory. This study contributes to the existing knowledge in this domain by using behavioural and linguistic tasks involving the spatial primitives in a diverse set of languages Khasi, Odia, and Telugu, towards understanding if this is a universal phenomenon. This study explores the correspondence in the organization of spatial relations between linguistic representation and perceptual representation of space based on two primitives: Axial structure and contact(Fig:1,2). Adapting Munnich(2001), the visual stimuli consist of 144 scenes, each featuring two objects. In each image, a 9x9 grid is used, placing the reference object at the center, with the figure positioned in one of the surrounding 72 sections for spatial analysis(Fig: 3).

The study is divided into two stages: first, linguistic representation of space in different languages are collected and compared based on their syntactic and semantic behaviour based on a visual-scene description task. The spatial information encoded in languages are analysed and it reveals the importance of axis, contact, and distance while assigning the adposition. Results of the first experiment show that proportions of horizontal and vertical adpositions in describing visual stimuli support our hypothesis, with higher proportions of vertical PP used in describing vertical axis and horizontal PP in the horizontal axis(Fig: 4&5). The second task investigates whether the cross- linguistic differences in categorization of space(Sentences:1-3) affect the linguistic processing. Participants perform a matching task, judging whether the visual stimuli match the linguistic description. This ongoing study also explores primacy of axial and contact positions comparing response times and accuracy in judgment among groups. This task examines the impact of linguistic categorization on linguistic processing of the spatial information, potentially contributing to our understanding of how language influences cognition and perception. These experiments explore factors influencing the distribution of adpositions, as well as the prototypical and boundary areas based on the linguistic classification of spatial scenes.

Figures: Ā

Figure 1: Visual stimuli from Experiment(Axial & contact)Ā

Figure 2: Visual stimuli from Experiment(Axial & Non-contact)Ā

Ā

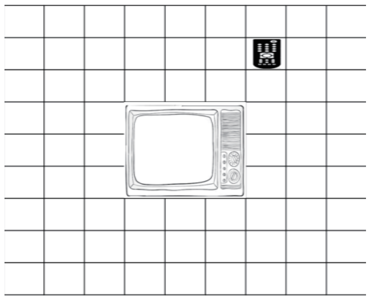


Figure 3: Design of the Visual stimuli from Experiment(Non-Axia & Non-contact). The grid section will not be visible in the experiment. These stimuli is used in both experimentsĀ

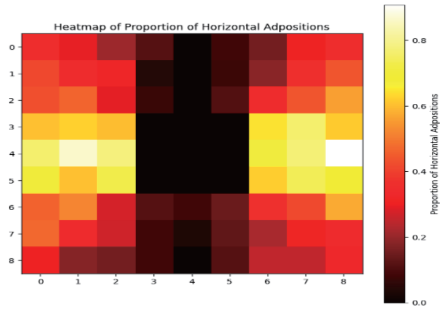
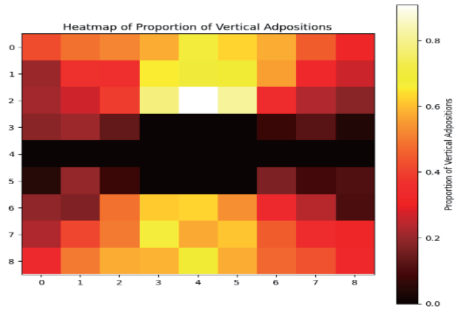


Figure 4 : Heatmap of proportion of Horizontal Adposition in OdiaĀ



Ā

Figure 5: Heatmap of proportion of Vertical Adposition in Odia Ā

Ā

Example sentences from the experiment for Figure 1 and Figure 2:

(1) Odia: biradi mɛdʒa upɔrɛ atʃʰi (Figure 1)   rimot tɪvɪ upɔrɛ atʃʰi (Figure 2)

cat table on be.PRS.3SG   remote TV above be.PRS.3SG

The cat is on the table.   The remote is above the TV.

(2) Telugu: pil:i tæbl mida undi (Figure 1)   rimot tɪvɪ mida undi (Figure 2)

cat table on be.PRS.3SG   remote TV above be.PRS.3SG

The cat is on the table.   The remote is above the TV.

(3) Khasi: *ka-miaw ka-don* *halor* *ka-mi* (Figure 1)

Fem -cat Fem-be.PRS.3SG on Fem-table

The cat is on the table.

u-rimot u-don *hadʒroŋ* *ka- tɪvɪ* (Figure 2)

Mas-remote Mas-be.PRS.3SG above fem-TV

The remote is above the TV.

About the Languages

Khasi is an Austroasiatic language spoken in Meghalaya, a north-east state of India. The language is rich in morphology, Characterized by complex word formation through affixation and compounding. It typically follows Subject-Verb-Object (SVO) word order.

Odia is an Indo-Aryan Language predominately spoken in Odisha, along with parts of neighbouring states like Jharkhand, Chhattisgarh, and West Bengal. It is one of the major languages spoken in India. The language has rich verbal agreement and follows SOV word order, although it can vary in certain contexts.

Telugu belongs to the Dravidian Language family. It is spoken in the southern part of India, primarily in the state of Telangana, Andhra Pradesh. It is an agglutinative language that follows SOV word order.

Each of these languages reflects their unique linguistic features, historical development, and cultural significance, contributing to the rich diversity of languages found in India.

OS.3.08

The influence of item repetition on bilingual language control

Authors: Hong Liu*, Adel Chaouch-Orozco, Darcy Sperlich
Correspondence author affiliation and email: Department of Applied Linguistics, Xi'an Jiaotong-Liverpool University, Suzhou, China; Hong.Liu@xjtlu.edu.cn

Language switching experiments, used to investigate bilingual language control, usually differ in the size of the stimuli lists, leading to variation in *item repetition* frequency. Critically, These variations can significantly influence the control mechanisms, potentially obscuring our understanding of the cognitive processes involved in efficient bilingual language use. As suggested by Kleinman and Gollan (2018), item repetition can contribute to the reversal of the commonly observed language dominance effect (i.e., overall larger response times in the L1)—allegedly, a marker of proactive control. Yet, this claim has never been tested. In addition, they showed that item repetition had little influence on language switching costs (i.e., longer response times in switch than in repeat trials)—a finding also observed in other studies (Costa & Santesteban, 2004; Shen & Chen, 2023). Crucially, all these studies tested highly proficient bilinguals, making it unclear how item repetition impacts language control in subjects with varying L2 proficiencies.

The present study addresses these two critical gaps, contributing to the debate on methodological considerations in the language control literature. (i) We examine how item repetition influences switching costs and the language dominance effect, and how these effects evolve dynamically as the task progresses. (ii) We also assess how item repetition affects bilinguals with varying L2 proficiencies. To do this, we conducted two experiments with two groups of 29 low- and 36 high-proficiency Chinese-English bilinguals performing a picture naming task in a cued mixed-language block. One experiment used a larger stimulus set, with each picture (n = 55) named once per language (minimal repetition); the other used a smaller stimulus set, with each picture (n = 9) repeated six times in each language (item repetition).

A linear mixed-effects model's analysis showed that switching costs differed in the two lists (Fig. 1). In the item-repetition list, switching costs were symmetrical across the two groups and languages. However, with minimal item repetition, a three-way interaction between trial type, proficiency, and language was observed (β = 63.77, t = 2.88, p < .01). Low-proficiency subjects yielded larger switching costs in the L1, while high-proficiency bilinguals had larger switch costs in the L2. Further, regardless of stimuli list, task progression led to symmetrical switching costs, caused by different effects on the responses to repeat trials in the two languages (Fig. 2). Finally, contrary to Kleinman and Gollan's (2018) claims, the language dominance effect did not change significantly with item repetition or task progression. These findings highlight the *critical methodological considerations* in language control experiments that often obscure our understanding of bilingual processing.

Figure 1. Predicted switch costs by experiment, language, and proficiency.

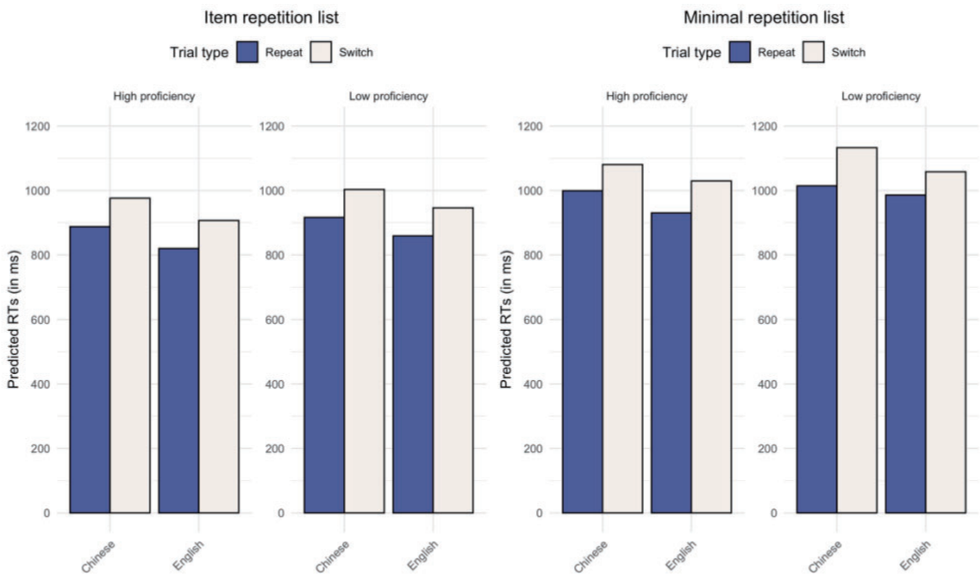
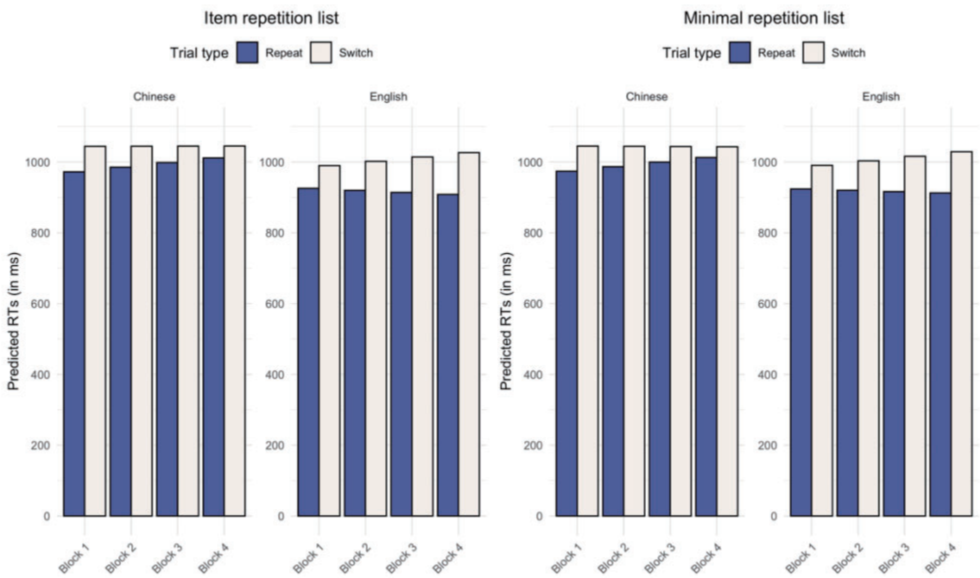


Figure 2. Predicted switch costs by experiment and task progression (four blocks).



References

Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language, 50,* 491–511.

Kleinman, D., & Gollan, T. H. (2018). Inhibition accumulates over time at multiple processing levels in bilingual language control. *Cognition, 173,* 115-132.

Shen, Q. & Chen Y. (2023). Investigating the modulation of stimulus types on language switching costs: Do semantic and repetition priming effect matter? *Frontiers in Psychology, 14,* 1090744.

OS.4.01

Mechanisms working behind the antecedent retrieval in the real-time processing of the reflexive dependency

Nayoun Kim (Sungkyunkwan University), Masaya Yoshida (Universitat Autòma de Barcelona)
nayoun@skku.edu

Introduction & Background Grammatically and semantically, the reflexive and the antecedent stand in a dependency relation ([1,2,3]). In this study, we investigate what information associated with the antecedent is retrieved when the reflexive is encountered ([4, 5, 6, 7]). We used an illusion of grammaticality to examine what is recovered in reflexive contexts, taking advantage of the finding that an ungrammatical verb following a local noun matching its morphological feature may often not be perceived as ungrammatical ([8], [9]). If the lexical, syntactic features, and structural information associated with the antecedent DP$_1$ is retrieved (**DP$_1$** the uncle [PP of [**DP$_2$** Thomas]]), then the head noun (or some content features of the head noun) embedded within the entire DP$_1$ is reactivated first, and the feature-matching operation under the content-addressable antecedent search ([4, 5, 6]) is deployed when the gender feature of the head of the DP$_1$ and the reflexive disagrees with each other. This would result in (1d) being read faster than (1b), with no differences in grammatical conditions in (1a) and (1c) (see Table 1). The second possibility is that the lexical and syntactic features are encoded in memory with no structural and relational information. Under this scenario, the head of the whole DP$_1$ (*uncle*) and the head of DP$_2$ (*Thomas*) which match the reflexive in lexical and syntactic features should be reactivated, where the retrieval of the two different DPs do not have a priority over one another. This would lead to the similarity-based interference effect ([10]). **Experiments** Two self-paced reading experiments were conducted. In the first experiment (N=76), we manipulated the *Gender* of the head noun in DP$_2$ (*Local Noun*: *Thomas/Katie* in Table 1) and the *Grammaticality* between the head noun of the DP$_1$ with the reflexive. To further examine whether the absence of the similarity-based interference effect in Experiment 1 supports the recovery of the structural information associated with the DP$_1$ including the DP$_2$, in the experiment 2 (N=84), we included another embedded noun within the DP$_2$ that matches lexical features of the reflexive (*Peter* in Table 2). Experiment 1 demonstrated a main effect of *Grammaticality* (*β = -0.10, SE= 0.02, t= -6.38*) as well as an interaction between *Local noun* and *Grammaticality* (*β = -0.07, SE= 0.03, t= -2.21*) at the critical region; further subset analysis revealed that an interaction was driven by the ungrammatical conditions (*β = 0.09, SE= 0.02, t= 3.53, p<0.001*) but not in grammatical conditions (Figure 1). Experiment 2 also demonstrated an interaction between *Local noun* and *Grammaticality* (*β = -0,06, SE= 0.02*, *t= -2.23*) driven by the ungrammatical conditions only (Figure 2). **Discussion & Conclusion** Taken together, the results of these experiments reveal that lexical and syntactic content as well as the structural information associated with the antecedent (e.g., [+Specifier of TP]), ([+Sister of P])) is retrieved at the reflexive. Local DPs that match the reflexives in gender feature provide an illusion of grammaticality, suggesting that the antecedent retrieval mechanism is sensitive to the structural details associated with the antecedent for the reflexives ([11], [12]).

References: [1] Chomsky (1981) [2] Chomsky (1986) [3] Reinhart (1976) [4] Lewis & Vasishth (2005) [5] Martin & McElree (2008) [6] Lewis, Vasishth, & Van Dyke (2006) [7] Kim, Brehm & Yoshida (2019) [8] Wagers, Lau, & Phillips (2009) [9] Lago, Shalom, Sigman, Lau, & Phillips (2015) [10] Jäger (2017) [11] Van Dyke & McElree (2011) [12] Kush, Lidz, & Phillps, 2015

Table 1: Example Stimuli for Experiment 1

Local Noun (Gender)	Grammaticality	Example sentence
1a. Match	Grammatical	**The uncle** of **Thomas** praised **himself** very harshly over lunch.
1b. Mismatch	Ungrammatical	**The uncle** of **Thomas** praised **herself** very harshly over lunch.
1c. Mismatch	Grammatical	**The uncle** of **Katie** praised **himself** very harshly over lunch.
1d. Match	Ungrammatical	**The uncle** of **Katie** praised **herself** very harshly over lunch.

Table 2: Example Stimuli for Experiment 2

Local Noun (Gender)	Grammaticality	Example sentence
2a. Extra Match	Grammatical	**The brother** of **Peter**'s uncle criticized **himself** very harshly at dawn.
2b. Match	Grammatical	**The brother** of **the uncle** criticized **himself** very harshly at dawn.
2c. Extra Match	Ungrammatical	**The sister** of **Peter**'s uncle criticized **himself** very harshly at dawn.
2d. Match	Ungrammatical	**The sister** of **the uncle** criticized **himself** very harshly at dawn.

Figure 1: Results of the Experiment 1 (the red box indicates the critical region right after the reflexive)

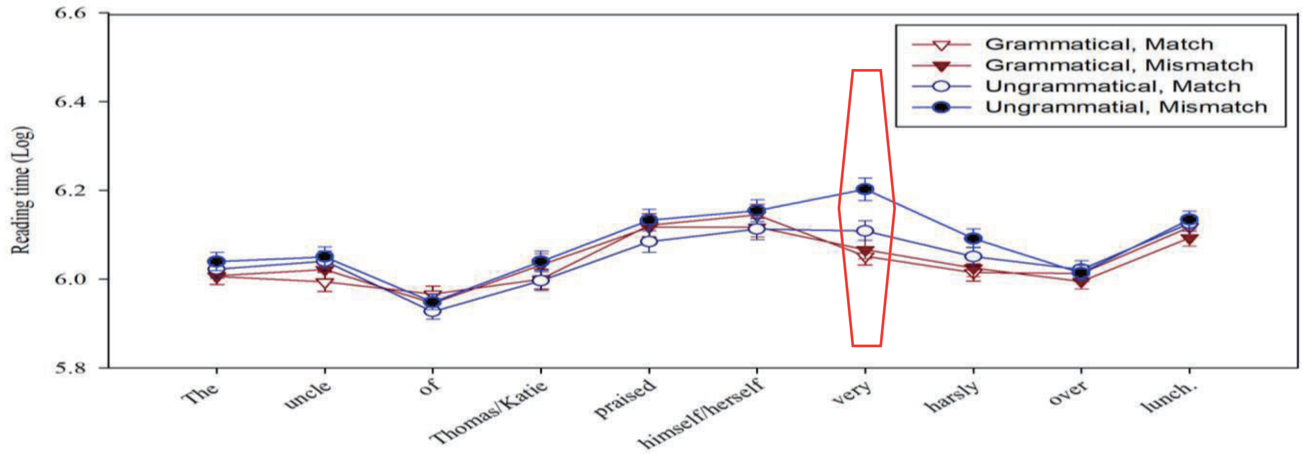
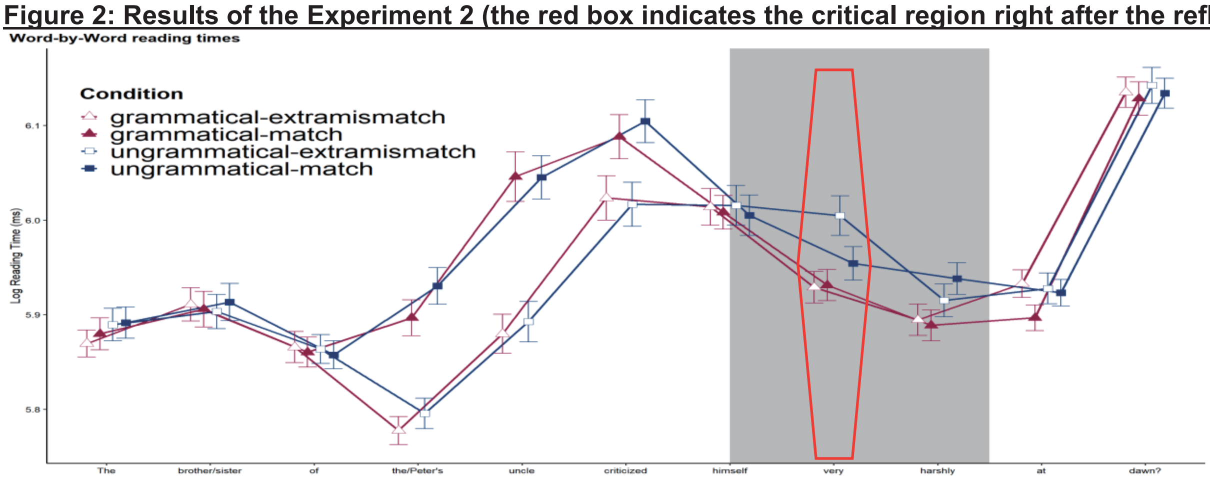


Figure 2: Results of the Experiment 2 (the red box indicates the critical region right after the reflexive)



OS.4.02

Structural constraints on processing compound reflexives in Mandarin: ERP Evidence

Chun-Chieh Hsu (cchsu@mx.nthu.edu.tw) & Chia-Hsuan Liao, National Tsing Hua University

This study investigated how structural constraints influence the processing of compound reflexives (*ta-ziji*) in Mandarin. Two recent eye-tacking studies adapting Sturt (2003)'s gender-(mis)match paradigm have reported inconsistent results. Chen (2018) observed a gender mismatch effect of local subjects only in the late measures, whereas Chang et al. (2020) found the same effect in the early measures. However, the latter finding was confounded by using the compound reflexive '*ta-ziji*' (他自己), which usually can refer to both male and female genders in Mandarin (Huang & Luh, 2012; Wu & Liang, 2008). We conducted two ERP experiments to examine how structural configuration may constrain online antecedent retrieval for female compound reflexives with a gender (mis)match paradigm. Exp.1 (n=29) involved a simple complement clause, with a matrix subject (MS) and an embedded subject (ES) (Table 1). Exp.2 (n=30) also had a matrix subject (MS), but with a relative clause subject (RS), that does not c-command the reflexive (Table 2). In both experiments, ERP analyses were time-locked to the onset of the female compound reflexive, *female-ta-ziji* (她自己), which unambiguously referred to a female antecedent. The results showed a P600 effect to gender mismatch from structurally accessible antecedents (Fig.1-2): in Exp.1, ES-mismatch conditions were more positive than ES-match ones; in Exp.2, MS-mismatch conditions were more positive than MS-match ones. Structurally inaccessible antecedents (MS in Exp.1/RS in Exp.2) did not modulate the brain responses, nor did they interact with structurally accessible antecedents. In addition, a referentially-related sustained frontal negativity, *Nref* (Van Berkum, 2007), was observed in both experiments, indicating possible online integration between structurally accessible antecedents and reflexives. Our findings provide strong evidence for immediate application of structural constraints on processing compound reflexives in Mandarin, and the lack of interference from structurally inaccessible antecedents argues against a standard cue-based account.

Table 1 Test paradigm for Exp.1

Condition	Matrix subject (R1)	Main verb (R2)	Embedded subject (R3)	Adverb (R4)	Embedded verb (R5)	Compound Reflexive (R6/Critical)	Conj. (R7)	(R8)	(R9)	(R10)
MS-match ES-match	張小姐 Ms. Zhang	說 says	李太太 Mrs. Li	曾經 once	推薦 recommend	她自己 herself	以及 and	其他 other	朋友 friends	參加比賽。 participate-in-contest
MS-match ES-mismatch	張小姐 Ms. Zhang	說 says	李先生 Mr. Li	曾經 once	推薦 recommend	她自己 herself	以及 and	其他 other	朋友 friends	參加比賽。 participate-in-contest
MS-mismatch ES-match	張先生 Mr. Zhang	說 says	李太太 Mrs. Li	曾經 once	推薦 recommend	她自己 herself	以及 and	其他 other	朋友 friends	參加比賽。 participate-in-contest
MS-mismatch ES-mismatch	張先生 Mr. Zhang	說 says	李先生 Mr. Li	曾經 once	推薦 recommend	她自己 herself	以及 and	其他 other	朋友 friends	參加比賽。 participate-in-contest
	Translation: *Ms/Mr. Zhang says that Mrs./Mr. Li once recommended herself and other friends to participate in the contest.*									

Note: MS = Matrix subject (structurally-inaccessible); ES = Embedded subject (structurally-accessible)

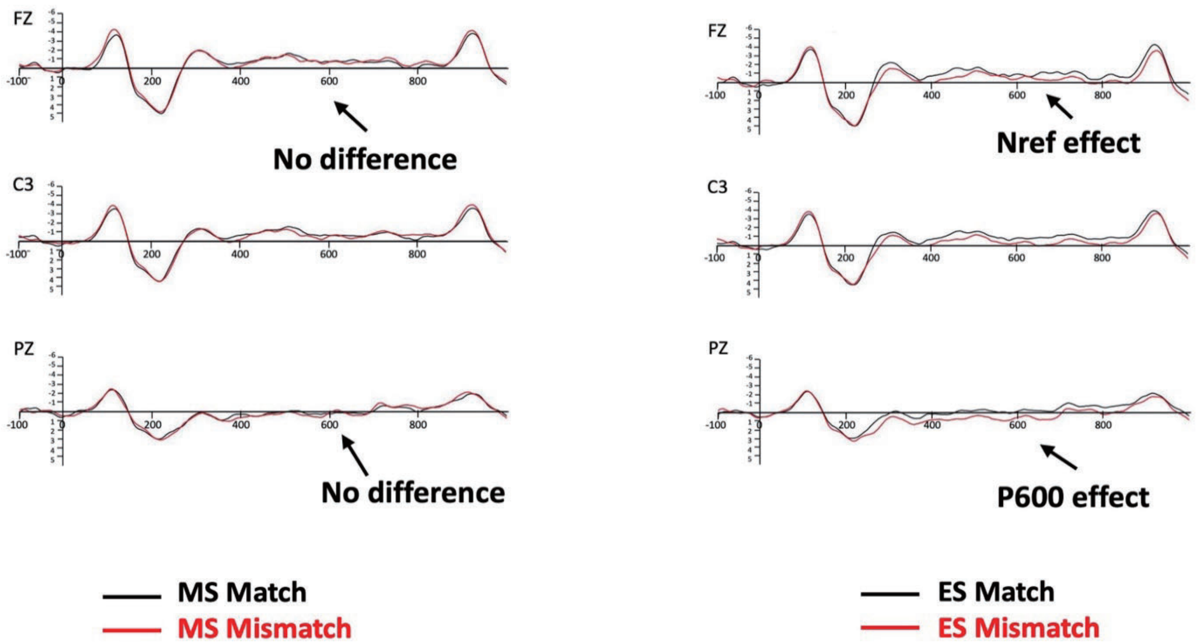


Fig.1 Grand averaged ERPs to gender mismatch effect at representative electrodes in Exp.1.

Table 2 Test paradigm for Exp.2

Condition	Matrix subject (R1)	Main verb (R2)	RC Subject (R3)	RC Verb (R4)	DE (R5)	Embedded Subject (Head noun) (R6)	Embedded verb (R7)	Compound Reflexive (R8/Critical)	Conj. (R9)	(R10)	(R11)
MC-match RC-match	王小姐 Ms. Wang	說 says	黃太太 Mrs. Huang	提供 provide	的 de	那份文件 that-CL document	幫助了 help	她自己 herself	並且 and	成為 become	證據。 evidence
MC-match RC-mismatch	王小姐 Ms. Wang	說 says	黃先生 Mr. Huang	提供 provide	的 de	那份文件 that-CL document	幫助了 help	她自己 herself	並且 and	成為 become	證據。 evidence
MC-mismatch RC-match	王先生 Mr. Wang	說 says	黃太太 Mrs. Huang	提供 provide	的 de	那份文件 that-CL document	幫助了 help	她自己 herself	並且 and	成為 become	證據。 evidence
MC-mismatch RC-mismatch	王先生 Mr. Wang	說 says	黃先生 Mr. Huang	提供 provide	的 de	那份文件 that-CL document	幫助了 help	她自己 herself	並且 and	成為 become	證據。 evidence
	Translation: *Ms/Mr. Wang says that [the document that Mrs./Mr. Huang provided ___] helped herself and became evidence.*										

Note: MS = Matrix subject (structurally-accessible); RS = Relative-clause (RC) subject (structurally-inaccessible)

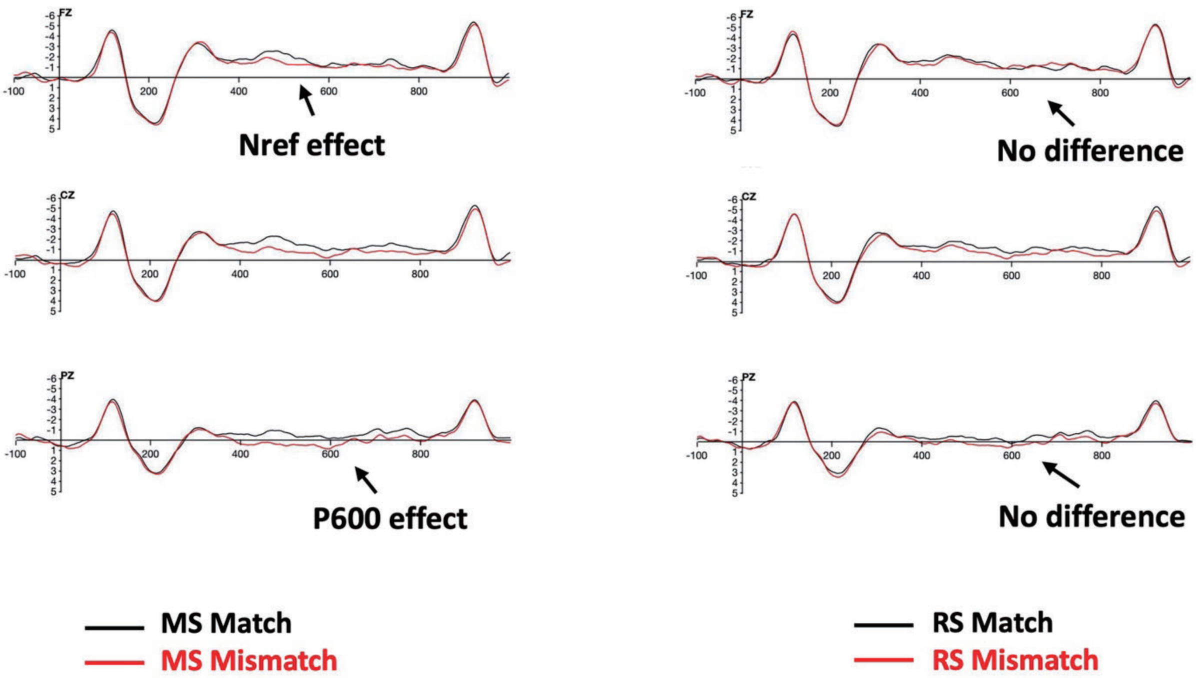


Fig.2 Grand averaged ERPs to gender mismatch effect at representative electrodes in Exp.2.

Selected References:

Chang, W., Duan, Y., Qian, J., Wu, F., Jiang, X., & Zhou, X. (2020). Gender interference in processing Chinese compound reflexive: evidence from reading eye-tracking. *Language, Cognition and Neuroscience*, *35*(10), 1355-1370.

Chen, Szu-En, (2018), 'Online processing of reflexives in L1 Mandarin and L2 English: An eye-tracking study', Master Thesis, National Tsing Hua University, Taiwan.

Huang, C.-H., & Luh, W.-M. (2012). Analysis of third person pronouns with gender semantic features in elementary school Mandarin textbooks. *Journal of Textbook Research.* 5, 85–113.

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of memory and language, 48*(3), 542-562.

Van Berkum, J. J. A., Koornneef, A. W., Otten, M., & Nieuwland, M. S. (2007). Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research, 1146*, 158–171.

Wu, D. S., & Liang, T. (2008). Chinese pronominal anaphora resolution using lexical knowledge and entropy-based weight. *Journal of the American Society for Information Science and Technology*, *59*(13), 2138-2145.

OS.4.03

Processing of Mental and Physical Experiencer Verbs in Complex predicates: Electrophysiological evidence from Malayalam.

S Shalu[1], Kamal Kumar Choudhary[1], R Muralikrishnan[2]
[1]Indian Institute of Technology, Ropar, India; [2]Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany.
Shalu.20hsz0009@iitrpr.ac.in; r.muralikrishnan@aesthetics.mpg.de

Understanding and categorizing entities and activities as experiencers or non-experiencers are fundamental to how we navigate and interpret the world. This cognitive significance of experiencers has consequences for the grammar of a language as well. Evidence from various languages indicates that the morpho-syntactic structures of experiencer verbs and experiencer arguments show significant variation across languages (Belletti & Rizzi, 1988; Dabrowska,1994; Jayaseelan, 2004). Previous ERP studies have investigated the processing of experiencer verbs, focusing particularly on grammatical function and thematic role reanalysis, compositional semantics, and conceptual representations, with a special emphasis on object experiencer verbs. But, to the best of our knowledge, no studies to date have addressed the question of whether different kinds of subject experiencer verbs, indicating distinct kinds of experiences are processed qualitatively similarly or differently. To this end, we conducted an ERP study on Malayalam, a Dravidian language in which both complex intransitive mental experiencer verbs (indicating experience of emotions such as happiness or anger) and physical experiencer verbs (indicating the bodily sensation and other bodily experiences) takes dative case on subject arguments (e.g., 'enikkə santhosham vannu- I(DAT) became happy' and 'enikkə pani vannu- I(DAT) became feverish') (Jayaseelan, 2004). We employed a 2 x 2 design, manipulating the factors CASE (Nominative Vs. Dative) and VERB (Mental Vs. Physical experiencer) to examine the possible processing differences between mental and physical experiencer verbs. Stimuli of the form in Table1 were presented in a rapid serial visual presentation. Conditions with nominative subjects were anomalous. EEG from 27 native-speakers of Malayalam (mean age: 27.4; 13 female; 14 male) was recorded when they read the sentences and performed acceptability judgement and probe detection tasks. The ERP results at the critical position i.e., the verb, revealed that the violation conditions with both types of experiencer verbs evoked an N400 effect as opposed to their respective correct counterparts. We interpret the N400 as resulting from a prediction mismatch, as follows. The prominence information extracted from the sentence-initial nominative subject in the violation conditions would ideally predict for an intransitive construction with an action verb (Droge, Maffongelli, & Bornkessel-Schlesewsky, 2014; Gattei, Tabullo, París, & Wainselboim, 2015). But when an experiencer verb that requires the subject to be in dative case is encountered, this results in a prediction mismatch (Gattei et al., 2015; Bornkessel-Schlesewsky & Schlesewsky, 2019). Since such a construction is irrecoverably anomalous, this results in an

N400 effect. This finding suggests that both physical and mental experiencer complex predicates in Malayalam are processed qualitatively similarly.

KEYWORDS: Mental experiencer verb , Physical experiencer verb, case, N400

Supplementary material.

Condition	Sample Stimulus
DAT-MENT	Ravi- kkə santhosham vannu. Ravi-DAT-An.M happiness come.PST Ravi became happy.
NOM-MENT	*Ravi santhosham vannu. Ravi-NOM-An.M happiness come.PST Ravi became happy
DAT-PHY	Ravi-kkə vishappu vannu. Ravi-DAT-An.M hunger come.PST Ravi became hungry.
NOM-PHY	*Ravi vishappu vannu. Ravi-NOM-An.M hunger come.PST Ravi became hungry.

Reference:

1. Belletti, A., & Rizzi, L. (1988). Psych-verbs and θ-theory. *Natural Language & Linguistic Theory*, 291-352.
2. Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2019). Toward a neurobiologically plausible model of language-related, negative event-related potentials. *Frontiers in psychology*, *10*, 435509
3. De Swart, P. J. F. (2007). *Cross-linguistic variation in object marking*. Utrecht<>: LOT<> UB Nijmegen Host.
4. Dąbrowska, E. (1994). Dative and nominative experiencers: two folk theories of the mind.
5. Droge, A., Maffongelli, L., & Bornkessel-Schlesewsky, I. (2014). Luigi piace a Laura?: Electrophysiological evidence for thematic reanalysis with Italian dative object experiencer verbs. In *Structuring the Argument* (pp. 83-118). John Benjamins..
6. Eberhard, D.M., Simons, G.F.,& Fenning, C. D. (Eds.). (2019). *Ethnologue: Languages of the World* (22nd ed.). SIL International. Retrieved from Ethnologue: Languages of the World
7. Gattei, C. A., Tabullo, Á., París, L., & Wainselboim, A. J. (2015). The role of prominence in Spanish sentence comprehension: An ERP study. *Brain and language*, *150*, 22-35.
8. Krishnan, G. G. (2019). Malayalam morphosyntax: Inflectional features and their acquisition. *Unpublished PhD dissertation. IIT: Bombay*.
9. Jayaseelan, K. A. (2004). The possessor-experiencer dative in Malayalam. *Typological Studies in Language*, *60*, 227-244.
10. Kaan E, Swaab TY (2002): The brain circuitry of syntactic comprehension. Trends Cogn Sci 6:350–356.
11. Kaan E, Swaab TY (2003): Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. J Cogn Neurosci 15:98–110.
12. Krishnamurti, B. (2003). *The Dravidian languages*. Cambridge University Press.

13. Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology*, *62* ,621-647.
14. Nizar, M. (2010). Dative Subject Constructions in South-Dravidian Languages. *Unpublished master's thesis). University of California,* Berkely.

Malayalam Language

Malayalam is an agglutinative South Dravidian language with a canonical SOV order, spoken by about 34.8 million speakers as their first language (Eberhard et al., 2019), mainly in the state of Kerala and union territories Lakshadweep and Puducherry in India. . Malayalam emerged as a separate language from the west coast dialect of Tamil around the 9th century (Krishnamurti, 2003), and is characterized by its rich inflectional morphology, including a three-way tense marking system (verbs are inflected for present, past and future), three genders (Masculine/feminine/neuter) and two numbers (singular/plural). Malayalam exhibits differential object marking, whereby only animate objects are marked accusative, with inanimates marked accusative only as an exception. For instance, inanimate objects can be marked when there is a potential ambiguity in interpreting the sentence if left unmarked. This situation is particularly common in sentences where the subject is also inanimate (de swart; 2007). Both the subject and object arguments can be omitted from a sentence if they can be understood from the context (Krishnan, 2019). The feature that makes Malayalam distinct from other sister languages in the Dravidian language family is the absence of subject-verb agreement.

In Malayalam, like in other Dravidian languages, grammatical relations and semantic roles are indicated through a series of case suffixes. Thus, changing the word order in a sentence typically does not change its meaning, as the roles and relations are primarily conveyed through these suffixes. The case system of Malayalam includes seven cases: nominative, accusative, dative, sociative, locative, genitive and instrumental. Malayalam exhibits non-nominative subject constructions (Nizar, 2010), a wide-spread areal feature of South Asian languages. Of these, the dative subject constructions are by far the most common and can be primarily found in three domains: experiencer predicates, possession constructions and modal constructions. Mental experiencer verbs (e.g., *'njaan santhoshichchu - I became happy'*) take nominative subjects when the predicate is simple but require dative subjects when the predicate is complex (e.g., *'enikku santhosham vannu - lit. Happiness came to me'*). By contrast, physical experiencer verbs require dative subjects in both simple (e.g., *'enikku vishannu - I became hungry'*) and complex constructions (*'enikku vishappu vannu - lit. Hunger came to me'*) (Jayaseelan, 2004).

OS.4.04

Processing the Fuzzy Meaning of Mandarin "Nong-Complement" Structure

Mei-Chun, Kuo, Doctor's Program in TCSL, National Chengchi University, Taiwan
Yao-Ying, Lai, Graduate Institute of Linguistics, National Chengchi University, Taiwan
110160501@nccu.edu.tw

In Mandarin Chinese (MC), the verb "*nong*" is semantically underspecified, denoting a vague action like "do," "make," or "deal with." In this study we examine the [*Nong*[V]-Complement] structure, where the complement indicates the result caused by the action that the verb *nong* denotes. For example, (a) indicates some action that causes a cup to break, without action specification. In contrast, the action in (b) is specified by the underlined morpheme.

(a) *nong-po beizi* nong-break cup 'broke a cup'
(b) *shuai-po beizi* drop-break cup 'broke a cup by dropping it'

The [*Nong*[V]-Comp.] structure underspecifies the manner of the action that results in the status denoted by the complement. We investigate how comprehenders process such fuzzy resultative verb compound to address the research question: do speakers access one—likely the preferred—reading for [*Nong*-Comp.] in line with the Multi-Entry account [2], or leave its meaning underspecified [1]?

We address the question via a norming study and a self-paced priming experiment. In the norming study, we conducted a fill-in-the-blank task in which the participants indicated the preferred causative action for a given [__[V]–Comp.] phrase. Responses from 52 native speakers of MC were collected to select the preferred (i.e., high-frequent, HF) action meaning and a less preferred (low-frequent, LF) one as the well-specified counterparts of [*Nong*-Comp.]. The ensuing self-paced priming experiment examined if speakers access a preferred reading for the [*Nong*-Comp.] phrase by probing the priming effect in the following HF vs. LF counterpart. Seventy native speakers of MC read two VP phrases in each trial (Fig.1). If comprehenders access a preferred action meaning for [*Nong*-Comp.], the HF condition would elicit shorter reading times (RTs) than LF at the critical Word3. A RepeatedV condition with a verb repeated in Word1 and Word3 was introduced for baseline comparison, along with other fillers. Results showed a significant effect of Condition at Word3 ($\chi^2(2) = 11.49$, $p < .01$), such that {HF≅LF} > RepeatedV in RTs (Fig. 2). The HF and LF did not differ significantly however, supporting the Underspecification account. The findings suggest that during the processing of a [*Nong*[V]-Comp.] with fuzzy verb meaning, comprehenders tend to retain an underspecified interpretation. This parallels the mechanism applied to polysemous words, where the meaning remains underspecified and adaptable to contextual nuances [3][4].

Norming study: Fill-in-the-blank task – sample sentence

(1) *Mínghuá ___pò yíge bēizi.*

Minhua ___-broke.COMP one-CL cup.

'Minhua broke a cup by ____.'

Figure 1. Trial procedure of self-paced priming experiment

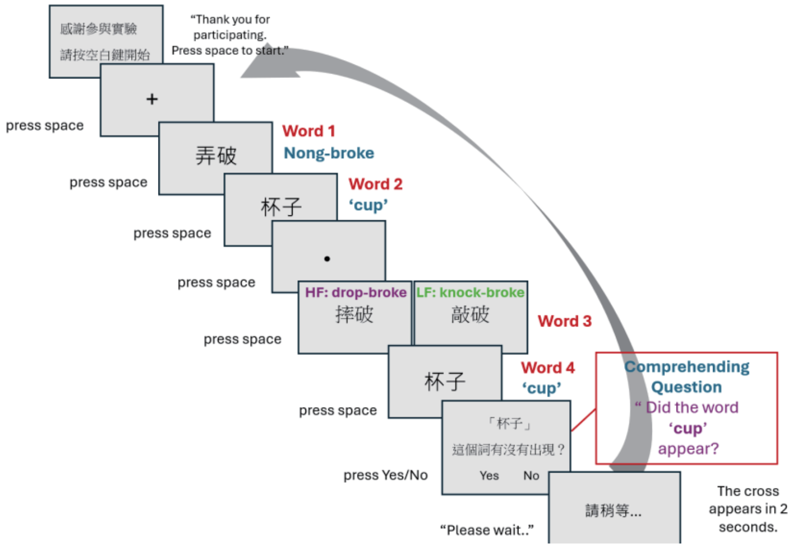
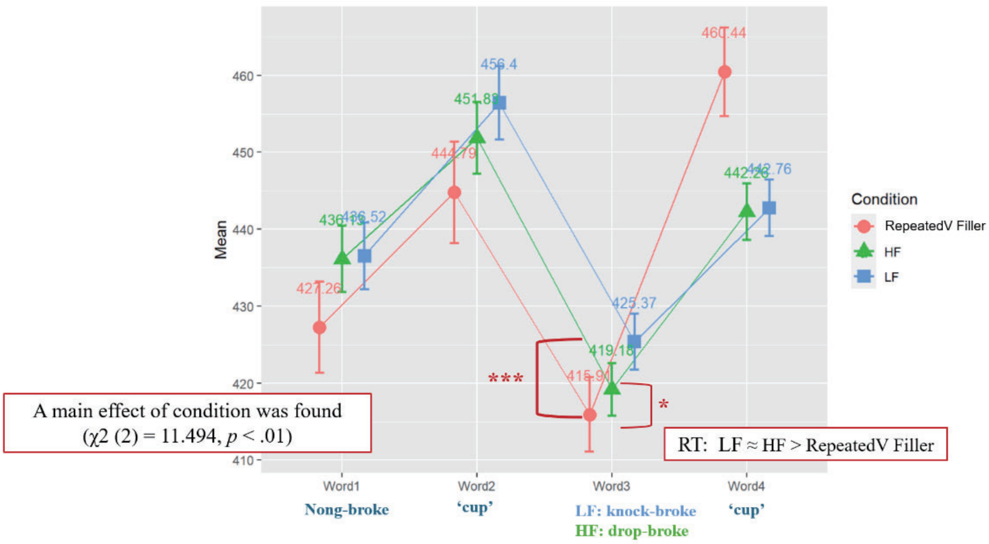


Figure 2. Result of self-paced priming experiment: RTs in millisecond (mean ±1 se) – exemplified by *nong-po* (nong-break)



References: [1] Frisson, S. (2009). Semantic Underspecification in Language Processing. *Language and Linguistics Compass, 3*(1), pp.111-127. **[2]** Klein, D. E., & Murphy, G. L. (2001). The Representation of Polysemous Words. *Journal of Memory and Language* 45, pp.259-282. **[3]** Wittenberg, E., & Piñango, M. M. (2011). Processing light verb constructions. *The Mental Lexicon*, 3(6), pp. 393-413. **[4]** Frisson, S. (2015). About bound and scary books: The processing of book polysemies. *Lingua*, 157, pp.17-35.

OS.4.05

Exploring connections between different higher-level language processes

Sonny. R. Wang & Craig. G. Chambers, University of Toronto

To what extent do distinct higher-level language processes draw on shared resources? Although this question has been explored using offline measures of reading performance and correlations with general cognitive ability, few studies **directly** assess relationships between language processes, and even fewer assess **online** comprehension. In two pre-registered, high-powered studies, we used an individual differences approach and real-time measures to investigate connections between the (i) generation of **bridging** inferences (argued to draw on sentence-based situational reasoning, cf. [1]) and two other processes, namely (ii) **instrument inference** (argued to draw on semantic association, cf. [2]), and (iii) **linguistic perspective-taking** (in which readers must track characters' knowledge states, [3]).

Expt. 1 **(*N*=150) tested** for correlations between **bridging** and **instrument** inferencing. Twelve critical passages, testing either bridging or instrument inferences (validated in previous studies), appeared one sentence at a time, randomized and intermixed with fillers. Target sentences were either congruent or incongruent with inferences that could be drawn upon reading earlier sentences (see Table 1). Results for each task replicated previous patterns [4,5], whereby participants read incongruent targets more slowly than congruent ones (see Fig. 1A/B), consistent with spontaneous inferencing. Importantly, individual performance in the two tasks were **positively correlated** (Table 2, left-side), suggesting that bridging and instrument inferences draw on shared resources.

Expt. 2 **(*N*=200) tested** for correlations between bridging inference (using similar materials as in Expt. 1) and **linguistic perspective-taking** (also validated separately), where a story character interprets a written message as either sincere or sarcastic [6], and where the 'correct' interpretation hinges on whether the reader effectively tracks characters' knowledge states (see **Table 1**). Results showed that, as with bridging inference, linguistic perspective-taking occurred spontaneously during reading (longer average RTs for incongruent- vs. congruent perspective, see **Fig. 1C**). Critically, however, **no correlation** was found between participants' performance on the bridging inference and perspective-taking materials (Table 2, right-side).

These are the first individual differences studies directly associating multiple types of higher-level language processes using validated real-time paradigms. Expt. 1 results suggest a shared processing substrate for bridging and instrument inferencing despite their (assumed) distinct underpinnings. Expt. 2 results suggests that, although sometimes labelled as 'inferencing', linguistic perspective-taking involves comparatively distinct resources. Together, the results provide new insights into the cognitive architecture supporting real-time comprehension.

References. [1] Singer & Halldorson, 1996. *Cognitive Psych*. [2] McKoon & Ratcliff, 1981. *Jnl of Verbal Learning & Verbal Behavior.* [3] Keysar, 1994. *Cognitive Psych.* [4] Singer et al., 1992. *Memory & Cognition.* [5] Vukic et al., 2009. *Disc. Processes*. [6] Weingartner & Klin, 2005. *Memory & Cognition*.

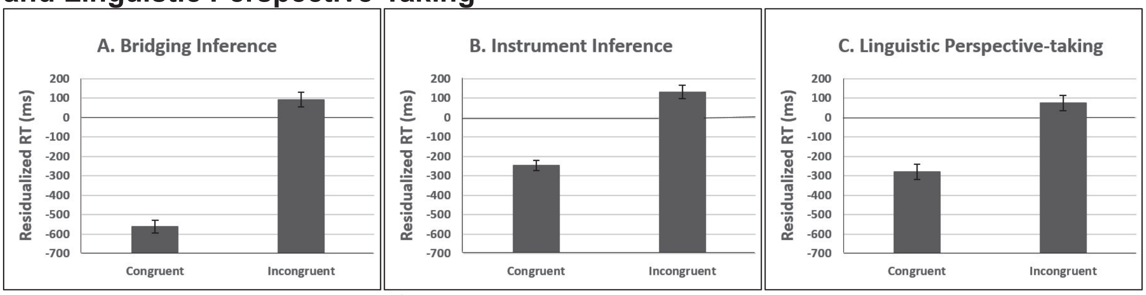
Table 1. Example Materials

Bridging Inference (Expt. 1 AND 2. Note: not all sentences are shown)	
Antecedent	Carol went outside and had forgotten about the bread she was baking in the oven. *(Bridging Inference –> oven is still on; bread is still inside)*
Target (congruent)	When she finally came back inside, Carol saw smoke coming from the oven.
Target (incongruent)	When she finally came back inside, Carol decided to turn on the oven.

Instrument Inference (Expt. 1. Note: there are 5 sentences in each passage, not all are shown)	
Antecedent	Dan was building a new shed he had been planning all winter. For the past ten minutes, he had been pounding nails into the boards. *Instrument Inference -> [Dan is using a] hammer*
Target (congruent)	"Lucky I got a hammer!" Dan thought to himself.
Target (incongruent)	"If only I had a hammer!" Dan thought to himself.

Linguistic Perspective Taking (Expt. 2. Note: not all sentences are shown)	
Background	…Susan had a bad experience watching Kurt's dog for him… …She left a note on Kurt's door saying, 'Wonderful dog sitting experience.'… After she left, a delivery man came to the door and saw the note. (*Correct perspective -> naïve delivery man should NOT perceive note as sarcastic*)
Target (congruent)	He was happy this dog sitter had really enjoyed the experience.
Target (incongruent)	He felt sorry that this dog sitter had not enjoyed the experience.

Figure 1. Condition Effects for Bridging Inferencing, Instrument Inferencing and Linguistic Perspective-Taking



All studies were conducted online using Gorilla. Data screening and preparation steps: [i] make sure all participants have full number of observations; [ii] eliminate inattentive participants; [iii] calculate residualized reading times by participant to control for sentence length effects and intra-individual differences in reading speed.

Table 2. Correlations based on rankings of participants' individual scores

Bridging Inference ~ Instrument Inference		Bridging Inference ~ Perspective-taking	
r	*p*	*r*	*p*
0.20	***	0.06	0.38

OS.4.06

Capturing context effects on garden-pathing and reanalysis with a latent process model

Dario Paape & Shravan Vasishth, University of Potsdam (paape@uni-potsdam.de)

A classic debate in sentence processing research concerns the question of whether discourse context can prevent or reduce garden-path effects. Consider the following discourses:

> A fireman was talking to [a man and a woman] / [two women]. He told the woman
> that he had rescued . . . . . . her dog as well. **Complement clause, CC**
> . . . to get a blanket. **Relative clause, RC**

The relative clause analysis is arguably disfavored by the parser because it contains a filler-gap dependency [1]. However, Altmann, Garnham & Dennis (1992) [2] reported that average reading times did not differ between the CC and RC continuations when the context mentioned two women, licensing the use of a restrictive RC. Two questions are left open by this result:

1. Does the RC-supporting context reduce the *probability* of garden-pathing by the RC continuation, or does it merely reduce the cost of syntactic reanalysis, or a mixture of both?
2. Does the CC continuation cause garden-pathing and reanalysis in the RC-supporting context, as there is no unique woman to refer to?

A recent computational model of garden-pathing and reanalysis [3] can answer these questions by looking beyond condition means. The model assumes nested latent processes that occur serially and probabilistically (Figure 1), and takes into account several insights from the literature:

- Garden-pathing may be stochastic [4]: different analyses are adopted in different trials.
- Reanalysis is optional ("triage", [5,6]) and can be overt (rereading) or covert [7].

We conducted a replication of [2] with 160 participants using bidirectional self-paced reading [8] and end-of-trial acceptability judgments. Based on the judgments, reading times at the disambiguating region, and presence or absence of a regression (anywhere after the disambiguation), the process probabilities and costs in the model can be estimated. We also investigate how disambiguation type and context/disambiguation match affect process probabilities and costs.

The parameter estimates in Figure 2 reveal several important insights:

- There is no baseline difference in the probability of garden-pathing between CC and RC continuations, but context boosts the matching analysis.
- Covert reanalysis (without rereading) is more costly for RC than for CC continuations, but its cost is not affected by contextual match.
- Overt reanalysis (with rereading) is more likely to succeed for RC continuations than CC continuations, and in matching compared to non-matching contexts.

Our modeling results suggest that the parser is not biased against relative clauses, contra [1] but in line with [9]. Scanpaths (Figure 3) reveal disruptions in reading progression due to context mismatch regardless of RC or CC disambiguation, but RC continuations show higher acceptability after rereading. Overt CC-to-RC reanalysis thus appears to be easier than overt RC-to-CC reanalysis, which casts further doubt on an intrinsic anti-RC bias.

Latent process modeling reveals that context affects first-pass parsing choices *and* reanalysis. A further novel finding is that the relative difficulty of CC-to-RC and RC-to-CC reanalysis, and the influence of contextual match, depends on whether rereading is involved or not.

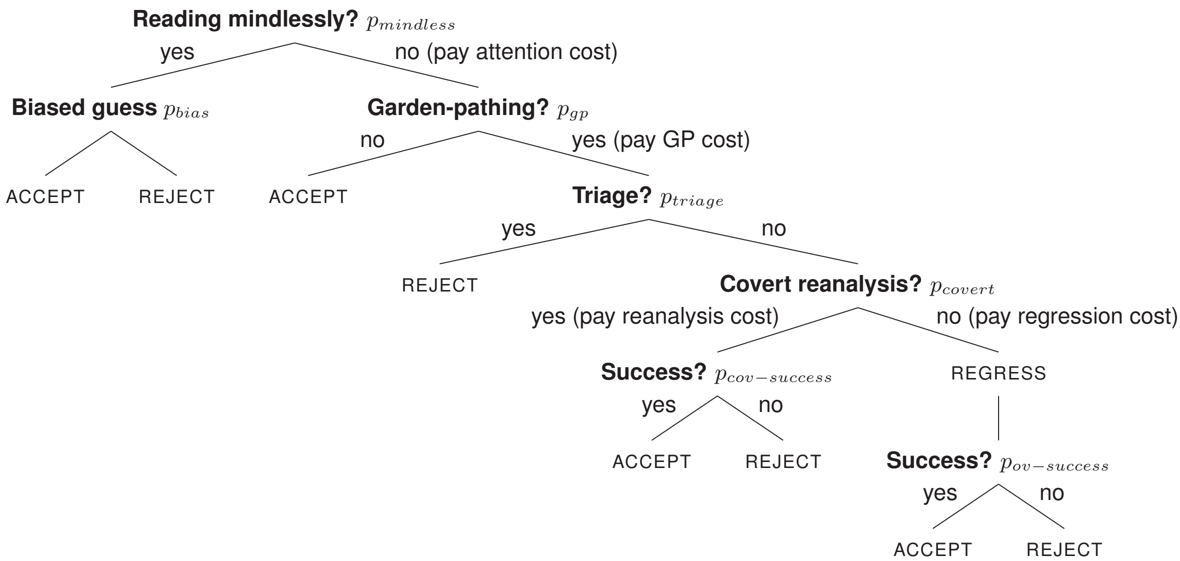


Figure 1. Multinomial processing tree model of garden-pathing and reanalysis.

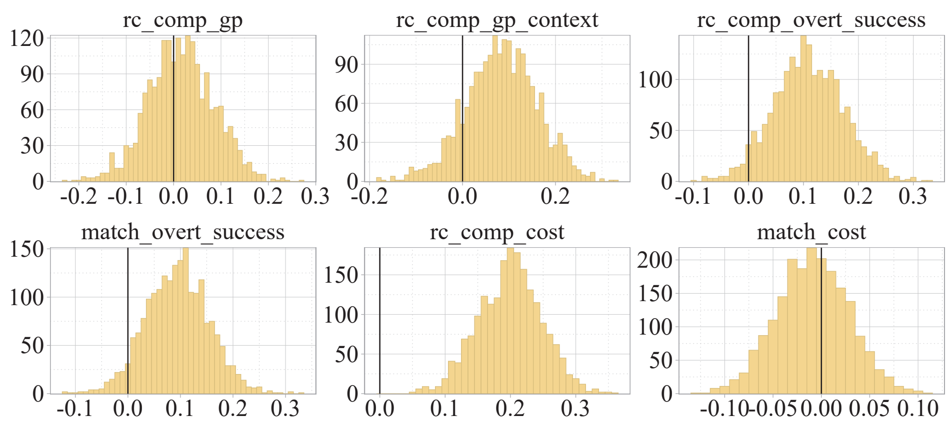


Figure 2. Posterior distributions of selected model parameters.

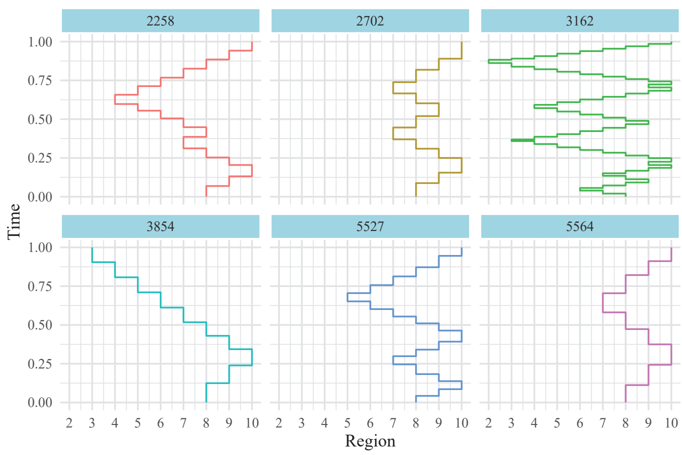


Figure 3. Example scanpaths starting at the disambiguating region in disambiguation/context mismatch conditions.

References. [1] Staub et al. (2018, J Mem Lang). [2] Altmann et al. (1992, J Mem Lang). [3] Paape & Vasishth (2022, Cogn Sci). [4] Traxler et al. (1998, J Mem Lang). [5] Meng & Bader (2000, Lang Speech). [6] Fodor & Inoue (2000, Lang Speech). [7] Frazier & Rayner (1982, Cogn Psychol). [8] Paape & Vasishth (2022, Glossa Psycholing) [9] Roland (2024, HSP poster).

OS.4.07

Theta-gamma coupling as an EEG signal for event processing in language

Wesley Leong (wesley.leong@uconn.edu), Gerry TM Altmann
University of Connecticut

Introduction

What do we represent in our minds when we process events in language? Altmann and Ekves (2019) proposed that events comprise objects and their histories (i.e. in the sentence “the chef chopped the mango”, we simultaneously represent both the intact and chopped mango). There is fMRI evidence of competition between these object-states in the left ventrolateral prefrontal cortex (Hindy et al. 2012; Solomon et al. 2015). However, little is known about the dynamics of these representations and how they are maintained simultaneously in the brain. Here we propose a neural code for event representation: theta-gamma phase-amplitude coupling (henceforth PAC). PAC occurs when neural oscillations of two different frequencies are coupled together. That is, the amplitude of a fast gamma wave waxes and wanes at the frequency of a slower theta wave. In temporal regions of the brain, the signal has typically been associated with working memory – for example, the strength of theta-gamma coupling increases with working memory demands (e.g. Heusser et al 2016, Reinhart and Nguyen 2019). What the individual oscillations represent is less clear, but one hypothesis is that PAC functions to sequentially encode multiple items in working memory (Lisman 2005). PAC may thus be a neural signal of events in one of two ways. On one hand, we might observe greater PAC when objects undergo substantial change, reflecting the encoding of multiple distinct object-states sequentially in working memory. Alternatively, PAC might increase in events which incur minimal change instead, representing the increased working memory demand of distinguishing two featurally similar object-states. This abstract presents PAC analyses of data from two existing experiments on events in language. We hypothesized that PAC would increase in the substantial change condition (e.g. “the chef chopped the mango”; corresponding to more distinct object-states) but not the minimal change condition (e.g. “the chef weighed the mango”).

Methods

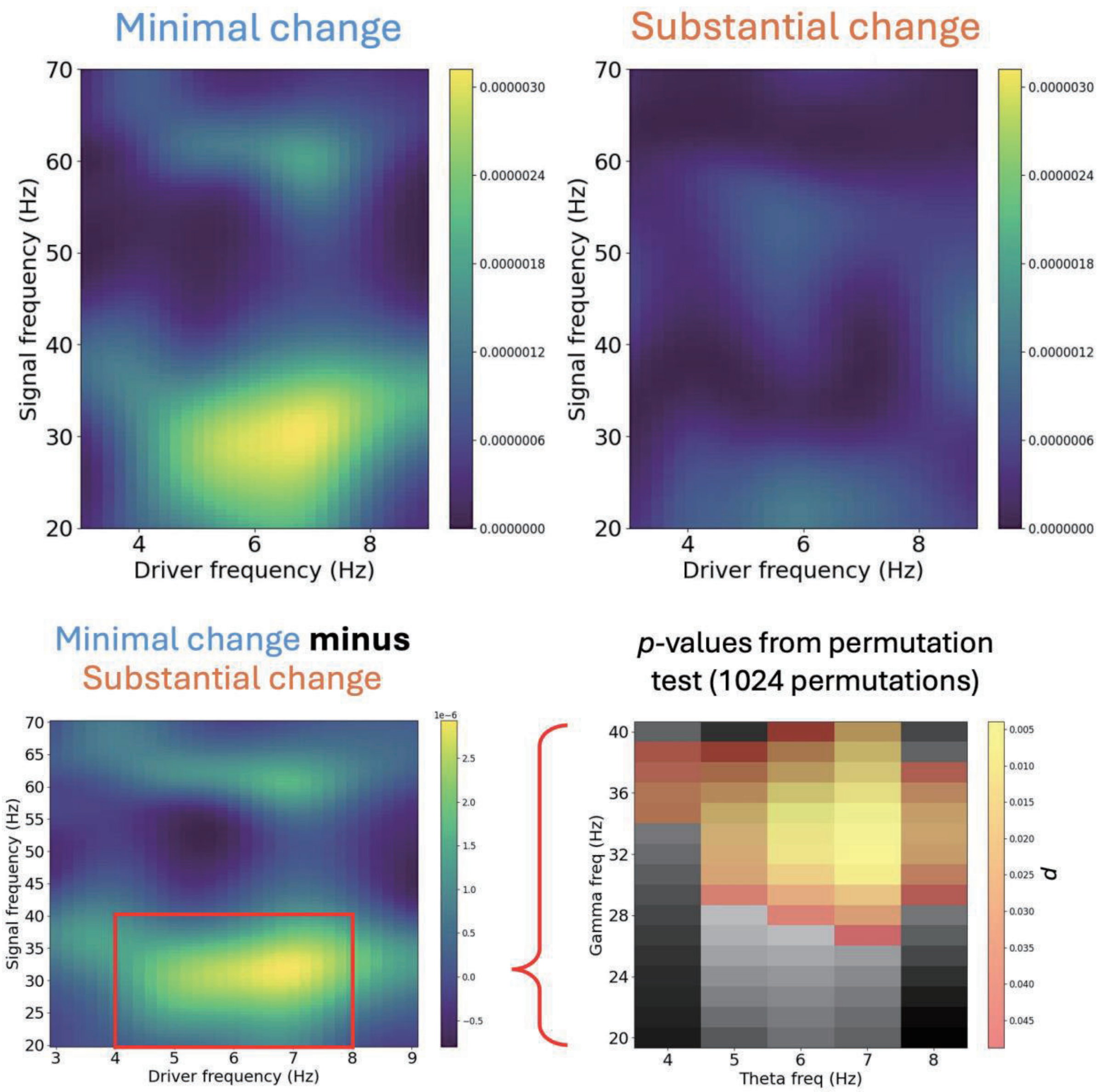
In both experiments (1: N=45; 2: N=31), participants read English sentences on a screen word-by-word (ISI: 600ms) while having their EEG data recorded. Stimuli consisted of 2 sentences each – in the first sentence, an object underwent either substantial or minimal state change, and in the second sentence, it was referenced again (e.g. “The woman will bite the plum. And then, she will squeeze the plum.”). EEG was recorded using 256-channel EGI HydroCel Geodesic Sensor Nets at a sampling rate of 1000 Hz. Raw data was bandpass filtered at 1-80 Hz then downsampled to 250 Hz. 7.2-second (12-word) epochs were extracted from a left temporal electrode (T7) following the onset of each trial. Data from both experiments was combined, and PAC at this channel was then measured using a Driven Auto-Regressive (DAR) model (Dupré la Tour et al 2017). The difference in PAC between conditions was quantified using a cluster-based permutation analysis (Maris and Oostenveld 2007).

Results

There was a significant difference in theta-gamma phase-amplitude coupling between the minimal change and substantial change conditions ($p < 0.05$). Minimal state-change sentences had, contrary to expectations, greater phase-amplitude coupling. The effect was visually replicated in a cluster of electrodes surrounding the electrode of interest, and was robust to various bootstrapping techniques.

Summary

Theta-gamma PAC is a promising correlate of event processing in language in the brain. We propose that events which incur minimal change also require finer representational distinctions to maintain separable versions of the same object across the event. These distinctions, reflecting the trajectory of the object through time, are diagnostic of the event itself. Thus maintaining diagnosticity in the face of greater representational overlap imposes additional working memory demands, which manifests as increases in PAC.



References

Altmann, G. T. M., & Ekves, Z. (2019). Events as intersecting object histories: A new theory of event representation. *Psychological Review*, *126*(6), 817–840.

Dupré la Tour, T., Tallot, L., Grabot, L., Doyère, V., van Wassenhove, V., Grenier, Y., & Gramfort, A. (2017). Non-linear auto-regressive models for cross-frequency coupling in neural time series. *PLoS Computational Biology*, *13*(12), e1005893.

Heusser, A. C., Poeppel, D., Ezzyat, Y., & Davachi, L. (2016). Episodic sequence memory is supported by a theta–gamma phase code. *Nature Neuroscience*, *19*(10), Article 10.

Hindy, N. C., Altmann, G. T. M., Kalenik, E., & Thompson-Schill, S. L. (2012). The Effect of Object State-Changes on Event Processing: Do Objects Compete with Themselves? *Journal of Neuroscience*, *32*(17), 5795–5803.

Lisman, J. (2005). The theta/gamma discrete phase code occuring during the hippocampal phase precession may be a more general brain coding scheme. *Hippocampus*, *15*(7), 913–922.

Reinhart, R. M. G., & Nguyen, J. A. (2019). Working memory revived in older adults by synchronizing rhythmic brain circuits. *Nature Neuroscience*, *22*(5), 820–827.

Solomon, S. H., Hindy, N. C., Altmann, G. T. M., & Thompson-Schill, S. L. (2015). Competition between Mutually Exclusive Object States in Event Comprehension. *Journal of Cognitive Neuroscience*, *27*(12), 2324–2338.

OS.4.08

Comprehending semantic and syntactic anomalies from an LLM versus human interlocutor: An ERP study

Xiaohui Rao(1155197282@link.cuhk.edu.hk), Hanlin Wu, Zhenguang Cai
The Chinese University of Hong Kong

As large language models (LLMs) increasingly engage in human-like conversations, understanding how people comprehend language produced by LLMs compared to humans is crucial. When processing sentences, individuals evaluate input against their world knowledge, with semantic anomalies eliciting an N400 effect and syntactic anomalies eliciting a P600 effect. These effects can be influenced by expectations about the interlocutor's identity (e.g., Hanulíková & Carreiras, 2015; Van Berkum et al., 2008). However, it remains unclear if expectations differ when interacting with LLMs versus humans. LLMs demonstrate strong grammatical competence but can sometimes produce nonsensical responses (Rawte et al., 2023). If so, people may show reduced sensitivity to semantic anomalies and increased sensitivity to syntactic anomalies when interacting with LLMs. To investigate this, we conducted two ERP experiments: Exp.1 focused on semantic anomalies, while Exp. 2 examined syntactic anomalies. Participants were informed that the sentences were produced by either an LLM or a human interlocutor, although the materials remained identical across both conditions.

Experiment 1 (N=64) utilized 40 semantically coherent/anomalous Chinese sentences and 80 fillers, presented at a fixed rate, followed by a sentence sensibility judgment task. Linear mixed-effects models revealed a significant main effect of Sentence Type at the critical word ($p < .001$), with an interaction between Sentence Type and Interlocutor Type ($p < .05$). Participants showed a reduced N400 effect when interacting with LLMs compared to humans, indicating lower sensitivity to semantic anomalies from LLMs, potentially due to expectations of LLMs' higher chance of illusion. Experiment 2 (N=64) focused on syntactic anomalies and revealed a main effect of Sentence Type ($p < .05$) and an interaction with Interlocutor Type ($p < .05$), where a P600 effect was observed only for LLM-produced syntactic anomalies. This suggests greater sensitivity to syntactic deviations from LLMs compared to humans, possibly because people don't expect LLMs to produce syntactically deviant sentences and are more tolerant to humans in this regard. Our findings provide the first piece of neural evidence for how people process language from LLMs versus human interlocutors.

References

Hanulíková, A., & Carreiras, M. (2015). Electrophysiology of subject-verb agreement mediated by speakers' gender. *Frontiers in Psychology, 6*(September), 1–12.

Van Berkum, J. J., Van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience, 20*(4), 580-591.

Rawte, V., Sheth, A., & Das, A. (2023). *A survey of hallucination in large foundation models*. arXiv preprint arXiv:2309.05922.

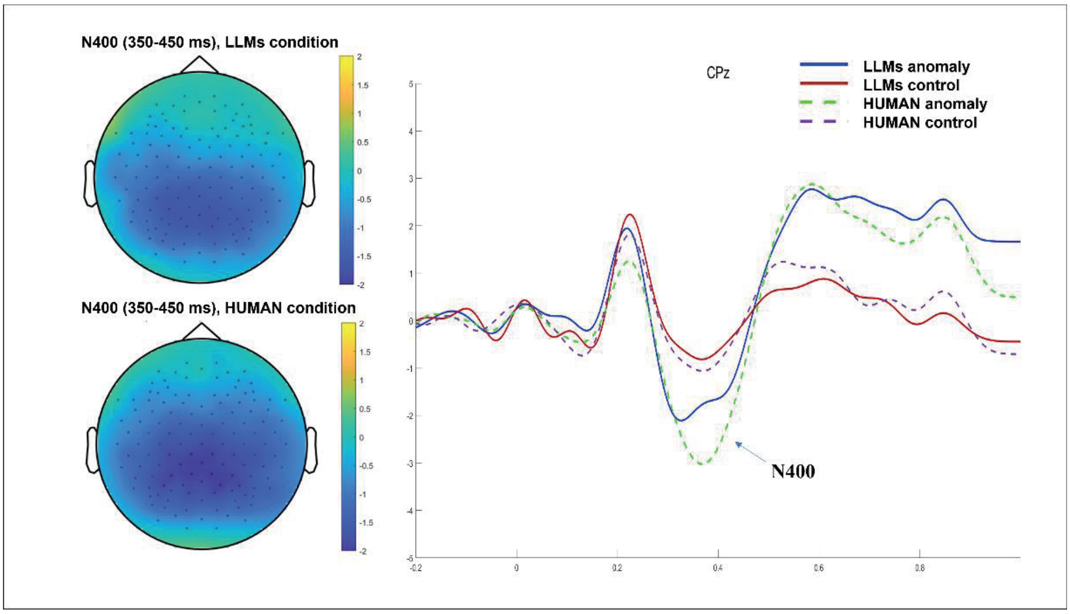


Figure 1. Topographic distributions of difference waves during 350-450 ms after the critical segment onset (amplitudes of anomalous minus control conditions); grand average waveforms for the anomalous and control conditions during 350-450 ms.

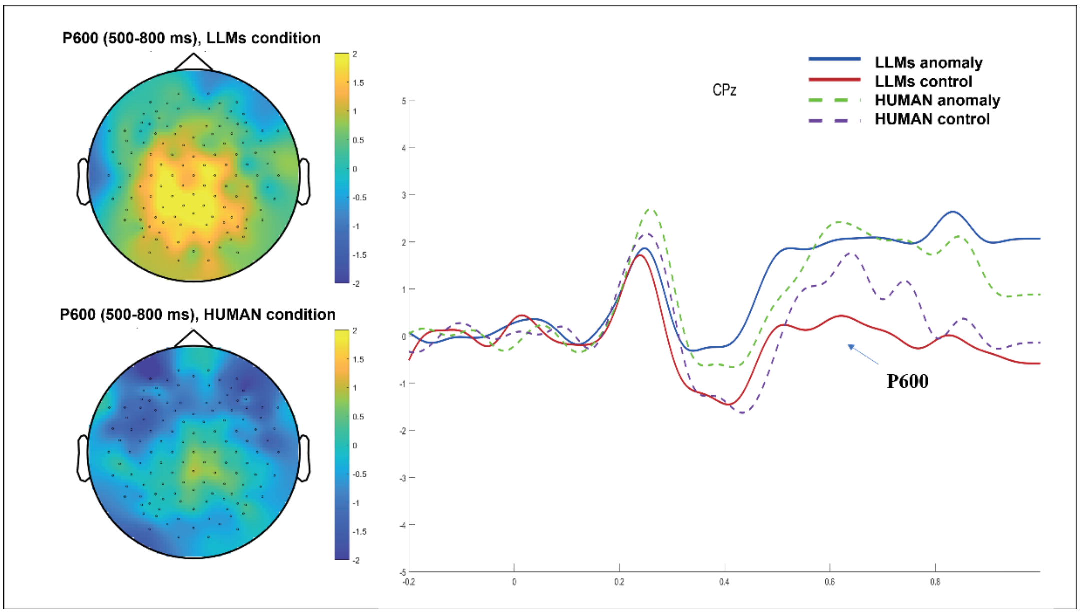


Figure 2. Topographic distributions of difference waves during 500-800 ms after the critical segment onset (amplitudes of anomalous minus control conditions); grand average waveforms for the anomalous and control conditions during 500-800 ms.

OS.4.09

Age-related differences in agreement processing: insights from an eye-tracking study

Victoria Cano-Sánchez[1], Itziar Laka[1], Jana Reifegerste[2], & Mikel Santesteban[1]
[1]University of the Basque Country (UPV/EHU), [2]Georgetown University
victoria.cano@ehu.eus

Introduction. Language processing in aging is understudied, especially regarding grammar. [1]. Prior research indicates that OAs take longer to process ungramm. sentences and make more agreement attraction errors (e.g., **The key to the cabinets were rusty)* than YAs [2, 3, 4]. We wanted to replicate this aging effect on agreement violations (aim 1) and to further explore OAs' and YAs' susceptibility to agreement attraction during sentence comprehension in Spanish (aim 2). We expected to replicate the larger grammaticality effects for OAs (prediction 1) [4]. Importantly, if these aging effects reflected weaker resources to process agreement, we expected OAs to show stronger agreement attraction effects than YAs (prediction 2) [2]. **Experiment.** We recorded the eye-movement patterns of 44 YAs (18-34 years; M=28.5, SD= 7.7) and 48 OAs (>65 years; M=64.0 SD= 5.6) healthy native Spanish speakers while they read sentences with varying grammatical and attractor conditions. We created 60 sentences in 4 experimental conditions (through an online likert scale norming study): with subject noun phrases whose head matched (pl.) or mismatched (sg.) in number with the pl. verb (Grammaticality: gramm./ungramm.) and attractor nouns that matched (sg.) or mismatched (pl.) the verb (Attractor: sg./pl.) (Figure 1). (G)LME analyses at the critical verb region revealed main effects of group, grammaticality and attractor showing larger fixations and more regressions for OAs vs. YAs, in ungramm. than gramm. sentences, and in sing. vs. pl. attractors. An interaction of grammaticality by group appeared at *Regression Path Durations* showing larger grammaticality effects (longer regression-path durations in ungramm. than gramm. sentences) for OAs vs. YAs (confirming prediction 1: see Figure 2). Another grammaticality by attractor interaction appeared at *Regressions-in* and *Total Time Durations,* showing that attraction effects were only significant in ungramm. sentences and unveiling a grammatical asymmetry of attraction effects [5]. Importantly, a 3-way interaction at *Total Time Durations* showed that this grammatical asymmetry of attraction effects was only significant for OAs at this measure (confirming prediction 2: see Figure 3) [2]. **Discussion & conclusion.** Group differences in *Total Time* suggest that OAs may reanalyze more often than YAs, particularly in ungrammatical-mismatch (SSP) sentences. This could reflect better attempts to resolve ungrammaticalities (confirming prediction 1). However, in the ungrammatical-match (SPP) sentences, OAs fall down more into the grammaticality illusion created by plural attractors (confirming prediction 2). Altogether, our findings indicate that age modulates agreement processing increasing agreement errors for the OAs.

References

[1] Harada et al., *Clin Geriatr Med,* 2013.

[2] Reifegerste, J., et. al., *Aging Neuropsuchol. Cogn.,* 2017.

[3] Reifegerste, J. et. al., *J Mem Lang.,* 2020.

[4] Authors, *AmLaP,* 2021.

[5] Wagers et al., *J. Mem Lang.,* 2009.

Figure 1. *Sample sentences of an item in the four experimental conditions resulting from the manipulation of the attractor number (singular vs. plural) and grammaticality (grammatical vs. ungrammatical sentences). The two elements involved in the attraction relation are bolded across all conditions, and the critical region for analysis (the verb) is framed in squares.*

(i) Grammatical-Mismatch (Plural subject-Singular attractor-Plural verb; PSP):

***Los pasteles**$_{PL}$ con el dibujo$_{SG}$ nunca **ganaron**$_{PL}$ un premio de repostería debido a la calidad de la masa.*

(ii) Grammatical-Match (PPP):

***Los pasteles**$_{PL}$ con los dibujos$_{PL}$ nunca **ganaron**$_{PL}$ un premio de repostería debido a la calidad de la masa.*

(iii) Ungrammatical-Mismatch (SSP):

****El pastel**$_{SG}$ con el dibujo$_{SG}$ nunca **ganaron**$_{PL}$ un premio de repostería debido a la calidad de la masa.*

(iv) Ungrammatical-Match (SPP):

****El pastel**$_{SG}$ con los dibujos$_{PL}$ nunca **ganaron**$_{PL}$ un premio de repostería debido a la calidad de la masa.*

The cake(s)$_{MASC.SING./PL}$ with the design(s)$_{MASC.\ SING./PL}$. never won$_{V°PL}$. a prize bakery due to quality dough.

Gloss: The cake with the designs never won a bakery prize due to its dough quality.

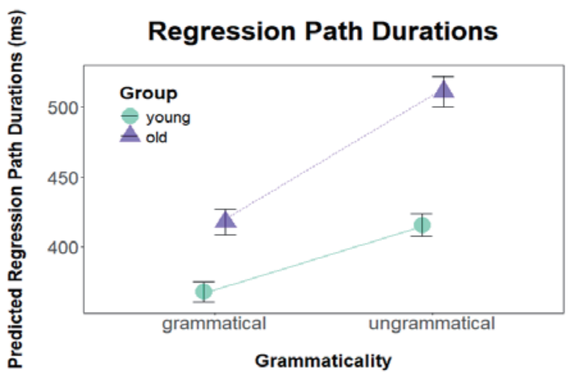


Figure 2. *Graph plotting the two-way interaction of grammaticality by group in Regression Path Duration at the verb region with larger grammaticality effects in the ungrammatical than grammatical sentences for the OAs compared to YAs.*

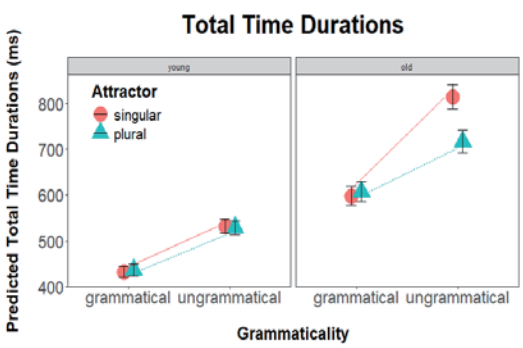


Figure 3. *Graph plotting the triple way interaction of grammaticality by attractor by group in Total Time Durations at the verb region with a significant effect of attraction only for the older adults in the ungrammatical-mismatch (SPP) condition.*

OS.4.10

Constructing dependencies with optional elements: Insights from Vietnamese

Linh Pham & Elsi Kaiser - University of Southern California – llpham@usc.edu

Introduction. The mechanisms underlying linguistic dependency formation, including its susceptibility to similarity-based interference, are still debated (e.g. cue-based retrieval [6], [7] vs. representation distortion [3, feature percolation [1], lossy compression [4]). However, prior work tested dependencies with two required elements, e.g. subject-verb agreement [2] and negative polarity [8], where encountering the first element allows comprehenders to reliably predict/expect the second element. This might increase the first element's activation (under cue-based retrieval accounts) or influence the degree of distortion (under representation distortion accounts). But what happens when the second element in a dependency is *optional*?

The current study examines similarity-based interference effect (e.g. [1,6]) in a new kind of dependency between (a) an overt Q(uestion)-particle *thế* in Vietnamese, which is optional and placed at the end of the sentence (see ex.1), and (b) wh-phrases in its domain.

Predictions. The cue-based retrieval account (e.g. [5]) predicts that the presence of an intervening existential *ai đó* (lit. who-that) 'someone' (Table 1) – an element that contains a wh-phrase morpheme but has no interrogative interpretation – might interfere with the retrieval process triggered by the Q-particle. Specifically, if the wh-component of this element is 'visible' to the retrieval process, then it should cause interference in grammatical conditions but facilitation in ungrammatical conditions (Table 1). In contrast, the distortion-based account predicts facilitation in both grammatical and ungrammatical condition.

Design. Our study manipulated two factors: (i) grammaticality (grammatical/ungrammatical) and (ii) attractor type ('wh-that' interference or not) which yields 4 conditions (Table 1)

Exp 1. (Self-paced reading, n=92). We used lmer to analyze log-transformed RTs. Word-by-word RTs (Fig.1) show a slowdown in the [ungrammatical-mismatch] condition at spillover3. Here, we found a main effect of grammaticality (p=0.05) and a grammaticality x attractor type interaction (p=0.039): We observe a facilitatory effect *only in ungrammatical conditions*.

Exp 2. (Acceptability judgment task, n=48 new participants). We also used lmer to analyze acceptability judgments (Fig.2). We found a main effect of grammaticality (p<.0001) and a grammaticality x attractor type interaction (p=0.0003). Planned comparisons show that presence of the 'who-that' intervener increases acceptability of ungrammatical sentences (p=0.001) *and* decreases acceptability of grammatical sentences (p=0.04).

Conclusion. Our findings align with a cue-based retrieval account. Crucially, the interference effects are stronger in offline data (Exp.2), contrary to patterns in prior work. This may be due to the Q-particle's optionality, suggesting that the time-course of dependency-formation crucially depends on whether comprehenders can reliably anticipate/predict the second element. Thus, the grammatical properties of Vietnamese allow us to gain new insights into properties of dependency-formation during processing.

Table 1. Sample item illustrating the 4 conditions.

Condition	Sentence
[grammatical match]	"Linh biết **cái gì** đã được ai đó chuyển ra khỏi căn phòng **thế**?" "Linh know **what** PST PASS who that move of the room **Q**?" 'What was the thing moved out of the room by someone that An knew?',
[grammatical mismatch]	"Linh biết **cái gì** đã được tổ trưởng chuyển ra khỏi căn phòng **thế**?" "Linh know **what** PST PASS manager move out of the room **Q**?" 'What was the thing moved out of the room by the manager that An knew?',
[ungrammatical match]	"Linh biết cái bàn đã được ai đó chuyển ra khỏi căn phòng **thế**?" "Linh know table PST PASS who that move out of the room **Q**?" *'An knew the table was moved out of the room by someone?',
[ungrammatical mismatch]	"Linh biết cái bàn đã được tổ trưởng chuyển ra khỏi căn phòng **thế**?" "Linh know table PST PASS manager move out of the room **Q**?" *'An knew the table was moved out of the room by the manager?'
Spill-over	….., whispered the boy.

Figure 1. Exp1. Average reading times per word

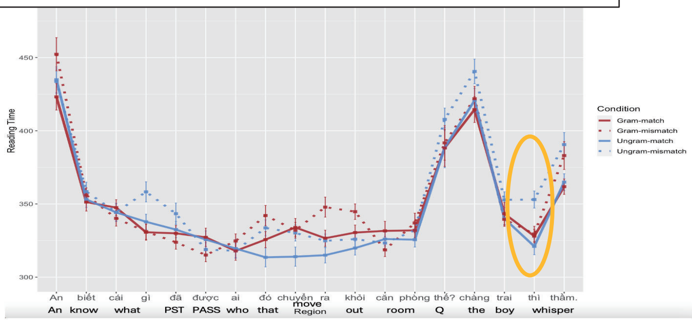
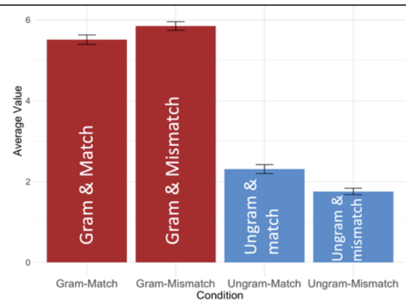


Figure 1. Exp2. Mean ratings by condition (1= unacceptable, 7=acceptable)



[1] Bock & Eberhard 1993. Meaning, sound and syntax in English number agreement. *LCP* || [2] Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora:Experimental and modeling evidence. *JML*.|| [3] Eberhard, K. M. (1997). The marked effect of number on subject–verb agreement.*JML*.|| [4] Futrell, R., Gibson, E.,&Levy, R.P.(2020): Lossy context surprisal:An information-theoretic model of memory effects in sentence processing.*CogSci*.|| [5] Gordon et al 2001. Memory interference during language processing. *JEP*|| [6] Lewis et al 2005. An activation-based model of sentence processing as skilled memory retrieval.*CogSci*.|| [7] Lewis, R.L., Vasishth, S., &Van Dyke, J.A.(2006). Computation principles of working memory in sentence comprehension.TCS.|| [8] Parker,D.,&Phillips,C.(2016). Negativity polarity illusions and the format of hierachical encodings in memory. *Cognition*.

Example 1. Q-particle is optional and forces an interrogative interpretation.

(1) a) *Wh-word without Q particle: Ambiguous*

Lan	biết	ai	vừa	ra	ngoài
Lan	know	who	PRF	go	out

"Lan knew who had gone out."/ "Who has gone out that Lan knew?"

b) *Wh-word with Q particle: Unambiguously a question*

Lan	biết	ai	vừa	ra	ngoài	thế?
Lan	know	who	PRF	go	out	Q?

"Who has gone out that Lan knew?"

OS.4.11

Non-plural 'some' in contexts: Mouse-tracking evidence for rapid real-time social reasoning

Wei Li[1], Hannah Rohde[2], Martin Corley[1]
[1] *Psychology, PPLS, University of Edinburgh, UK*
[2] *LEL, PPLS, University of Edinburgh, UK*
Contact Author: Wei Li (s1833067@ed.ac.uk)

Listeners' interpretations of the scalar quantifier 'some' vary from the semantic (some and possibly all) to a pragmatically-reinforced (some but not all) meaning, depending on context. Loy et al. (2019) showed that listeners are more likely to make an early commitment to the semantically-allowed meaning of 'some' as 'all' if it follows disfluent 'uh' in a context where larger values are socially undesirable (I ate, uh, some oreos).

Here, we varied the context to one where smaller values are socially undesirable interpretations of 'some'. In two mouse-tracking experiments, we recorded mouse movements from 150 (experiment 1) and 173 (experiment 2) participants in a web-based task, in which we manipulated disfluency (present vs. absent) within-subjects in a set of 12 target trials. In each target trial, participants saw four images with different numbers of qualifications, each representing one of four potential interpretations of the meaning of 'some' (Fig. 1), and heard an interviewer ask an interviewee about their qualifications (Example 1). We measured both the final click results (i.e., which image each participant clicked at the end) as well as the trajectories of participants' mouse movement during each trial.

Here, disfluency has the opposite effect, reducing the value participants associate with 'some': We found that participants are more likely to select images corresponding to one, or zero, qualifications, following disfluent utterances (Fig. 2). However, their mouse movements (Fig. 3) show they are quick to commit to one qualification (experiment 1) and slow to commit to zero (experiment 2).

Analyses of mouse trajectories further confirm these differences. For each experiment, we calculated the perpendicular distance over time in each trial from the mouse position to a diagonal line running from the two- to the four-check image, passing through screen centre. The perpendicular distance from this line depicts the extent to which a participant moves away from common, plural, understandings of 'some'. We analysed this distance over 101 time-normalised timesteps to map listeners' *non-plural tendencies* when hearing disfluent/fluent utterances, independently for each experiment. For participants who clicked the one-check image, their mouse-movements diverged around 355ms after disfluency onset. The pattern for the zero-check experiment was different: Despite the difference in eventual responses, a difference in mouse-movements only emerged some 2013ms after disfluency onset.

These findings suggest that social context and manner of speech can combine to affect the interpretation of 'some' as an utterance unfolds. Extending its meaning to 'one' is relatively easy, but overwriting it with 'zero' (in effect, deciding that a speaker is lying) is more demanding.

Example 1 Interviewer: "How many 'A's have you got for your psychology courses?"
Interviewee: "I've got **some** 'A's." (fluent) *or*
"I've got, **uh, some** 'A's." (disfluent)

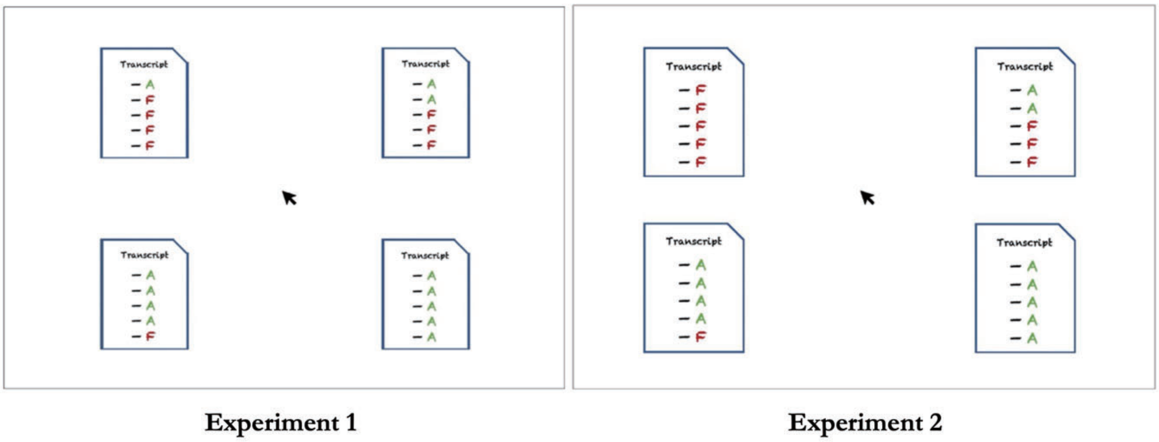


Fig. 1 Examples of screen displays in Experiment 1 (left) and Experiment 2 (right)

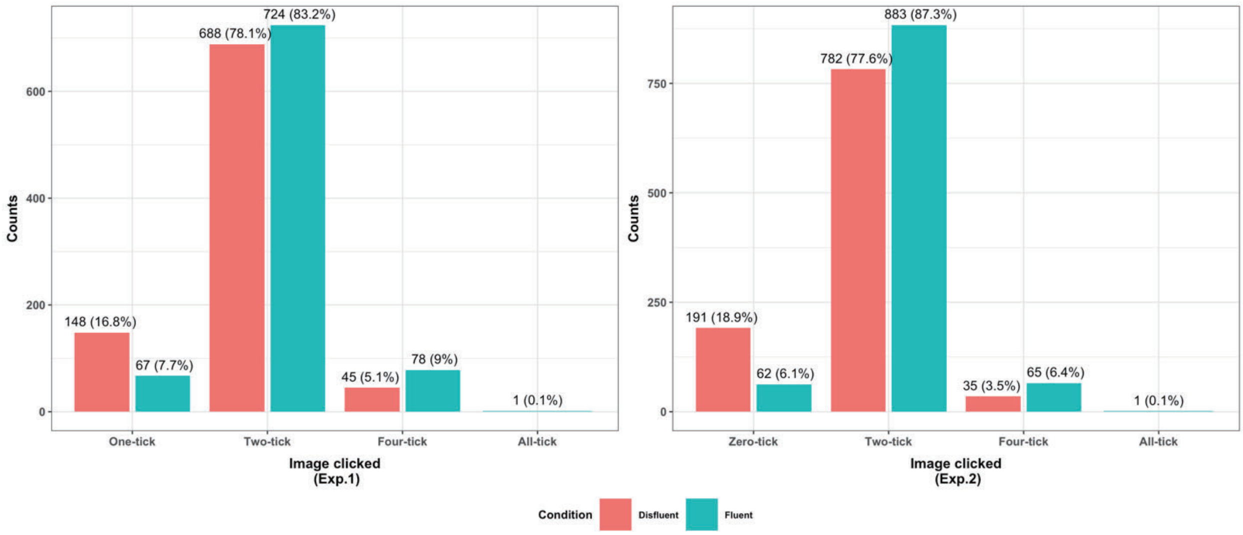


Fig. 2 Total number and distribution of mouse clicks recorded on each image (one/zero, two, four, or all) by manner of delivery (disfluent/fluent) in Exp. 1 (left) and Exp. 2 (right).

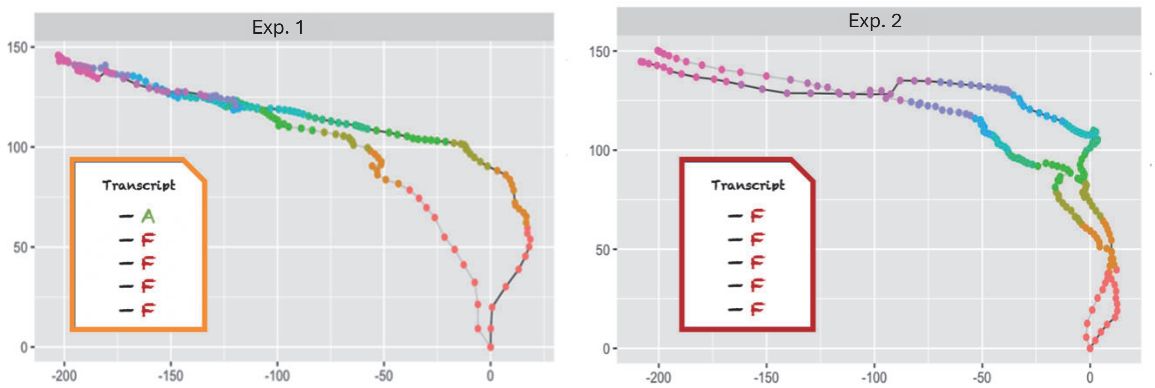


Fig. 3 Aggregated mouse trajectories towards image representing one (left) and zero (right) by condition (disfluent/fluent); the colours of points (from red to violet) indicate 10%, 20%, 30%...100% of trial time.

Reference Loy, J. E., Rohde, H., & Corley, M. (2019). Real-time social reasoning: The effect of disfluency on the meaning of some. *Journal of Cultural Cognitive Science*, *3*(2), 159-173.

Poster Presentations

PS.1.01

Spoken Word Recognition of Malay-Derived Singlish Concepts Among Indonesians Living in Singapore

Andrea Mae[1] & Cynthia S. Q. Siew[1]* (email: cynthia@nus.edu.sg)

[1]Department of Psychology, National University of Singapore

This study investigated the potential impact of L1 Indonesian knowledge on the processing of L2 Singapore Colloquial English (SCE; also referred to as "Singlish") among Indonesians who are currently living in Singapore. Because Singlish emerged as a contact language from prolonged interactions between speakers of English, Malay, Hokkien, Tamil, and other languages in early Singapore (Leimgruber, 2011), the SCE lexicon consists of lexical items that originated from these languages. As Indonesian was adapted from Bazaar Malay, Malay and Indonesian share many lexical similarities (Simpson, 2021). Language transfer theories (Faerch & Kasper, 1987) predict that L1 Indonesian speakers should be able to make use of their L1 to facilitate acquisition and processing of their L2 (Singapore English)—although this facilitation is also hypothesized to be limited to Singlish concepts of Malay origin.

41 Indonesians and 41 Singaporeans whose native languages were Indonesian and Singapore English respectively took part in this study. Indonesian participants were currently living in Singapore and have done so for at least 1 to 20 years. Participants completed the Language Experience and Proficiency Questionnaire (LEAP-Q) and the auditory lexical decision task (ALDT). Participants were presented with 100 Singlish words and 100 nonwords auditorily and indicated whether they heard a word or nonword by pressing the corresponding button. Of the 100 Singlish words, 50 originated from Malay, and the other 50 were of non-Malay origin (i.e., English, Hokkien, Tamil). Examples of the stimuli are provided in Table 1. Reaction Time (RT) and Accuracy data from the ALDT were analyzed using linear and generalized mixed-effects models respectively. The models included fixed effects of Word Origin and Participant Group and its interaction, lexical covariates, random intercept effects of Participant and Item, and random slopes where appropriate. For both models, the interaction effect between Word Origin and Participant Group was statistically significant and further probed with planned contrasts. Interaction effects are depicted in Figure 1.

The interaction effect indicated enhanced ALDT performance for Malay Singlish concepts than non-Malay concepts among Indonesians, suggesting that their L1 knowledge of Indonesian facilitated processing of SCE, but this may be limited to the part of the Singlish lexicon with Malay origin words. In contrast, there was no difference for Malay and non-Malay Singlish concepts among Singaporean participants. Across participant groups, Singaporeans were faster than Indonesians, and more accurate for non-Malay Singlish words, although both groups were similar in their accuracy for recognizing Malay Singlish words. Taken together, the results provide insights into the nature of L2 acquisition and processing of SCE among immigrants to Singapore, with potential implications for language integration among immigrants—the contribution of diverse languages to the Singlish vocabulary may provide multiple "points of entry" for diverse L1 speakers to leverage upon as they begin to learn the host country's language.

References

Faerch, C., & Kasper, G. (1987). Perspectives on Language Transfer. *Applied Linguistics, 8*(2), 111–136.

Leimgruber, J. R. E. (2011). Singapore English. *Language and Linguistics Compass*, *5*(1), 47–62. https://doi.org/10.1111/j.1749-818X.2010.00262.x

Simpson, A. (2021). Language and the building of nations in Southeast Asia. In *The Languages and Linguistics of Mainland Southeast Asia* (pp. 927–956). De Gruyter. http://dx.doi.org/10.1515/9783110558142-038

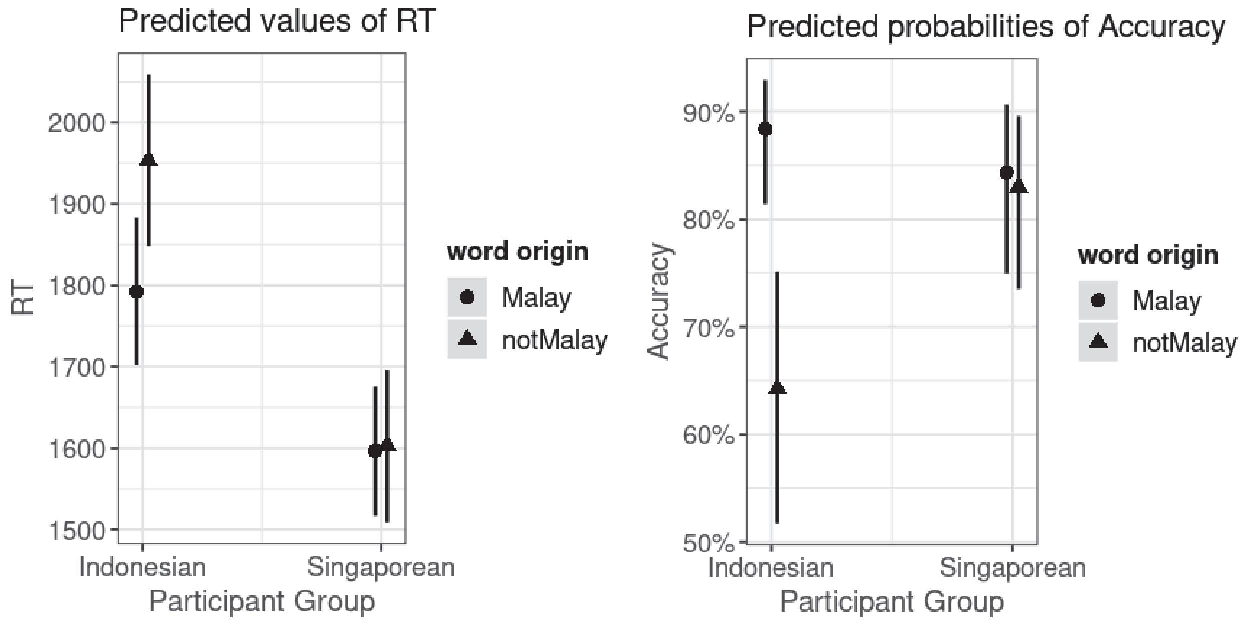


Figure 1. Interaction effect of Word Origin and Participant Group. Left: RT model. Right: Accuracy model.

Malay-Derived Singlish Word	Corresponding Nonword	Non-Malay-Derived Singlish Word	Corresponding Nonword
atas	atus	confirm [English]	confarm
bodoh	bodoy	dabao [Cantonese]	dabiu
terbalik	cerbalik	goondu [Tamil]	goonvu
gila	giha	humji [Hokkien]	humki

Table 1. Examples of word and nonword stimuli used in the ALDT. Nonwords were created by substituting one phoneme or syllable to create word forms that did not exist in Singapore English. We note that some Singlish Malay concepts in our study are also Indonesian words; however, there are meaning and usage differences of those words in SCE that Indonesian participants are in fact sensitive to (as assessed in an offline vocabulary test not reported here). This provides some supporting evidence that Indonesian participants are not simply better at recognizing words in their L1 on the ALDT, since they have also acquired the “Singlish” meanings of those words.

Cross-language co-activation during concurrent comprehension and production: Evidence from simultaneous interpreters

Xueni Zhang[1] and Binghan Zheng[1]
[1]School of Modern Languages and Cultures, Durham University, Durham, UK
(binghan.zheng@durham.ac.uk)

It is widely evidenced that both languages are simultaneously active in bilinguals when only one language is being used (Chen et al., 2017; Lagrou et al., 2011; Thierry & Wu, 2007). However, little is known about how languages are activated during simultaneous interpreting, a demanding task involving explicit use of both languages and concurrent operation of comprehension and production (Christoffels & de Groot, 2004). This study examined the time-course of cross-language co-activation triggered by similarity in form and meaning. Twenty-seven professional interpreters (Chinese as L1 and English as L2; age: *mean* = 31.33, *SD* = 5.88; interpreting experience: *mean* = 7.11, *SD* = 4.02) were recruited to complete a cross-language task (English-to-Chinese simultaneous interpreting) and a within-language task (English-to-English shadowing) while their eye movements were tracked. Participants heard English passages containing critical spoken words while being visually presented a display of printed Chinese words. The displays contained three unrelated distractors and a competitor which resembled the translation equivalent of the spoken word in either form (e.g., 服务 - 服装, *service - clothes*) or meaning (e.g., 经济 - 金融, *economy – finance*). Fixation proportions were calculated to index visual bias towards competitors. We found that participants directed more visual attention on word-form competitors compared to distractors immediately after spoken word onset in a transient manner, and the effect reoccurred during/prior to production. Activation of semantic information was long-lasting and large in magnitude. Stronger co-activation of meaning was found for shadowing than simultaneous interpreting, while there was no task difference in the word-form competitor condition. Our findings suggest that the time lag between input and output modulated the time-course of co-activation, providing implications for the serial versus parallel account of interpreting processing.

Keywords: Bilingual processing, simultaneous interpreting, cross-language co-activation, visual word paradigm

References

Chen, P., Bobb, S. C., Hoshino, N., & Marian, V. (2017). Neural signatures of language co-activation and control in bilingual spoken word comprehension. *Brain Research*, *1665*, 50–64. https://doi.org/10.1016/j.brainres.2017.03.023

Christoffels, I. K., & de Groot, A. M. B. (2004). Components of simultaneous interpreting: Comparing interpreting with shadowing and paraphrasing. *Bilingualism: Language and Cognition*, *7*(3), 227–240. https://doi.org/10.1017/S1366728904001609

Lagrou, E., Hartsuiker, R. J., & Duyck, W. (2011). Knowledge of a second language influences auditory word recognition in the native language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(4), 952–965. https://doi.org/10.1037/a0023217

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences*, *104*(30), 12530–12535. https://doi.org/10.1073/pnas.0609927104

Typed Tongue-Twisters as Evidence for Interactivity in Production

Martin Corley, Psychology, University of Edinburgh

A continuing debate in language production concerns the origins of the well-attested biases in segmental speech errors. According to the self-monitoring view (Levelt, 1989) activation feeds forward during production and obvious errors, such as nonwords or dissimilar phonemes, are edited out via feedback from the comprehension system. According to the interactive view (Dell, 1986), activation flows back as well as forward in the production system, increasing the likelihood that words which share phonemes with an intended target word will become activated in error, as will phonemes which share features with target phonemes. Importantly, on the interactive but not the self-monitoring view, biases should emerge before the comprehension system is implicated: However, since editing can precede any observable output, determining which is the case is difficult.

Here, we use a typed tongue-twister paradigm in which participants type in four-word ABBA tongue twisters. Repairs of typed errors can be made quickly and without conscious awareness (Pinet & Nozari, 2021), suggesting that they may be analagous to pre-articulatory repairs.

In two online experiments (the second including nonwords), 76 and 95 participants each typed 32 ABBA tongue twisters four times apiece (following two learning trials), from memory. Keyboard effects (proximities of keys, etc.) were controlled for. Analyses of errors using logistic mixed-effects models show robust effects of lexical bias, and of phonemic similarity, in the submitted responses. By recording keypresses, we were able to investigate what was typed prior to any correction; again, there were robust effects of lexical bias and of phonemic similarity. There were no robust patterns in the likelihoods of making overt edits (whether or not those edits led to correct responses).

The evidence from this study makes it harder to sustain the view that an important contributor to segmental error biases is the comprehension system. Lexical and phonemic biases are evident in the first keys participants press, prior to any subsequent editing. Following previous work, we assume that at least some of these edits were not made consciously, and that initial keyings therefore provide a close analogue to the early state of the production plan. The fact that there was a phonemic bias in the errors produced provides additional evidence that typing tongue-twisters is closely related to speaking them aloud. This allows us to add to the growing evidence for production-based interactivity in the early stages of language production, and to suggest that the processes underlying speech production are likely to be a close analogue of those uncovered here.

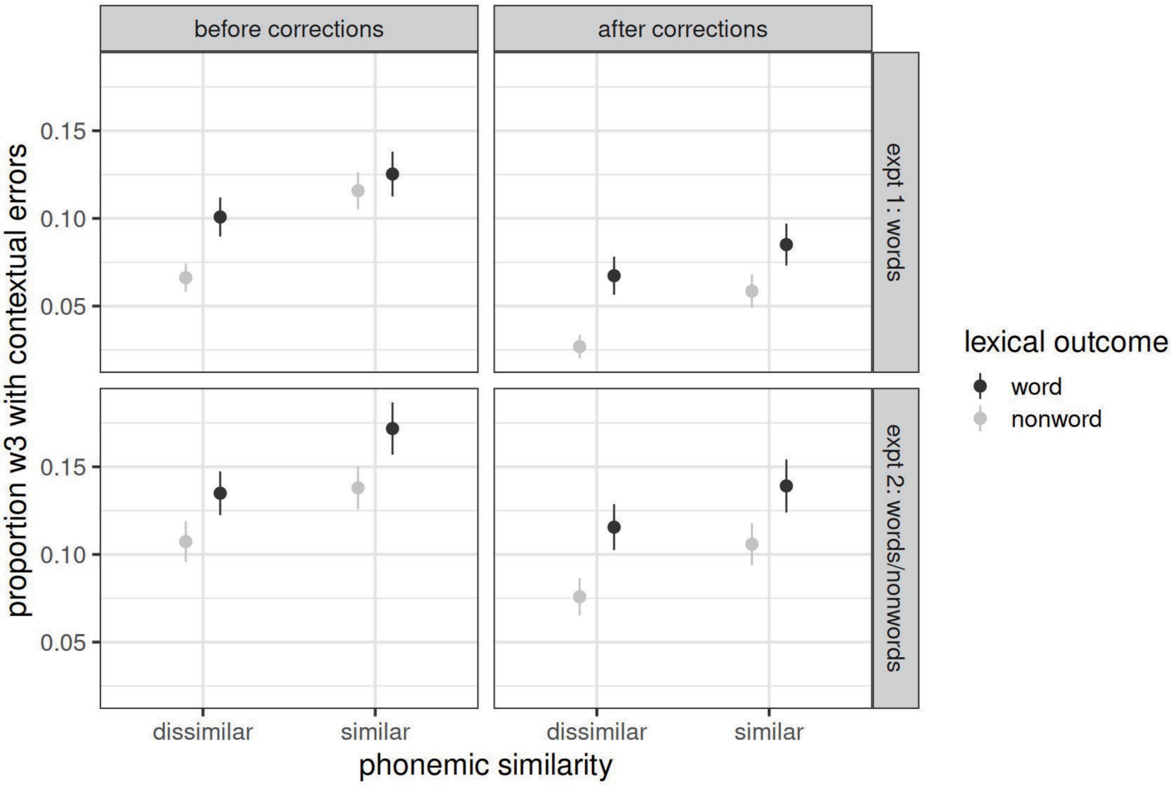


Figure: By-participant mean proportions of responses with contextual errors, where the onset of Word 3 is replaced with the onset of Word 4. Left: before corrections, right: as submitted after corrections; top: experiment 1, bottom: experiment 2. Lines represent one standard error of the mean.

	Similar, Nonword	Similar, Word	Dissimilar, Nonword	Dissimilar, Word
Expt 1	bag dad damp back	bag dad dash back	sag dad damp sack	sag dad dash sack
	dull budge but dove	dull budge buck dove	lull budge but love	lull budge buck love
Expt 2	baz daff damp back	baz daff dash back	saz daff damp sack	saz daff dash sack
	dove gudge gut dutt	dove gudge gull dutt	love gudge gut lutt	love gudge gull lutt

Table: Example materials from each Experiment. NB., phonemic similarity is measured between the onsets of words 1,4 and words 2,3; lexicality of outcome is defined via the substitution of the onset of word 3 with that of word 4 (referred to as a "contextual" error).

PS.1.04 **[Withdrawal]**

Investigating the Mechanism underlying Self-Monitoring in Chinese Word Production

Wang Ziyi & Zhao Chen (Guangdong University of Foreign Studies)

wangyuze77@126.com

Self-monitoring is vital in word production for error detection and correction, yet the level at which it occurs remains unclear. The Perceptual Loop Theory proposes that both inner and overt speech are monitored via the speech comprehension system, comparing formulated utterances with intended ones at the conceptualizer (Levelt, 1989), yet the comparison process is underspecified. Prior research yields conflicting findings on whether monitoring operates primarily at the phonological level in English. This study aims to investigate the level at which self-monitoring occurs in Chinese word production.

In our experiments, the modified stop-signal paradigm was used with Chinese as the target language. Participants named pictures while occasionally halting speech in response to auditory (Experiment 1) or visual (Experiment 2) words that did not match the picture name. Stimuli varied in semantical or phonological similarity/dissimilarity to the intended word of the picture to be named. Stopping accuracy was selected for data analysis.

Results from Experiment 1 showed that phonological similarity significantly influenced halting accuracy, with more errors in phonological similar trials than in other trials, while semantic similarity did not show a significant effect. In contrast, Experiment 2 showed a significant effect of semantic similarity, with more errors in semantically similar trials compared to other trials. Additionally, phonological similarity also had a significant effect, as evidenced by the higher error rate in phonologically similar trials. However, the interaction between phonological and semantic similarity was not significant in either experiment.

Our findings indicate overt speech monitoring and inner speech monitoring showed divergent mechanism, with overt speech monitored by phonological comparation and inner speech monitored by phonological and semantic comparison in Chinese. Moreover, this also implies speakers flexibly monitor their speech at either semantic or phonological level separately based on the availability of information. The evidence could contribute to refining comprehension-based monitoring theories and advancing our understanding of self-monitoring mechanisms.

References:

Levelt, W. J. M. (1989). Speaking: From intention to articulation. Cambridge, MA: MIT Press.

PS.1.05

Missing /y/: Vowel perception in bilinguals whose languages differ in whether the high front rounded vowel is phonemic

Lei Pan*, Centre for Research in Child Development, National Institute of Education, Singapore
Suzy J Styles, School of Social Sciences, Nanyang Technological University, Singapore
*Corresponding Author Email: lei.pan@nie.edu.sg

Previous studies of perception for /b/ and /p/ in English and Mandarin Chinese have shown that the phonemic processing of stop consonants in bilingual speakers differs across a bilingual speaker's languages (Ke et al., 2021; Pan et al., 2022). The high front rounded vowel /y/ occurs in Mandarin as a category but not in other official languages of Singapore (Pan et al., 2023), leading to a mismatch in the vowel inventories of different bilinguals in Singapore. A 16-step vowel continuum ranging from /i/ to /y/ was created, by synthesizing intermediate steps between two tokens of natural speech using the Tandem-STRAIGHT approach (Kawahara et al., 2008). This study reports investigations of vowel perception in English-Mandarin bilinguals (N = 76) and a subgroup comparison between two groups of bilinguals: English-Mandarin bilinguals (N = 51) versus English-Malay bilinguals (N = 51) using a lexical identification task for Singapore Mandarin words 椅 (/i2/ 'chair') and 鱼 (/y2/ 'fish'), as illustrated in Figure 1. For early bilinguals of English and Mandarin, a Principal Component Analysis derived 'bilingual balance' factor did not predict individual identification function in linear regression models of vowel perception. However, the two bilingual groups exhibited different perceptual gradients across the /i/ to /y/ vowel spectrum, due to the differences in their linguistic experience (see Figure 2 for individual patterns of language exposure), with no differences in the location of the perceptual boundaries. As shown in Figure 3, the steepness of the identification functions differs among bilinguals with different linguistic experience, with steeper slopes for early English-Mandarin bilinguals (for whom the /y/ vowel is phonemic) and shallower slopes for early English-Malay bilinguals (for whom /y/ is not phonemic, but is largely discriminable in the forced choice task). With nuanced language background information, this finding suggests that exposure to both /i/ and /y/ in early development shapes phonemic perception. In addition, a continuous variable of bilingualism highlighting Mandarin experience was implemented in exploratory analyses, revealing that a continuous model of 'bilingual balance' is more sensitive for detecting the effects of linguistic experience on bilingual phoneme perception than a categorical model. These findings suggest that early language exposure patterns, in terms of continuous timing and amount of linguistic experience, have impacts on perception for high front vowels in adult bilinguals with different language backgrounds. Continuous measures of bilingual exposure are therefore highlighted as useful tools in the investigation of phoneme perception.

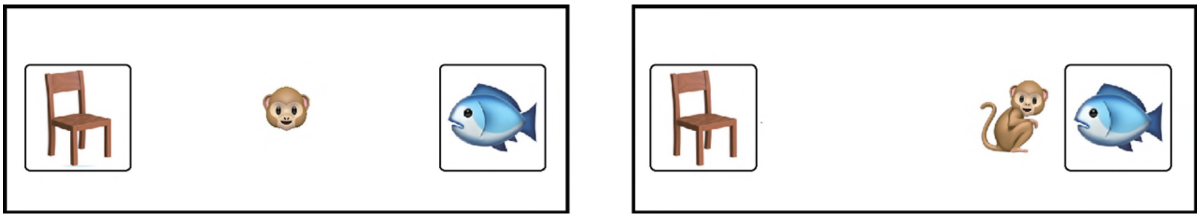


Figure 1 Example screens from the lexical identification task for testing high front vowel perception. Left: example of stimuli /i2/ and /y2/ presented on screen. Right: feedback following choice of target image /y2/.

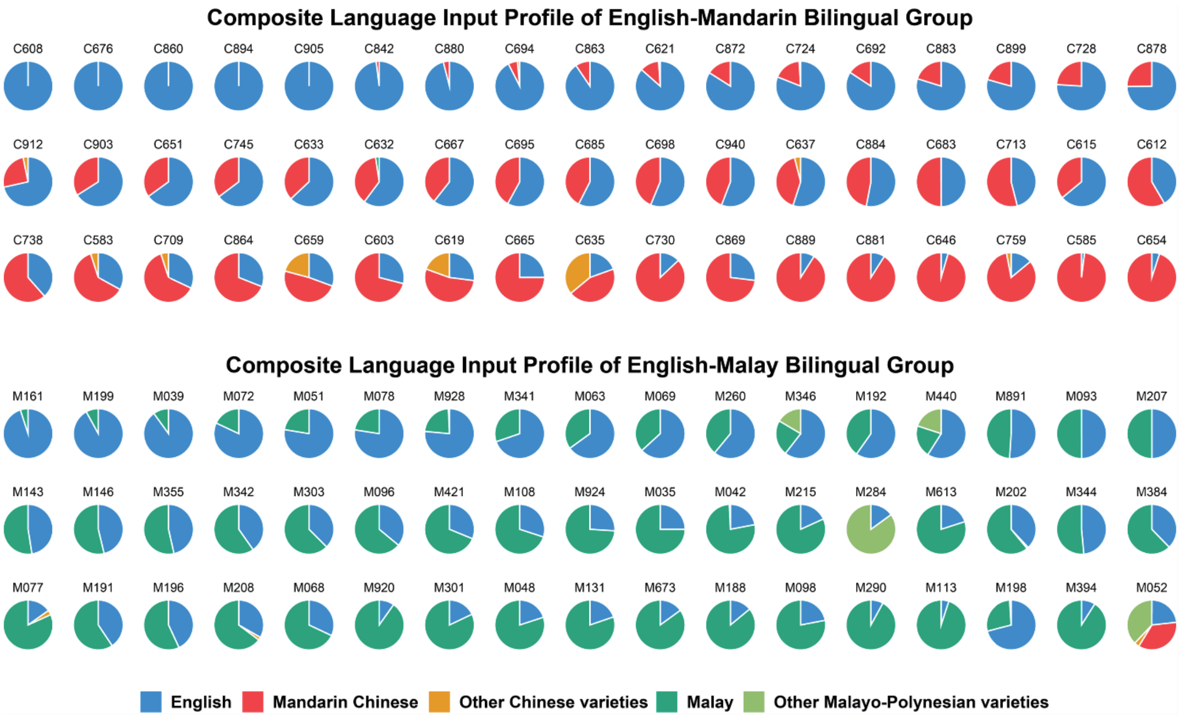


Figure 2 Composite Language Input Profile for each participant separated by language group. Upper: English-Mandarin bilinguals (N = 51) ordered by bilingual balance. Bottom: English-Malay bilinguals (N = 51) ordered by PCA-derived bilingual balance.

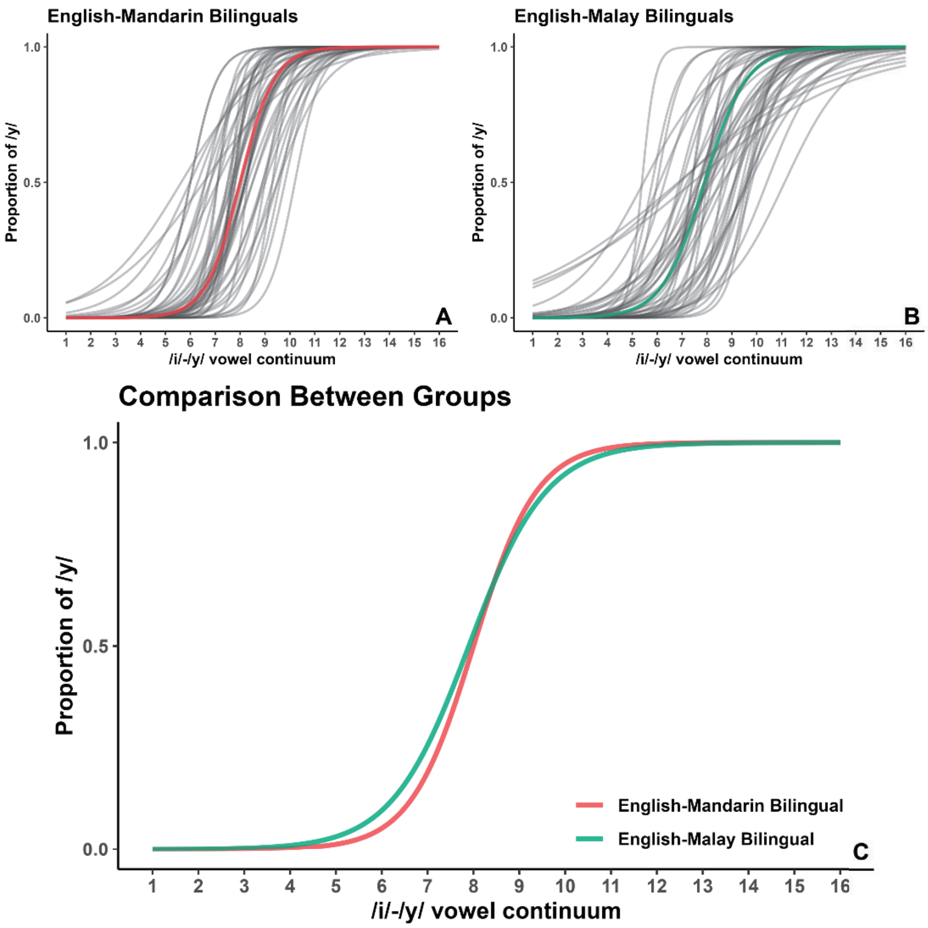


Figure 3 Psychometric curves in the high front vowel identification task. Two groups shown separately, with median slope participant highlighted. A: English-Mandarin bilinguals (N = 51); B: English-Malay bilinguals (N = 51) ; C: Medians for each group shown together for comparison.

Reference

Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., & Banno, H. (2008). Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, 3933–3936. https://doi.org/10.1109/ICASSP.2008.4518514

Ke, H., Pan, L., O'Brien, B. A., & Styles, S. J. (2021). *Phoneme Identification in Bilingual Children in Singapore – A preregistered investigation using the CROWN Game*. OSF. https://doi.org/10.31234/osf.io/3cgjh

Lisker, L., & Abramson, A. S. (1964). A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *WORD*, *20*(3), 384–422. https://doi.org/10.1080/00437956.1964.11659830

Pan, L., Ke, H., & Styles, S. J. (2022). Early linguistic experience shapes bilingual adults' hearing for phonemes in both languages. *Scientific Reports*, *12*(1), 4703. https://doi.org/10.1038/s41598-022-08557-7

Pan, L., Moisik, S. R., & Styles, S. J. (2023). *Being a round /y/: An acoustic description of high front vowels in Singapore Mandarin elicited by speakers with different bilingual balance in Mandarin and English*. https://doi.org/10.31234/osf.io/rdc6f

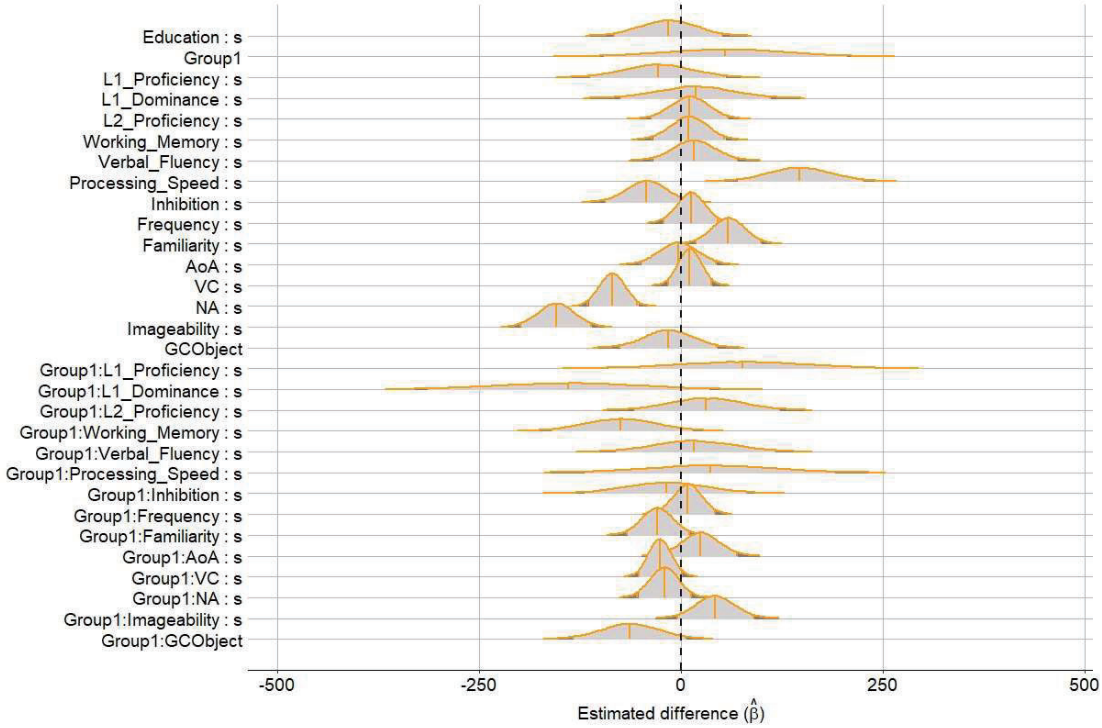
PS.1.06

Different cognitive and linguistic skills affect language production in younger and older adults

Mohammad Momenian
The Department of Chinese and Bilingual Studies, Hong Kong Polytechnic University
mohammad.momenian@polyu.edu.hk
Florian Hintz
Deutscher Sprachatlas, Philipps University of Marburg
Max Planck Institute for Psycholinguistics

While, there has been some research showing which aspects of language deteriorate during healthy aging (Shafto & Tyler, 2014), not much research has been done on the role of both cognitive and linguistics skills in the same study (see Hintz & Momenian, 2024). The purpose of this study was to determine what linguistic and cognitive functions have the biggest effect on language production in younger and older Cantonese speakers. 113 native speakers of Cantonese (67 people > 65 years old and 46 people < 65 years old) were recruited from Hong Kong for this study. Cognitive functions such as working memory (WM) (Forward digit span task), domain-general processing speed (Neutral condition of Flanker task), verbal fluency (Semantic fluency task) and inhibitory control (IC) (Incongruent condition minus Congruent condition in Flanker task) were measured. Psycholinguistic properties of the words such as age of acquisition (AoA), imageability, name agreement (NA), objective visual complexity (VC), familiarity and frequency as well as participants' language experiences such as L1/L2 proficiency and L1/L2 dominance were included as linguistic variables. Picture naming task, including both object and action words, was used to measure language production. We used Bayesian mixed effects analysis to analyse the data. In the analysis, we looked at the interaction between group and the above-mentioned variables. The findings show that among all the cognitive functions tested, domain-general processing speed and IC show similar effects on language production in both young and older adults. WM has opposing effects in the two groups: while higher WM capacity is associated with lower RT in the younger group, in the older group higher WM capacity resulted in slower RT. When it comes to linguistic variables, imageability, NA, and familiarity showed similar significant effects in both groups. However, VC and AoA showed effects only in the older group. L1 proficiency was only significant in the older group, while L1 dominance showed opposing effects across the two groups. See Figure 1 for the posterior probabilities. The findings of this study demonstrate that a mix of cognitive and linguistic skills are important predictors of language production in the older population. Our study demonstrates that researchers working on healthy aging need to measure several individual differences in order to have more valid findings.

Figure 1: posterior probability



References

Hintz, F., & Momenian, M. (2024, 1st revision). Domain-general processing speed influences structural assembly but not lexical retrieval during language production. *Psychonomic Bulletin & Review*.

Shafto, M. A., & Tyler, L. K. (2014). Language in the aging brain: the network dynamics of cognitive decline and preservation. *Science, 346*(6209), 583-587.

The role of musical aptitude in L2 tonal word learning: Evidence from cross-situational statistical learning

Xia Wang[1], Yuxin Ge[1], Padraic Monaghan[1], and Patrick Rebuschat[1,2]

[1] Lancaster University, UK. [2] University of Tübingen, Germany

Second language (L2) learners acquire words in natural environments by tracking the co-occurrence of sounds and referents over multiple times and eventually mapping between them. This process is called Cross-situational statistical learning (CSL) (Yu & Smith, 2007). In addition to tracking the co-occurrence, learners must perceive the non-native sound contrasts to discriminate different L2 sounds. Due to the absence of tones in learners' native languages and insufficient learning instructions during CSL, learners find it hard to perceive and apply tonal cues. Cross-domain transfer (Ong et al., 2016) suggests that a strong perception ability in music processing can benefit language processing if similar abilities are required in both domains. Based on this, we have explored whether musical aptitude could affect the CSL performance in learning Chinese words. Smit et al. (2022) found the role of musical aptitude in CSL of native-like pseudowords. However, there is no study about whether musical aptitude can facilitate CSL on non-native sound words, and the current study aims to fill this gap.

49 native English speakers were asked to learn 12 disyllabic Chinese pseudowords (see Table 1) in a CSL paradigm, where participants saw two pictures corresponding to minimal pair words and heard a sound simultaneously in a trial. Initially, they had to guess which picture matched the sounds, but after several trials, they might find the statistical co-occurrence of sounds and pictures and map the words to their referents. There were four trial types of minimal pairs (see Table 2): One at the segmental level (v-MP) and three at the suprasegmental level (p-MP, d-MP, dp-MP). These types allowed us to test participants' perception ability of Chinese phonological details at both levels during learning. After the CSL task, participants completed retrospective verbal reports to assess their awareness of non-native sounds and five musical aptitude tests (timbre, tempo, pitch, loudness, and melody). Results are visualized in Fig.1, 2.

Results demonstrate: 1. Non-native word pairs (e.g., p-MP) are more difficult for learners during CSL. 2. Melodic perception ability facilitates CSL on word pairs differing in tonal cues, while loudness perception ability predicts CSL on word pairs contrasting in segmental sounds. This shows the cross-domain transfer of perception abilities between music and language.

Supplemental Material

Table 1 *Pseudowords for CSL task*

	Tone 1	Tone 2	Tone 4
Short tone	/pa1lu1/	/pa2lu1/	/pa4lu1/
	/pi1lu1/	/pi2lu1/	/pi4lu1/
Long tone	/pa:1lu1/	/pa:2lu1/	/pa:4lu1/
	/pi:1lu1/	/pi:2lu1/	/pi:4lu1/

Table 2 *The examples of minimal pairs for each trial type*

v-MP		p-MP		d-MP		dp-MP	
minimal pairs contrasting on one vowel		minimal pairs contrasting on the pitch of tones		minimal pairs contrasting on the duration of tone		minimal pairs contrasting on pitch and duration	
/pi1lu1/	/pa1lu1/	/pi1lu1/	/pi2lu1/	/pi1lu1/	/pi:1lu1/	/pi1lu1/	/pi:2lu1/

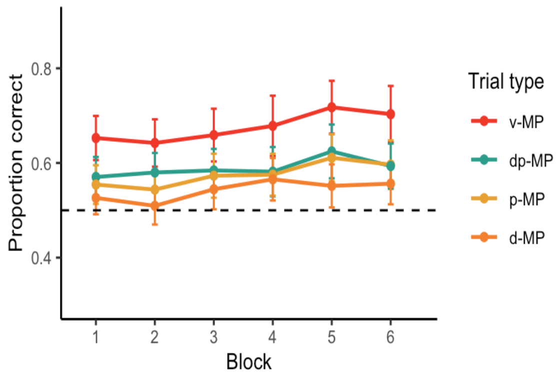


Fig. 1 *Learning trajectory*

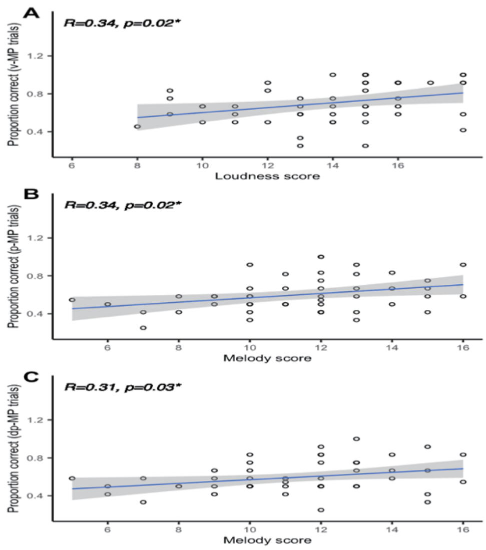


Fig. 2 *The relation between musical aptitude and learning performance in different trial types.*

Reference

Ong, J. H., Burnham, D., Stevens, C. J., & Escudero, P. (2016). Naïve learners show cross-domain transfer after distributional learning: The case of lexical and musical pitch. *Frontiers in psychology*, 7, 1189.

Smit, E. A., Milne, A. J., & Escudero, P. (2022). Music perception abilities and ambiguous word learning: Is there cross-domain transfer in non-musicians? Frontiers in Psychology.

Yu, C., & Smith, L. B. (2007). Rapid Word Learning Under Uncertainty via Cross-Situational Statistics. Psychological Science, 18(5), 414–420.

PS.1.08

Does prior knowledge benefit Mandarin-speaking older adults' memory consolidation in Cantonese tone learning?

Kangdi Liu [1] (kliubc@connect.ust.hk), Yan Feng [2], Zhen Qin [1]
1 The Hong Kong University of Science and Technology; 2 Nanjing University of Science and Technology

Sleep supports young adults' perceptual learning of nonnative speech categories through memory consolidation (i.e., memory is preserved better during sleep than during waking) [1]. Research showed that newly learned information aligned with prior knowledge was consolidated more rapidly than that does not [2]. In Mandarin-speaking young adults' Cantonese tone learning, prior knowledge was manipulated according to L1 [3]. The pitch contour cue in the Cantonese contour-level contrast (henceforth, contour tones) is consistent with their prior L1 knowledge, while the pitch height cueing the level-level contrast (henceforth, level tones) is not as L1 lacks the contrast. They consolidated contour tones better than level tones (i.e., improved in the trained-talker posttest after sleep). Older adults, however, forgot equally in both intervals due to age-related changes in consolidation and sleep [4]. Given the age-related deficits in memory consolidation, this study examined whether prior knowledge would benefit tonal consolidation among older adults.

Following the design of [3], 60 Mandarin-speaking older adults (aged 55-78) were recruited (Fig 1). 30 participants were trained to learn Cantonese contour (T5 vs. T6 cued by pitch contour, aligned with their prior knowledge) and level tones (T3 vs. T6 cued by pitch height, not aligned) produced by a male talker in the evening (with sleep after training), while 30 trained in the morning (without sleep), which followed by three posttests: immediate identification (ID1), 12-hour delayed ID2, and novel-talker (female) ID3. A Fitbit recorded sleep to test whether it facilitated learning [5].

Two mixed-effects logistic regression models were run on accuracy. Three variables, Posttest (ID1/ID2/ID3), Group (Evening/Morning), and Tone (Contour/Level), with by-subject random slope and intercept, were entered in the models. The model regarding ID1 and 2 did not yield a three-way interaction (Fig 2). Crucially, the one involving ID2 and 3, revealed a three-way interaction (Fig 3). Post-hoc analysis indicated, in contour-tone learning, the evening group did not show a difference between ID2 and ID3 ($z=0.62$, $p=.53$), but others exhibited lower accuracy in ID3 than in ID2 (Contour-Morning, $z=4.74$, $p<.001$; Level-Evening, $z=4.53$, $p<.001$; Level-Morning, $z=5.40$, $p<.001$). It suggests a generalization between trained and novel talkers for contour tones in the evening group. Sleep parameters positively predicted the generalization effect (ID 3-2; in Fig 4).

Unlike young adults (aged 19-33) showing a beneficial effect of prior knowledge on memory consolidation by improving performance in the trained-talker posttest after sleep [3], older adults benefited through talker generalization (although nonnative tones are challenging for L2 learners due to significant variabilities across talkers [6], compared to other speech categories [1]). This study further demonstrates the role of prior knowledge in memory consolidation across the lifespan.

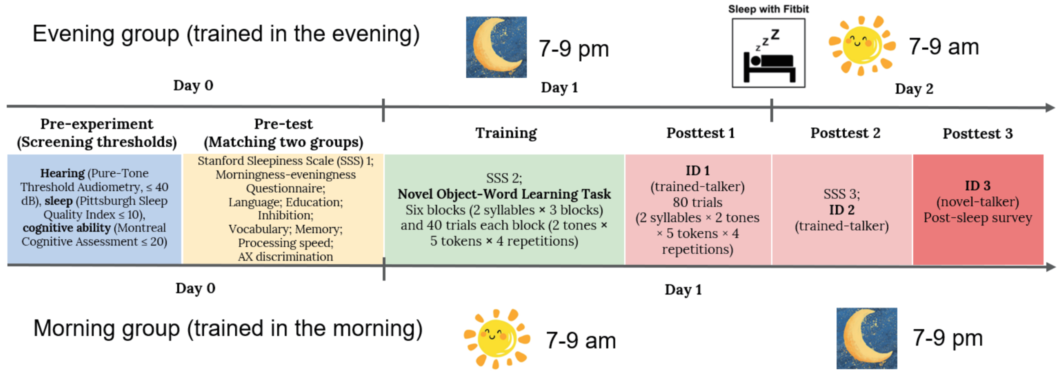
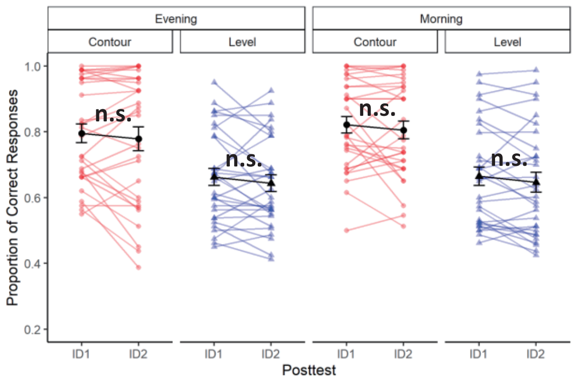


Figure 1 Experimental procedure



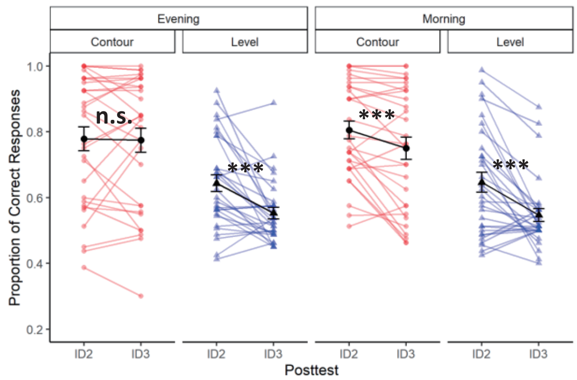


Figure 2 Accuracy in ID1/2 (immediate vs. 12-hour) Figure 3 Accuracy in ID 2/3 (trained vs. novel talker)

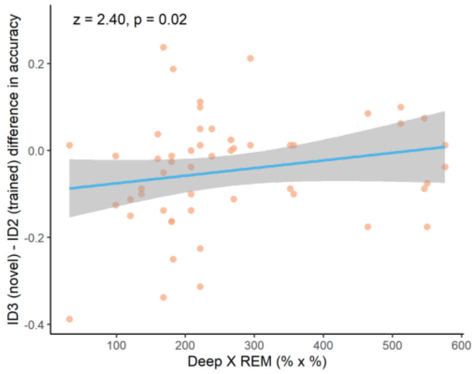


Figure 4 Sleep parameters (percentage product of Deep X REM) positively predicted the ID3-ID2 difference in accuracy (novel-trained talkers). Deep: deep sleep (Stage 3 of non-rapid eye movement sleep); REM: rapid eye movement sleep

References

[1] Earle, F. S., & Myers, E. B. (2015). Overnight consolidation promotes generalization across talkers in the identification of nonnative speech sounds. *Journal of the Acoustical Society of America*, 137(1), EL91–EL97.

[2] Hennies, N., Lambon Ralph, M.A., Kempkes, M., Cousins, J.N., & Lewis, P.A. (2016). Sleep spindle density predicts the effect of prior knowledge on memory consolidation. *Journal of Neuroscience*, 36, 3799 - 3810.

[3] Qin, Z., Jin, R., & Wu, R. (2023). The role of prior knowledge in second-language learners' overnight consolidation of Cantonese tones. In R. Skarnitzl, & J. Volín (Eds.), *Proceedings of the 20th International Congress of the Phonetic Sciences (ICPhS 2023)* (pp. 1881-1885). Prague: Guarant International.

[4] Kainec, K. A., Paracha, A. W., Ali, S., Bussa, R., Mantua, J., & Spencer, R. (2022). Age-related changes in sleep-dependent novel word consolidation. *Acta Psychologica*, 222, 103478.

[5] Batterink, L. J., Oudiette, D., Reber, P. J., & Paller, K. A. (2014). Sleep facilitates learning a new linguistic rule. *Neuropsychologia*, 65, 169–179.

[6] Qin, Z. & Zhang, C. (2019). The effect of overnight consolidation in the perceptual learning of non-native tonal contrasts. *PLoS ONE*, 14(12), e0221498.

न Shabd: A Pseudoword Generator for Hindi and other Indian languages

Selection of word/non-word stimuli is crucial in lexical decision tasks relying on the assumption that participants must access the mental representation of a word to classify it. Various programs such as Wuggy (Keuleers, 2010), LINGUA (Westbury, 2007), WordGen (Duyek, 2004), MCWord (Medler, 2005), WordCreator, WordPars (Esmaeelpour, 2021), are primarily designed for alphabetic languages like English. The current study introduces a tool specifically designed to generate pseudowords for Indian languages. Pseudowords generated through this platform have undergone validation through lexical decision tasks and neural network analysis, filling a notable gap in the availability of resources for pseudoword research in the Hindi.

Pipeline for pseudoword generation: The program takes positional bigram chain with their frequency as an input for any selected language. To build such bigram chain, each item in the corpus is segmented into subsyllabic units. (a) First, the input word is segmented at sub syllabic level; (b) Now, to generate pseudowords from given template, the algorithm replaces $unit_{N+1}$ element in its bigram chain with another unit. (c) Bigram frequencies are found in language module database. (d) Based on bigram frequency at each position, all possible combinations are found for first three items of the tuple i.e. ($unit_N$, n, $Total_N$) and different fourth item ($unit_{N+1}$). It ensures that a particular length of pseudowords are generated based on similar length of words. (e) All plausible combinations are searched based on nearest neighbourhood frequency deviations or with highest frequency. Each unit in chain is swapped with 5 such best units to give a list of pseudowords. Example for आंकड़े, generated pseudowords are (**ब**कड़े, आं**ग**ड़े, आंक**र**).

Validation: 280 participants (183 males; 93 females) were presented with 180 trials were shown (120 words, 60 pseudo-words) on screen. The words were selected from a Hindi psycholinguistic corpus, Shabd (Verma et al., 2021) and drawn from a wide range of frequencies. The task was self-paced, with Yes/No responses if the words are meaningful.

Neural network-based validation: Since NN are very good in finding statistical relationships among data, we designed a deep NN architecture to learn more about relationships between words and pseudo-words generated by our algorithm. We trained two NN, (a) with the equal proportion of non-words as words, and (b) 4:1 ratio of non-words to words. We used the trained network to classify both versions of the pseudowords.

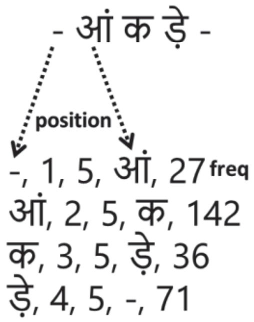
Results: We observed a significant difference in mean accuracy between words (M=0.88, SD=0.09) and pseudowords (M=0.36, SD=0.22); $t(254) = 29.681$, $p < 0.001$. There was significant difference in mean RT between words (M=1320.18, SD=496.2) and pseudowords (M=3194.83, SD=1630.6); $t(254) = -16.2$, $p < 0.001$.

References:

- Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. Behavior Research Methods, 42(3), 627–633. https://doi.org/10.3758/brm.42.3.627
- Westbury, C., Hollis, G., & Shaoul, C. (2007). LINGUA. The Mental Lexicon, 2(2), 271–284. https://doi.org/10.1075/ml.2.2.09wes
- Duyck, W., Desmet, T., Verbeke, L. P. C., & Brysbaert, M. (2004). WordGen: A tool for word selection and nonword generation in Dutch, English, German, and French. Behavior Research Methods, Instruments, & Computers, 36(3), 488–499. https://doi.org/10.3758/bf03195595
- Medler, D.A., & Binder, J.R. (2005). MCWord: An On-Line Orthographic Database of the English Language. http://www.neuro.mcw.edu/mcword/
- Esmaeelpour, E., Saneei, S., & Nourbakhsh, M. (2021). WordPars: A tool for orthographic and phonological neighborhood and other psycholinguistic statistics in Persian. Behavior Research Methods, 54(4), 1902–1911. https://doi.org/10.3758/s13428-021-01712-4
- Verma, A., Sikarwar, V., Yadav, H., Jaganathan, R., & Kumar, P. (2021). Shabd: A psycholinguistic database for Hindi. Behavior Research Methods, 54(2), 830–844. https://doi.org/10.3758/s13428-021-01625-2

Table-1: Bigram chain instance for single word. Frequencies are indicative and calculated based on whole corpus

$unit_N$	n	Total N	$unit_{N+1}$	Freq.
-	0	4	सं	188
सं	1	4	भा	15
भा	2	4	वि	5
वि	3	4	त	10



Additional information about language (Hindi):

Most Indian languages adhere to the concept of a character unit, known as *akshara*, serving as the fundamental linguistic building block. An *akshara* [CV] typically comprises consonants [C] and a vowel [V], with words being constructed from more than one *aksharas*.

Hindi differs from other languages in terms of how syllables and vowels are defined in the language. Consonants in the language have hidden vowels at their end, also called as *schwa*, for example, क /ka/ is just *akshara* in Hindi but can be considered as a whole syllable in some cases. However, in most cases, diacritic markers i.e. vowels are added to such *aksharas* or consonants to add to its phonology. Such markers are called *matras* and can be added to the *aksharas* in both linear and non-linear fashion, example (क+ा = का /kaa/ and क+ि = कि /ki/). Such markers (*matras*) can also be written to top or bottom of the consonant (*akshara*), each changing the pronunciation in different ways. Thus, we defined sub syllabic unit in Hindi as a combination of *akshara* (consonant) + *matra* (diacritic marker or vowel). (if *matra* is present), or just *akshara* (consonant) (if *matra* is absent).

The program takes positional bigram chain with their frequency as an input for any selected language. To build such bigram chain, each item in the corpus is segmented into subsyllabic units, wherein each unit is defined in Hindi as akshara + matra (see above). From each unit, a bigram tuple (Table 1) is constructed with four attributes ($unit_N$, n, $Total_N$, $unit_{N+1}$, Freq). Here $unit_N$ - $unit_{N+1}$ forms a bigram with two subsyllabic units, n represents the position of such bigram in any word, while $Total_N$ represents the total number of bigrams possible in that particular word. Similarly, we calculated frequency of all possible tuples in the corpus of 96000 words from Shabd (Verma et al., 2021). This ensures that bigram unit chain consists of position information along with frequency of occurrence in a particular language.

These bigram chains are used as a language module in subsequent steps. Similar to Wuggy algorithm (Keuleers & Brysbaert, 2010), our pseudoword generator also uses template word(s) to create pseudowords. We swapped second element ($unit_{N+1}$) in bigram chain successively to create pseudoword.

We started replacing with the weakest bigram first. The weakest bigram was the one with least bigram frequency, and maximum number of potential replacement candidates. The candidate with least frequency deviation from the original bigram was selected. This ensured that frequent bigrams were replaced with frequent ones and less frequent bigrams with less frequent ones. We repeated these steps to iteratively replace multiple bigram elements. The number of iterations depends on the word length of input template. We replaced minimum 33% to maximum of 50% of bigrams. For example, for sub-syllabic word length 2 and 3, we replaced 1 unit; for length 4, 5 and 6 we replaced 2 units; for 7, 8 and 9 we replaced 3 units and so on.

We have segmented words by considering *akshara* + *matra* as a single unit. However, since Indic languages are alpha syllabic, and sub syllabic units are not distinctly defined, future experiments should compare between subsyllabic and syllabic levels. This tool is likely to improve the quality of the nonwords used in lexical decision tasks and other psycholinguistic experiments. The procedure computes matching nonwords in little time.

The uniformity observed in Indian languages, where words are formed using character units known as *aksharas* – comprising consonants and a vowel – implies that our program can generate pseudowords for any language with a corresponding bigram language module. Additionally, we've developed a script to construct this module for any language based on a dictionary of words specific to that language. Future plans involve compiling corpora and adding modules for all possible Indian languages, with an upcoming large-scale multilingual validation study for the generated pseudowords. This article serves as a proof of concept, expanding on previous algorithms like Wuggy, and specifically explores the *akshara-matra* combination as a sub-syllabic unit in pseudoword generation.

The role of semantic transparency in visual word recognition of compound words in Korean: A megastudy approach

Say Young Kim[1,2,*] & Hyoung Sun Kim[1]
[1]Hanyang University, Seoul, South Korea ([*]sayyoungkim@hanyang.ac.kr)
[2]Hanyang Institute for Phonetics and Cognitive Sciences of Language

As compound words consist of two (or more) independent lexemes, they provide an ideal platform for testing how morphemic units are processed, frequently by estimating the relative contribution of semantic relatedness between a compound (e.g., *limelight*) and its constituents *(lime* and *light*) over and above lexical properties such as length and frequency (e.g., Gagné et al., 2019; Günther et al., 2020; Juhasz et al., 2015; Kim et al., 2019). Korean, a less studied language on this matter, is known for its alphasyllabary system, offering the opportunity to examine whether semantic transparency interacts with the writing system. In particular, salient syllable units in Korean are often homographs, resulting in a high degree of ambiguity for each syllable within a multisyllabic word. For instance, in the disyllabic compound, 비옷 *pi-os* (meaning “raincoat”), the first constituent 비 *pi* (meaning “rain”) can also serve as a constituent of other compounds (e.g., 비극 *pi-kuk*, meaning “tragic drama”). In relation to this issue, our previous study (Kim et al., 2015) demonstrated asymmetric processing of prefixed versus suffixed derived words, implying greater ambiguity for monosyllabic prefixes due to their delayed resolution in contrast to monosyllabic suffixes.

To further address this issue, we used a megastudy approach to overcome some methodological constraints typically encountered in factorial designs. Semantic transparency scores for 1,687 Korean compounds were collected at the constituent level, with native speakers judging how strongly the compound is related to each constituent in terms of meaning. To assess the relative contribution of semantic transparency between a compound and its constituents on compound word recognition, we conducted item-level hierarchical regression analyses on compound lexical decision data from the Korean Lexicon Project (Yi et al., 2017). Our analyses indicated that both the first and second constituents play a significant role in compound word recognition, as semantic transparency at the constituent level accounted for unique variance in lexical decision on compounds. In addition, the interaction between the transparency of the two constituents was significant, suggesting that the influence of one constituent’s transparency increases when the other constituent is semantically opaque, not transparent. In conclusion, this study provides clear evidence for the significant role of each constituent in compound word processing in Korean.

References

Gagné, C.L., Spalding, T.L. & Schmidtke, D. (2019). LADEC: The Large Database of English Compounds. *Behavior Research Methods,* 51, 2152–2179

Günther, F., Marelli, M. & Bölte, J. (2020). Semantic transparency effects in German compounds: A large dataset and multiple-task investigation. *Behavior Research Methods,* 52, 1208–1224

Juhasz, B. J., Lai, Y.-H., & Woodcock, M. L. (2015). A database of 629 English compound words: Ratings of familiarity, lexeme meaning dominance, semantic transparency, age of acquisition, imageability, and sensory experience. *Behavior Research Methods, 47*, 1004– 1019.

Kim, S.Y., Wang, M., & Taft, M. (2015). Morphological decomposition in the recognition of prefixed and suffixed words: Evidence from Korean. *Scientific Studies of Reading, 19*, 183-203.

Kim, S. Y., Yap, M. J., & Goh, W. D. (2018). The role of semantic transparency in visual word recognition of compound words: A megastudy approach. *Behavior Research Methods, 51*, 2722–2732.

Yi, K., Koo, M., Nam, K., Park, K., Park, T., Bae, S., …Cho, J. R. (2017). The Korean lexicon project: A lexical decision study on 30,930 Korean words and nonwords. *The Korean Journal of Cognitive and Biological Psychology, 29*, 395–410.

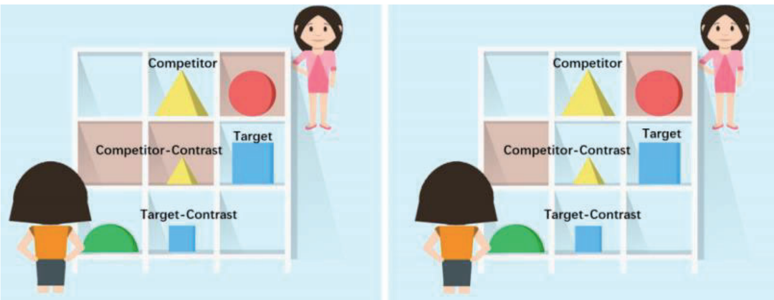
Perspective-taking in L1 vs L2 Referential Processing and Benefit of Speech Speed

Yipu Wei[1]; Chao Sun[1]; Yingjia Wan[2]
[1] School of Chinese as a Second Language, Peking University;
[2] Chinese Academy of Sciences, Institute of Psychology; wanyj@psych.ac.cn

Perspective-taking, the ability to understand others' viewpoints[1], is crucial to communication. While native speakers spontaneously use perspective information to anticipate referents[2,3], it remains unclear whether second language (L2) speakers do the same in real-time comprehension. Advanced L2 listeners might benefit from enhanced inhibitory control[4] and multicultural exposure[5], increasing their sensitivity to perspectives. On the other hand, L2 listeners may prioritize bottom-up linguistic cues over higher-level perspective cues[6], and have limited predictive processing[7], which could restrict perspective-taking performance[8].

We monitored participants' eye movements in a referential communication task using scalar adjectives. In trials assessing *spontaneous perspective-taking* (Fig.1), using perspective information enabled pre-identification of the target. We compared L2 listeners to L1 listeners at both normal and slow speech rates, with the slow rate set at two-thirds of the normal rate to provide extra processing time, and possibly enhance spontaneous perspective-taking. We also included trials where listeners must use perspectives to disambiguate target from competitor, indicating *strategic perspective-taking* (Fig.2).

We collected data from 32 advanced L2 learners of Mandarin Chinese (mean age: 23.3; accuracy rate of Chinese cloze test: 85%; HSK level: Grade 5-6, i.e. C1-C2 in CEF) and 31 native Chinese speakers. A significant interaction of language*ground reveals distinct patterns of L1 and L2 listeners in spontaneous perspective-taking (measured by target ratio; $\beta = -0.15$, $SE = 0.05$, $z = -3.23$, $p = .001$). At the scalar, L1 listeners exhibited ground effects—more looks to the target in the privileged-ground condition compared to shared-ground and benefited more from slower speed (Fig.3). However, L2 listeners did not show spontaneous perspective-taking, even with slower speech rate. In trials requiring strategic perspective-taking, L2 listeners performed similarly to L1s ($\beta = -0.03$, $SE = 0.03$, $z = -0.89$, $p = .38$), indicating that they can strategically use perspectives when necessary for referential decision-making. The findings suggest that L2 listeners rely more on bottom-up processing and utilize high-level cues like perspectives only when linguistic ambiguity persists.



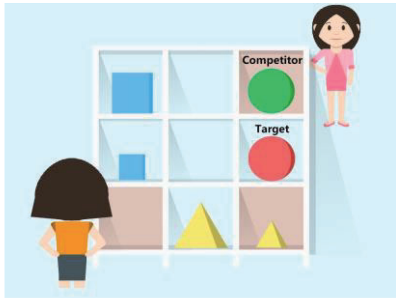


Fig. 1 Spontaneous perspective-taking trial: privileged-ground condition (left) vs shared-ground condition (right). Auditory input (Chinese): *Qing ba* | *da de na kuai* | *fangxing jimu gei wo.* 'Please give me |that big | cubic block.' Paradigm adapted from Heller et al. (2008). In the privileged-ground condition where the competitor-contrast (the small triangle) is not visible to the speaker, listeners who take the speaker's perspective, could exclude the competitor (the big triangle) and predict the target (the big cube) at the scalar *dade* 'big', as it would be infelicitous for the speaker to use *big* to refer to an object whose size contrast is not accessible.

Fig. 2 Strategic Perspective-taking trial: Auditory input (Chinese): *Qing ba* | *na kuai* | *qiuxing jimu gei wo.* 'Please give me | that | sphere block.' The target can only be distinguished from the competitor by perspective: the speaker cannot see the green sphere and must be referring to the red sphere.

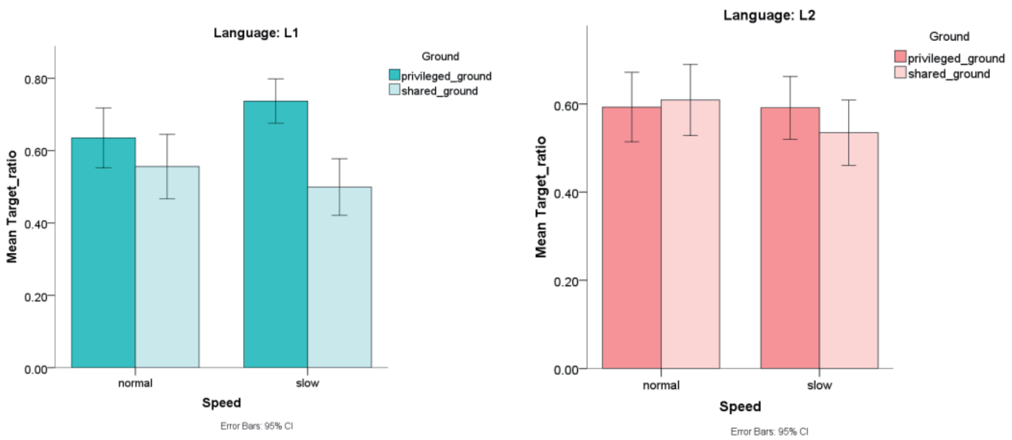


Fig. 3 Mean target ratio (fixation proportion on the target divided by fixation proportion on the target and competitor) under the normal-speed condition and slow-speed condition during the critical time window (from the onset of scalar adjective till the onset of shape) for L1 and L2 groups in spontaneous perspective-taking trials.

Reference

1. Apperly, I. *Top. Cogn. Sci.* **10**, 133–139 (2018).
2. Heller, D., Grodner, D. & Tanenhaus, M. K. *Cognition* **108**, 831–836 (2008).
3. Wei, Y., Wan, Y. & Tanenhaus, M. K. *Sci. Rep.* **14**, 8031 (2024).
4. Bialystok, E., Klein, R., Craik, F. I. M. & Viswanathan, M. *Psychol. Aging* **19**, 290–303 (2004).
5. Todd, A. R. & Galinsky, A. D. *J. Exp. Soc. Psychol.* **48**, 1394–1398 (2012).
6. Tsui, A. B. M. & Fullilove, J. *Appl. Linguist.* **19**, 432–451 (1998).
7. Grüter, T. & Hopp, H. *J. Mem. Lang.* **121**, 104281 (2021).
8. Ryskin, R. A., Brown-Schmidt, S., Canseco-Gonzalez, E., Yiu, L. K. & Nguyen, E. T. *J. Mem. Lang.* **74**, 46–76 (2014).

Testing Classifiers as Early Cues in Mermaid Constructions in Mandarin Chinese

Yupei Zhang
The University of Tokyo, yupeizhang.ling@gmail.com

Introduction. The Mermaid Construction (MMC) in Mandarin Chinese has been paid much attention to (Ono 2013, 2015). An example of MMC in Mandarin Chinese is given in (1) with its properties illustrated in (2) following Ono (2013, 2015).

Ono (2015) has argued that in the case of MMC, the sentence would be less natural when the pre-AC (Adnominal Clause) [Numeral+Classifier] complex is dropped.

Research question. It has been suggested that the Classifier-Noun mismatch would function as early cues for the processing of relative clauses (Wu et al. 2018). This study would examine whether MMCs in Mandarin Chinese exhibit early cue effects.

Proposal. I argue that early cue effects (Wu et al. 2018) are based on Classifier-Noun compatibility and could be applied when analyzing MMCs in Mandarin Chinese. MMCs have received much attention due to the property that the MMC subject and the MMC noun are not coreferential (see 2). In (4a), there is a mismatch between the MMC subject and the pre-AC classifier *fu* while the MMC noun is compatible with the classifier *fu*. Conversely, a noun that is coreferential to the subject would be ruled out due to constraints on Classifier-Noun compatibility (4b).

On the other hand, in (5), the general classifier *ge* is compatible with the subject, therefore a noun that is coreferential to the subject would not be ruled out (5b).

Evaluation experiment. One possible account would be that a pre-AC classifier that is incompatible with the MMC subject would function as an early cue for predicting the MMC noun [**Hypo. 1**]. Meanwhile, when the pre-AC classifier is compatible with the subject, the pre-AC classifier would hinder the prediction of the MMC noun [**Hypo. 2**].

This study tested the two hypotheses by conducting an experiment in which two factors were crossed, namely Numeral (Numeral vs. No Numeral) and Classifier (Matching Classifier vs. Mismatching Classifier vs. No Classifier). This study adopted 16 MMC nouns and constructed 4 sets of test suites for each noun, therefore 384 sentences in total. Baseline MMCs of [-Numeral, -No classifier] have been constructed to follow the [NP COP VP DE Noun] structure.

This study adopted a computational approach on the basis of surprisal (Hale 2001; Levy 2008). Specifically, Hypo. 1 expects a lower surprisal of the MMC noun when there is a pre-AC classifier which is incompatible with the MMC subject. Hypo. 2 expects a higher surprisal of the MMC noun when the pre-AC classifier is compatible with the MMC subject.

For sentence scoring, this study used a pre-trained large language model (LLM) uer/gpt2-chinese-cluecorpussmall (Zhao et al. 2019) and the minicons library (Misra 2022).

This study built linear mixed-effects models using R (R Core Team 2021) and lme4 (Bates et al. 2015). Variables examined are illustrated in Table 1.

Results. Results suggest that a pre-AC classifier which is incompatible with the MMC subject would function as an early cue for predicting MMC noun ($p < 0.05$) and a pre-AC classifier which is compatible with the MMC subject would hinder the prediction of MMC noun ($p < 0.05$). It has also been suggested by the results that there is an interaction between the numeral and the pre-AC classifier which is incompatible with the MMC subject ($p < 0.001$).

(1) *Ta* *shi* *ai* *guan xianshi* *de* *piqi*.
3SG COP love mind other.people's.business DE temper
'He likes to meddle in affairs of others.'
(Lit. 'He is the temper that he likes to meddle in affairs of others.')

(2) [$_{NP}$ *Ta*][$_{COP}$ *shi*][$_{Clause}$ *ai guan xianshi*][$_{DE}$ *de*][$_{NP}$ *piqi*] (NP COP Clause DE Noun) =(1)
Non-coreferential

(3) [$_{NP}$ *Ta*][$_{COP}$ *shi*][$_{NUM}$ *yi*][$_{CL}$ *fu*][$_{Clause}$ *ai guan xianshi*][$_{DE}$ *de*][$_{NP}$ *piqi*]
mismatch *match*

(4) a. MMC: [$_{NP}$ *Ta*][$_{COP}$ *shi*][$_{NUM}$ *yi*][$_{CL}$ *fu*][$_{Clause}$ *ai guan xianshi*][$_{DE}$ *de*] [$_{NP}$ *piqi*]
b. Non-MMC: *[$_{NP}$ *Ta*][$_{COP}$ *shi*][$_{NUM}$ *yi*][$_{CL}$ *fu*][$_{Clause}$ *ai guan xianshi*][$_{DE}$ *de*][$_{NP}$ *ren*]

(5) a. MMC: [$_{NP}$ *Ta*][$_{COP}$ *shi*][$_{NUM}$ *yi*][$_{CL}$ *ge*][$_{Clause}$ *ai guan xianshi*][$_{DE}$ *de*] [$_{NP}$ *piqi*]
b. Non-MMC: [$_{NP}$ *Ta*][$_{COP}$ *shi*][$_{NUM}$ *yi*][$_{CL}$ *ge*][$_{Clause}$ *ai guan xianshi*][$_{DE}$ *de*][$_{NP}$ *ren*]

(6) a. *Ta shi ai guan xianshi de piqi*. No Numeral*No classifier
b. *Ta shi* ***ge*** *ai guan xianshi de piqi*. No Numeral*Matching classifier
c. *Ta shi* ***fu*** *ai guan xianshi de piqi*. No Numeral*Mismatching classifier
d. *Ta shi* ***yi*** *ai guan xianshi de piqi*. Numeral*No classifier
e. *Ta shi* ***yi ge*** *ai guan xianshide piqi*. Numeral*Matching classifier
f. *Ta shi* ***yi fu*** *ai guan xianshi de piqi*. Numeral*Mismatching classifier

Variables	**Type**	**Description**
Surprisal	Outcome	Surprisal of the MMC noun
NUM	Fixed	1=pre-AC numeral, otherwise=0
Mismatching_CL	Fixed	1=pre-AC CL which is incompatible with MMC subject, otherwise=0
Matching_CL	Fixed	1=pre-AC CL which is compatible with MMC subject, otherwise=0
Noun	Random	The idiosyncrasy of the MMC noun
Test_Suite	Random	The idiosyncrasy of the test suite

Table 1: Variables examined.

References

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

Hale, John. 2001. A Probabilistic Earley Parser as a Psycholinguistic Model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies* NAACL '01, 1–8. USA: Association for Computational Linguistics.

Levy, Roger. 2008. Expectation-based Syntactic Comprehension. *Cognition* 106(3). 1126–1177.

Misra, Kanishka. 2022. minicons: Enabling Flexible Behavioral and Representational Analyses of Transformer Language Models. *arXiv preprint arXiv:2203.13112* .

Ono, Hideki. 2013. Mermaid Construction in Mandarin Chinese. In Tasaku Tsunoda (ed.), *Adnominal Clauses and the 'Mermaid Construction': Grammaticalization of Nouns*, 677–679. Tachikawa, Japan: National Institute for Japanese Language and Linguistics.

Ono, Hideki. 2015. Hironri-teki dootei kankei o arawasu kopyurabun no nittyuutaisyoo. *Gendai chuugokugo kenkyu [Contemporary Research in Modern Chinese]* 17. 13–25.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria. `https://www.R-project.org/`.

Wu, Fuyun, Elsi Kaiser & Shravan Vasishth. 2018. Effects of Early Cues on the Processing of Chinese Relative Clauses: Evidence for Experience-Based Theories. *Cognitive Science* 42. 1101–1133.

Zhao, Zhe, Hui Chen, Jinbin Zhang, Xin Zhao, Tao Liu, Wei Lu, Xi Chen, Haotang Deng, Qi Ju & Xiaoyong Du. 2019. UER: An Open-Source Toolkit for Pre-training Models. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations*, 241–246. Hong Kong, China: Association for Computational Linguistics.

A Network Approach to Examine Gender Differences in Word Associations for Singapore English Words

Wong Jin Jye and Cynthia S. Q. **Siew** (Department of Psychology, National University of Singapore; wongjinjye@gmail.com)

This research describes the differences between genders in word associations for words unique to Singapore Colloquial English ("Singlish"), building upon our existing efforts to study the structure of the Singlish lexicon (Wong & Siew, 2023). We investigated the differences in how men and women ascribe meanings to words which may not exist in more widely used forms of English. The word association task has been previously shown to be a viable method for investigating gender differences in word meanings (De Deyne, Warner, & Pefors, 2023). However, instead of examining word associations directly, a network science approach was used to compare association networks constructed from word association data, since association networks have demonstrated additional utility in measuring semantic relatedness by incorporating spreading activation (De Deyne et al., 2019).

Responses from 2,646 participants (1,287 male, 1,359 female) were considered. 2,477 (93.6%) of the participants were born in Singapore, and 2,632 (99.4%) of the participants resided in Singapore at the time of the study. 2,451 (92.6%) of the participants spoke English as a first language; the most common second languages spoken were Mandarin Chinese (n = 2186, 82.6%) and Malay (n = 652, 24.6%). 238 Singlish words and phrases, sourced from Wikipedia's page on Singlish Vocabulary, were presented to the participants; approximately half of the participants provided responses to half of the cues, and the other half of the participants to the other half of the cues. Wikipedia was chosen due to its crowd-sourced nature: the terms present were deemed more likely to be known due to them being user-submitted. When presented with a cue, participants were instructed to provide up to three responses that came to their mind. Participants could also indicate if they did not know the word or if they recognized the word but were unable to think of any associations to it.

Two word association networks were constructed; one based on responses from men, the other based on responses from women. Network metrics which characterized global structural patterns of word associations was obtained from both networks. To evaluate if numerical differences between the network structures were statistically meaningful, 1,000 random male and female word association networks were generated by randomly re-assigning the original gender associated with each cue-association pair, creating a distribution of possible differences. Observed differences between the male and female network measures were statistically significant relative to random expectation exemplified by the bootstrapping procedure. Overall, the results further reinforce that gender differences, both in the structure of the mental lexicon and in the perception of individual words, exist.

References

De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2019). The “Small World of Words” English word association norms for over 12,000 cue words. *Behavior research methods, 51*, 987-1006.

De Deyne, S., Warner, S., & Perfors, A. (2023). Common words, uncommon meanings: Evidence for widespread gender differences in word meaning. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 45, No. 45).

Wong, J. J., & Siew, C. S. Q. (2023). Word Association Network of English Words Unique to Singapore English. In P. L. Villagrá, & X. Li (Eds.), *Proceedings of the International Workshop on Cognitive AI 2023 co-located with the 3rd International Conference on Learning & Reasoning (IJCLR 2023)*. https://ceur-ws.org/Vol-3644/

Network metric difference (*male - female)*	Generated distribution		Actual value
	M	*SD*	
Edge count	-642	256	1470***
Vertex count	-96.2	77.6	398***
Mean edge weight	0.000886	0.00183	0.0146***
Mean vertex degree	-0.0455	0.0739	-0.140***
Average shortest path length	-0.0122	0.862	0.113***
Diameter	-0.0249	3.20	1.34***
Global clustering coefficient	-0.0000653	0.000537	-0.000163***
Local clustering coefficient[1]	-0.000577	0.00618	-0.00456***

****p* < .01.
[1] the mean local clustering coefficient of every node.

Table 1: Actual values of differences in network metric values compared with the differences obtained from the randomly generated networks. Differences were calculated by subtracting the female network value from the male network value: positive values denote that the male network value was higher, while negative values denote that the female network value was higher.

Male	Female
atas, good, no, handsome, lucky	*money, market, red, good, lucky*

Table 2: Some responses to the Singlish phrase "white horse". Note how male responses capture more of the Singaporean-specific connotation "the son of a VIP in National Service". "Atas" means "refined" or "high-class".

Human and AI Processing Animal Metaphors in English and Chinese

Ben Yue Shi Yang and jenny Yichun Kuo
National Chung Cheng University, Taiwan
National Chiayi University, Taiwan

This study examines how humans and ChatGPT process “ox” and “dragon” metaphors in English and Chinese, revealing cultural values, language structures, and AI challenges. While prior research (Yang & Kuo, 2024) explored human processing of these metaphors, table 1shows the mean and standard deviation of niu (牛), long (龍), ox, cow, bull, buffalo, and dragon uniquely compares human and AI processing to uncover similarities and differences.

Thirty sentences containing Chinese “niu 牛" (ox) and "long(龍)” dragon, and English ox, cow, bull, buffalo, and dragon were collected from the Academia Sinica Balanced Corpus of Modern Chinese and the Corpus of Contemporary American English. Fifteen Chinese and fifteen English speakers rated the semantic prosody of ox and dragon metaphors on a 1 to 5 scale, and ChatGPT performed the same task. Results showed that both humans and ChatGPT identified positive connotations of strength in ox metaphors (e.g., “built like an ox”) and negative connotations of stubbornness (e.g., “stubborn as an ox”). However, humans provided more context-sensitive interpretations, especially for nuanced uses like “having a cow” (overreaction), which ChatGPT missed. For dragon metaphors, while Chinese dragons (龍) symbolize power and prosperity, English speakers often see dragons as symbols of danger. Interestingly, ChatGPT’s ratings highlighted these cultural contrasts more than the human participants did, raising questions about AI’s role in interpreting cultural differences.

English speakers rating “cow” more negatively than “bull,” and ChatGPT rating “dragon” more negatively than humans, suggesting that AI may amplify perceived cultural patterns. However, the lack of clear differentiation between English and Chinese speaker ratings of dragons challenges the assumption that these perceptions are deeply culturally ingrained. The findings show that ChatGPT aligns with human ratings but often lacks the personal and contextual depth humans provide.

This study underscores the effectiveness of ChatGPT in capturing broad language patterns but highlights its limitations in grasping the nuanced interpretations that come from personal and cultural experiences. Future research should explore how AI can better simulate the complex cultural and experiential elements that underpin human metaphor processing.

Keywords: : conceptual metaphors, semantic prosody, human processing, AI processing, sociocultural context

References:

Chinese Knowledge and Information Processing. (1994). *Academia Sinica Balanced Corpus of Modern Chinese.* Taipei City, Taiwan.

Davies, Mark. (2008-) *The Corpus of Contemporary American English (COCA).* Available online at https://www.english-corpora.org/coca/.

Schmid, H.-J., & Ungerer, F. (2006). *An introduction to cognitive linguistics: Friedrich Ungerer, Hans-Jorg Schmid.* Pearson Longman.

Yang, Ben Yue Shi and Kuo, Jenny Yichun (2024). *Conceptual Differences of Animal Metaphors Between English and Chinese.* Paper presented in The 11th Conference on Language, Discourse and Cognition. National Taiwan University, Taipei, Taiwan.

Table 1. Mean and standard deviation of human evaluations of conceptual animal metaphors in English and Chinese, and ChatGPT's evaluation of the metaphors.

Animal metaphors	Ox "Niu 牛"	Dragon "Long 龍"	Ox	Cow	Bull	Buffalo	Dragon
Mean (by human)	2.77	3.46	3.05	2.05	2.99	3.22	3.52
Standard deviation	0.77	0.65	1.32	0.9	1.15	0.86	0.83
Mean (by ChatGPT)	3.5	4.2	3.5	3.2	3.1	3.3	2.8

Processing of Modern Chinese Proverbs in High-Functioning Autistic Children

Ben Yue Shi yang
National Chun Cheng University, Taiwan

This study investigates the comprehension of Mandarin idioms in high-functioning autistic (HFA) children compared to typically developing (TD) children, addressing notable gaps in the literature regarding non-literal language processing in autistic populations. Although extensive research exists on idiom comprehension in autism, studies specifically targeting HFA children remain limited, with few focusing on cultural and linguistic contexts like Mandarin (Morsanyi & Stamenković, 2021). This research aims to fill these gaps by examining how HFA children understand idioms, especially metaphorical expressions involving human, animal, and daily-life subjects.

Participants included six HFA and six TD children from grades three to six, assessed through the Peabody Picture Vocabulary Test, False Belief Task, and Metaphorical Proverbial Story Test. The main experiment involved the Metaphorical Idiom Story Test, consisting of twenty-four short stories with proverbial or metaphorical endings. Comprehension was measured by asking participants to select from three images—literal, metaphor-related, or unrelated.

Results revealed that TD children significantly outperformed HFA children in vocabulary, particularly in sensory-emotional cognition. In the False Belief Task, HFA children showed slightly lower performance, likely due to differences in language development. In the Metaphorical Proverbial Story Test, TD children had a 90.65% correctness rate, compared to 75.45% for HFA children. The most pronounced deficit for HFA children was observed with animal metaphors, where they performed 25% worse than TD children, whereas the groups performed similarly on daily-life metaphors. These findings align with existing literature suggesting difficulties in processing figurative language among autistic individuals but highlight a particular challenge with metaphors that involve non-human subjects.

This study underscores the linguistic difficulties HFA children face, especially with non-human metaphors, suggesting that these children may struggle more with abstract and less familiar metaphorical content. The results emphasize the need for targeted interventions and specialized teaching strategies to support HFA children's language comprehension. Future research should explore the cognitive and developmental factors influencing idiom processing in HFA populations and assess the effectiveness of tailored educational approaches to improve metaphor comprehension.

Keywords: high-functioning autism, non-literal language, modern proverbs, cognitive linguistics, storytelling experiments

References:
Chen, Yuan Heng and Lin, Yu Hua (2008). A Theory of Mind Study (ToM) on Patients with High- Functioning Autisms and Asperger Syndromes. ***Taiwan Association of Clinical Psychology*.** National Cheng Chi University, Taipei, Taiwan.
Morsanyi, K., Stamenković, D. (2021). Idiom and proverb processing in autism: a systematic review and meta-analysis. ***Journal of Cultural Cognitive Science***, Volume 5, p.367–387.
Perner, J., Frith, U., Leslie, A. M., & Leekam, S. (1989). Exploration of the autistic child's theory of mind: knowledge, belief and communication. ***Child Development***, 60, 689-700.
Sandrine Le Sourn-Bissaoui, Marc Aguert, Pauline Girard, Claire Chevreuil, Virginie Laval. (2013). Emotional speech comprehension in children and adolescents with autism spectrum disorders, ***Journal of Communication Disorders***, Volume 46, Issue 4,309-320.

Table1: Example of Metaphorical Proverbial Story Test

Story line in Chinese	Story line translated into English	Question in Chinese	Question translated into English	Answer in Chinese	Answer in English
哥哥考試忘記帶筆，著急得像熱鍋上的螞蟻。	My brother forgot to bring a pen to the exam and was as anxious as an ant on a hot pan.	熱鍋上的螞蟻是什麼意思?	What does it mean by "as anxious as an ant on a hot pan" ?	很焦急	Very anxious

Competition and Coordination of Efforts: A Study of the Effectiveness of Symbols in Consecutive Interpreting Note-Taking

Yue Wu (wuyue8940@gmail.com), Qianxi Lv, Jingchao Yang (Shanghai Jiao Tong University, China)

Symbols are easier to write and read than language notes in consecutive interpreting (CI) [1]. Previous studies report a positive correlation between symbols and interpreting quality [2], but few have explicated whether all symbol categories are effective and the extent to which note-taking choices affect interpreting quality. We investigated the note-taking choices and symbol categories of 34 interpreting beginners performing CI tasks from Chinese to English. The notes were divided into word, abbreviation, and symbol forms, with symbols further categorized into VT, VGD, VMC, VR, G, CM, CV, and S [3] (Table 1). Target speech accuracy was rated at the element and the clause levels according to the source speech (Table 2). An *element* was an expression that could be noted down through symbols (e.g., 'country' recorded as □). A *Clause* consisted of a subject and its predicate, and it was the smallest independent grammatical unit to express an intended meaning.

The results showed that at the element level, symbols and full words contributed to significantly higher target-speech accuracy than abbreviations ($z=-6.173$, $p<0.001$), and significantly higher accuracy was achieved by using CM, CV, and S, compared to VT ($p<0.05$). These symbol categories were estimated to be 6.16 times, 4.20 times, and 5.22 times more effective for an accurate interpretation (Fig 1). At the clause level, symbols continued to demonstrate superior performance, while words became the least effective ($z=-2.492$, $p=0.013$). Meanwhile, only VR was estimated to be significantly more effective at facilitating accurate clauses compared to VT ($z=1.994$, $p=0.046$) (Fig 2).

In summary, the findings provided empirical support for the advantage of using symbols in memorization and speech reformulation. Additionally, the results validated the hypothesis that certain symbol categories, especially VR, CM, and S, could reduce processing demands, enhance the cooperation among Efforts, and facilitate target-speech accuracy. By using these symbols encapsulating logical meaning and larger processing units, the interpreters prioritize information analysis to note-taking and are thus more engaged in active listening.

Table 1. Note categorization

Form	Category	Definition		Tag	Example
Word		All notes represented as full words		Word	企业，Spain
Abbreviation		Part of the letters of a long English word/part of the characters of a long Chinese word		Abbreviation	"企" for "企业" "Sp" for "Spain"
Symbol	Vector	Arrows or graphic lines that represent	(i) the timeline of an event	VT	last year
			(ii) growth and decline	VGD	↑ increase
			(iii) a movement or consequence	VMC	← attract
			(iv) repeated information	VR	Europe
	Geometrical Shape	Pictorial representation of the (underlying) meaning of a word or expression		G	China
	Composition	Margin	A place in the left-hand margin for noting logical links	CM	∴ therefore
		Vertical List	A structure that visually classify information with equal importance	CV	US Fran Ita Spain America, France, Italy, Spain
	Salience	Things noted in an eye-catching way with bigger letter, underlying or special markers		S	! very important

Table 2. Assessment of accuracy

Checklist for rating		
Level	Criteria	Score
Element	Accurate rendition of the exact word or compositional structure that has been prescriptively annotated	0/1
Clause	1) No opposite meanings 2) Accurate rendition of main ideas 3) No unjustified change in meaning	0/1

Figure 1. The average **element-level** accuracy rate achieved by various symbols categories

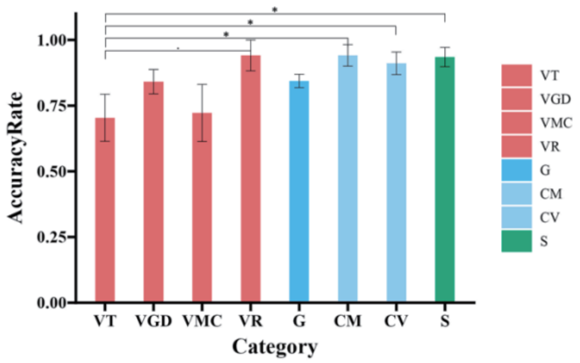
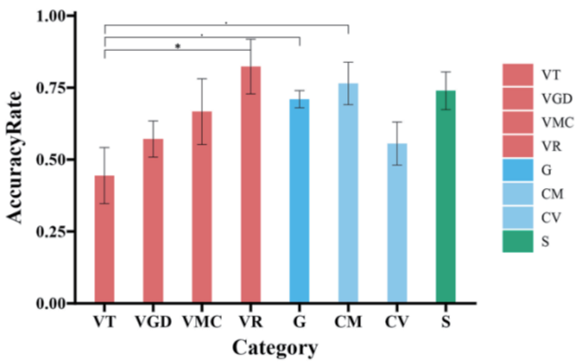


Figure 2. The average **clause-level** accuracy rate achieved by various symbols categories



References:

[1] Gillies, A. (2005). Note-taking for consecutive interpreting: A short course. St. Jerome.

[2] Chen, S. (2020). The process of note-taking in consecutive interpreting: A digital pen recording approach. Interpreting: international journal of research and practice in interpreting, 22(1), 117-139.

[3] Chang, L. (2014). Investigating note-taking in consecutive interpreting – Using the concept of visual grammar (Unpublished doctoral dissertation). University of Manchester

Co-activation of Semantic and Phonological Representations in Interpreting Learners

Qianxi Lv, Zhaokun Jiang, Yue Wu lvqianxi@sjtu.edu.cn

Individuals rely on mentally representing their combination of sounds, i.e., phonological representations [1] or readily evoking semantic representations [2] in memorization. Though phonological representations are more automatically activated, individuals can use semantic representation exclusively [3] or adaptively switch to semantic representations when the former is inhibited [4], such as under articulatory suppression (AS). It is also favored for bilingual tasks that requires meaning retrieval, such as interpreting. Forty Chinese-English bilinguals undergoing intensive interpreting training were asked to perform a shadowing or a simultaneous interpreting (SI) task, followed with a serial recall of English words(via the Word Length Effect (WLE), so as to examine how their memorization processes differ under requirement, and how their capacity to integrate semantic meaning facilitate retrieval in SI.

Exp1 investigated WLE in a serial recall task using a 2(Imaginability) × 2(word length) × 2(Inhibitory control) × 2(stimulation task) mixed design (Table1). It revealed significant effects of AS: $F_{(1, 362)}$ =6.602, p<.001; word length: $F_{(1, 362)}$=5.560, p<.001; AS × word length: $F_{(1, 362)}$= 2.135, p=.033, AS × word length × imaginability: $F_{(1, 362)}$ =2.018, p=.0443. WLE existed under AS ($F_{(1,38)}$=6.089,p<.001), and highly-imaginable words were memorized better for longer words($F_{(1,435)}$=5.560,p<.001). There is no significant effect of stimulating tasks (Fig1,2).

Exp2 investigated how individual's capacity for maintaining semantic representations affected WLE. Participants were asked to identify the synonym for one of the ten precedingly-studied words. Group differences were observed of capacity (Higher：Mean=9.15, SD= 1.040; Lower: Mean= 5.9, SD=1.071; t=-2.342, p=0.03). The lower-capacity participants show WLE under AS for highly-imaginable words only, whereas higher ones did not show WLE under silent, but did show WLE under AS, and word imaginability did not affect their performance (Fig3,4).

In conclusion, the findings suggest that participants exclusively use semantic representations to facilitate phonological reliance, instead of switching between the two. Moreover, the capacity of using semantic representations significantly affect the memorization behavior, such that people of higher capacity are more likely to commit to activating semantic concept. It also emphasizes the importance of developing strategies that foster the adaptive use of semantic representations in interpreting training.

Table1. Mean values for word imageability and length

	Short words			Long words		
	Imaginability	Syllables	Frequency	Imaginability	Syllables	Frequency
Low-imaginable	328.32	1.58	249.00	332.25	3.46	240.33
	(53.90)	(0.48)	(211.24)	(45.34)	(0.74)	(103.91)
High-imaginable	506.05	1.47	241.91	479.70	3.32	234.44
	(58.05)	(0.50)	(194.21)	(58.58)	(0.56)	(155.35)

The values in brackets indicate standard deviations.

Table2. Serial recall score of both groups

	Higher capacity		Lower capacity	
Control (Silent)				
High-imaginable	4.2	(0.837)	3.873	(1.071)
Low-imaginable	4.375	(0.518)	3.584	(1.077)
Articulatory Supression				
High-imaginable	4.000	(0.000)	3.400	(1.192)
Low-imaginable	2.667	(1.155)	3.009	(1.255)

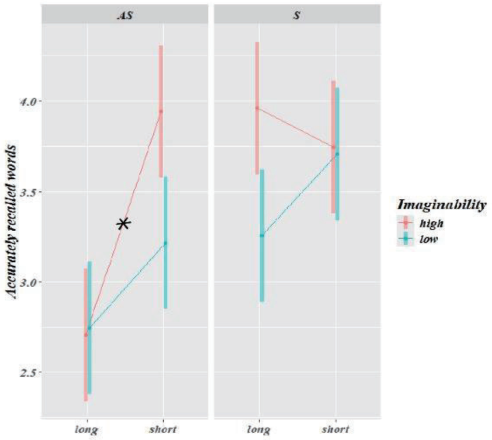


Figure1. Predicted score for the serial recall post-shadowing

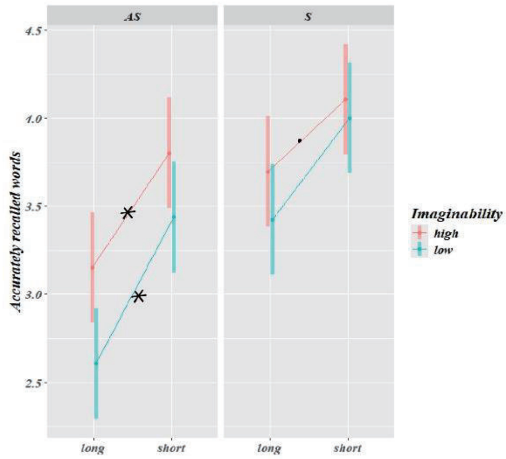


Figure2. Predicted score for the serial recall post- SI

AS: Articulatory suppression; S: Silent control; SI: simultaneous interpreting

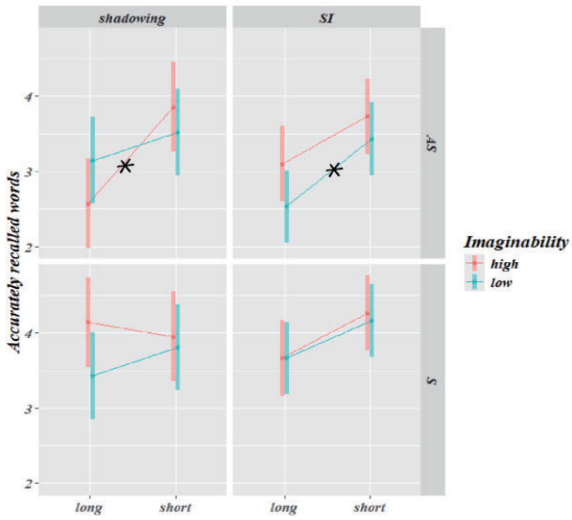


Figure 2. Predicted score for the serial recall post-shadowing & SI for participants of higher capacity

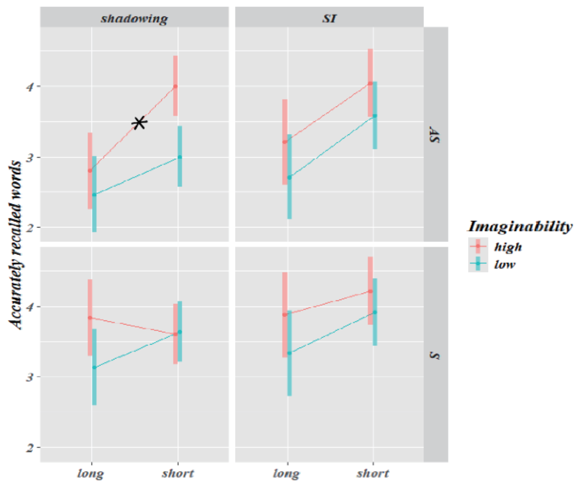


Figure 3. Predicted score for the serial recall post-shadowing & SI for participants of lower capacity

Reference:

[1] Baddeley, A. (1986). Working memory. Oxford: Oxford Univ. Press.

[2] Paivio, A. (1990). Mental representations: A dual coding approach. Oxford: Oxford Univ. Press.

[3] Martin, Randi C., Shelton, J. R., & Yaffee, L. S. (1994). Language processing and working memory: Neuropsychological evidence for separate phonological and semantic capacities. Journal of Memory and Language, 33(1), 83–111.

[4] Shivde, Geeta, & Anderson, M. C. (2011). On the existence of semantic working memory: Evidence for direct semantic maintenance. Journal of Experimental Psychology: Learning, Memory, and Cognition, 37(6), 1342–1370.

COMPOSITIONALITY IN LARGE LANGUAGE MODELS: EVIDENCE FROM CHINESE

Large language models (LLMs) segment text into strings of bytes called tokens, and they produce language by predicting what token comes next in a sequence. Tokens sometimes correspond to words, but many words are segmented into multiple tokens. For example, GPT-2, -3, and -4 segment *rhinoceros* into four tokens: *rh* + *in* + *oc* + *eros*. Subword tokenization supposedly allows LLMs to represent unfamiliar words as combinations of meaningful tokens[1, 2], but examples such as *rhinoceros*, comprising meaningless and misleading constituent tokens, seem to bely this claim. On the other hand, the apparently meaningless token *rh* recurs in words that are related to *rhinoceros*, such as *rhino* and *rhinoplasty*, and we recently found that across a wide variety of languages, there is at least some semantic information available in tokens[3]. Do LLMs extract this information, decomposing multi-token words to meaningful constituents, or do they memorize the combination of arbitrary tokens that spells out *rhinoceros*, for instance, like the letters of an alphabet, hindering their ability to interpret unfamiliar words as combinations of tokens?

Chinese characters present an opportunity to investigate this issue. Each character comprises three bytes, and GPT-2 segments almost all characters into multiple tokens (4,864 of the 4,895 characters in the Chinese lexical database[4]). Sub-character tokens, comprising one or two bytes, lack surface-level representations and never occur in isolation; they appear meaningless. But for historical reasons, characters that begin with the same two bytes share a semantic radical in about 80% of cases (as we show in the online supplement), so tokens can derive meaning from patterns in word form (as with *rh* in *rhino*). For example, the characters for "hit" and "hold" (打 and 把) share the "hand" radical (扌), and 打, 把, and 扌 begin with the same byte. Crucially, words that begin with same GPT-2 token share a semantic radical in about 37% of cases (compared to 2% as the at-chance baseline for sharing semantic radicals). For example, 打 and 扌 begin with same token, while 把 and 扌 do not (see Table 1). If LLMs decompose multi-token words to meaningful constituents, then constituent tokens that approximate semantic radicals should influence representations.

We therefore had GPT-2 identify semantic radicals in 4,895 Chinese characters. LLMs rank the likelihood that a token will come next in a sequence, and because GPT-2 is an open source model, the transformers Python package allows us to extract the rank of each token in each radical for each character in the context of **prompt A**; see page 2. GPT-2 has a vocabulary of 50,257 tokens, so ranks range from 1 to 50,257. We found, first, that GPT-2 ranks tokens in target radicals as more likely than in distractors. For example, GPT-2 ranks 扌 higher than 亻 in the context of 打 and 把. (As we show in the supplement, larger versions of GPT-2, with more parameters, rank tokens in target radicals as more likely than the base model.) Second, crucially, performance at identifying semantic radicals improves when the first token in the target radical is the same as the first token in the corresponding character. For example, GPT-2 ranks 扌 higher in the context of 打 (both beginning with token ID 33699) than the context of 把 (beginning with token 162). This suggests that GPT-2 is sensitive to the identity of sub-character tokens and that it associates the meanings of characters which share tokens, as implied by compositionality. Third and finally, we investigated whether the higher ranks for tokens which are identical to the first tokens in characters are a byproduct of GPT-2 predicting the character rather than the radical. When predicting the second token in the target radical, following prompt A plus the first token in the target radical / character, the second token in the radical is ranked substantially more likely than the second token in the character. For example, following prompt A plus token ID 33699, GPT-2 ranks token ID 234 (from 扌) higher than token ID 241 (from 打). See Figure 1. All differences are highly significant and hold when including the frequency and family size of characters and semantic radicals as covariates in linear regression models.

Prominent AI companies assume that LLMs decompose multi-token meanings to meaningful constituents[2]. Such compositionality is essential to processing the indefinite variety of words that LLMs may encounter, but little evidence supports this assumption. We have shown that GPT-2 does extract information from tokens which appear meaningless.

Prompt A
汉字包含语义部首。例如，「他」字包含「亻」,而「CHARACTER」字包含「RADICAL
English gloss: *Chinese characters contain semantic radicals. For example, the "he" character contains "person", and the "CHARACTER" character contains "RADICAL*

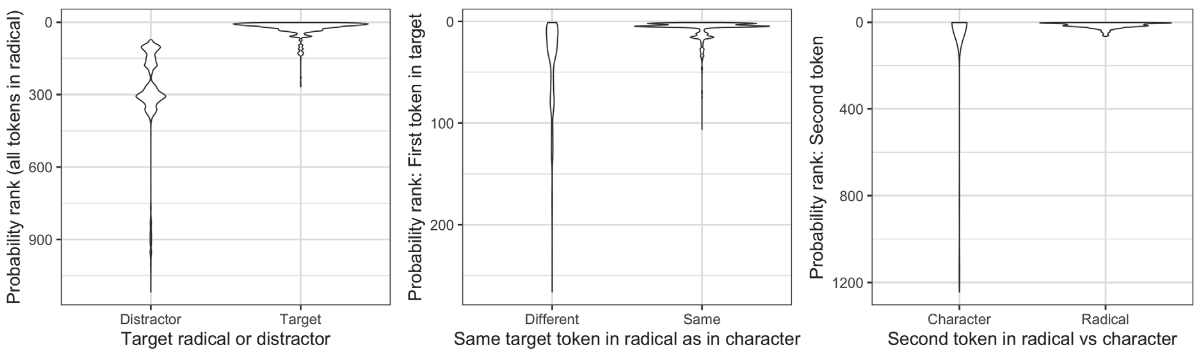
Online supplement (anonymized):
https://osf.io/c6bvm/?view_only=67b4d81ff8e84dc39c0d9016543d97a4

Table 1: Chinese characters as bytes (hexadecimal digits) and GPT-2 tokens

Character	Bytes	Token IDs	Radical	Bytes	Token IDs
打 (hit)	e6 89 93	33699, 241	扌(hand)	e6 89 8c	33699, 234
把 (hold)	e6 8a 8a	162, 232, 232	扌(hand)	e6 89 8c	33699, 234
手 (hand)	e6 89 8b	33699, 233	手 (hand)	e6 89 8b	33699, 233
他 (he)	e4 bb 96	19526, 254	亻(person)	e4 ba bb	12859, 119
佬 (guy)	e4 bd ac	19526, 105	亻(person)	e4 ba bb	12859, 119
人 (person)	e4 ba ba	21689	人 (person)	e4 ba ba	21689

Note: GPT-2 token IDs less than 256 are single bytes. Larger Token IDs are two or three bytes. Different representations of the same radical (e.g.,亻 and 人) have different byte-level representations and different tokens. We used the semantic radicals identified by [4].

Figure 1: **Left panel**: rank of next-token likelihood for target radicals versus distractors (e.g., 扌 versus 亻 in the context of 打 or 把; averaging all tokens per radical, following prompt A plus any preceding tokens in that radical). **Middle panel**: rank of next-token likelihood for target radicals that have the same versus a different first token as the corresponding character (e.g., 扌 in the context of 打 versus 把; first token only, following prompt A). **Right panel**: rank of next-token likelihood for the second token in radicals versus characters that begin with the same token (e.g., the second token of 扌, ID # 234, versus 打, ID # 241, following ID # 33699).



References
[1] Sennrich et al. (2016). Neural machine translation of rare words with subword units. *ACL 2016.* arXiv:1508.07909
[2] Wolf et al. (2020). HuggingFace's Transformers: State-of-the-art Natural Language Processing. huggingface.co/docs/transformers
[3] Anonymized for peer review.
[4] Sun et al. (2018). Chinese lexical database (CLD): A large-scale lexical database for simplified Mandarin Chinese. *Behavior Research Methods*.

PS.2.02

How does ChatGPT interpret implausible sentences?

Fei Yuan, The Chinese University of Hong Kong. E-mail: 1155195616@link.cuhk.edu.hk

Zhenguang Cai, The Chinese University of Hong Kong.

Xufeng Duan, The Chinese University of Hong Kong.

Abstract The growing interest in psycholinguistics has led researchers to question whether large language models (LLMs), such as ChatGPT, truly resemble human language competence. Despite the clear evidence of LLMs' remarkable language comprehension and production abilities (Cai et al., 2023; Hu et al., 2022; Mahowald et al., 2023), it remains unclear how they achieve these tasks and if their processing resembles human cognitive processes.

Results from Cai et al. (2023) showed that ChatGPT can reinterpret implausible sentences that were likely to have been corrupted by noise with plausible alternatives, but how? Current explanations for humans' ability to derive plausible meanings from implausible sentences are grounded in the noisy channel account (Gibson et al., 2013, 2017) and the plausibility-driven syntactic prediction account (Cai et al., 2022).

In the current research, we investigated whether ChatGPT resembles humans in its ability to interpret implausible sentences into plausible alternatives. In particular, we examined (1) whether ChatGPT can adapt the noise rate when communicating with individuals from diverse language backgrounds and (2) whether ChatGPT can arrive at the plausible meaning of an implausible sentence by either revising the syntactic structure.

Two experiments were conducted on ChatGPT to address the questions above. In experiment 1, ChatGPT was presented with an instruction including the profile of the speaker (native or non-native), then one sentence (plausible or implausible; in three pairs of structures: DO-PO; transitive-intransitive; active-passive) followed by a comprehension question. In experiment 2, we presented ChatGPT with prime sentences (plausible or implausible; DO or PO), followed by a comprehension question and a preamble completion task. The answers to the comprehension question would show how ChatGPT interprets the sentence (literal or not) and the preamble completion responses were collected to calculate the structural priming effect.

For Experiment 1, we conducted mixed-effects modeling on the trial-level responses (literal vs. nonliteral interpretation), with nativeness and structure as predictors. The results showed a lower rate of literal interpretation in the non-native than the native condition for transitive-intransitive structures and active-passive structures but not for the DO-PO structures. There is also an effect of structure for the DO-PO structures and transitive-intransitive structures (with more literal interpretations for PO/transitive than for DO/intransitive but not for the active-passive structures). These findings partially align with Gibson et al. (2017) in a similar study involving human participants.

For Experiment 2, the results showed that ChatGPT had a higher tendency for non-literal interpretations than literal interpretations when reading implausible sentences. The analysis of structural priming on preamble completions revealed a significant main effect of the syntactic structure and plausibility of the prime sentence (with more DO responses following a DO/plausible prime than PO/implausible prime) and a significant interaction effect between plausibility and structure (see Figure.2). However, the main effect of the interpretation of the prime sentence and the interaction effect between interpretation and structure were missing, indicating that there was no difference in the priming effect of an implausible prime regardless of how the implausible prime was interpreted. Furthermore, a significant interaction effect between structure and plausibility was observed when focusing solely on implausible primes, indicating that ChatGPT conducted non-literal syntactic analysis even when interpreting them literally. Therefore, this evidence supports the plausibility-driven syntactic prediction account (Cai et al., 2022) and provides additional support for the presence of syntax-like representations in LLMs, suggesting that such a syntactic layer operates independently from the intertwined semantic information, which contradicts the claim made by Piantadosi (2023).

Appendix A

1. System prompt of Exp.1

Native condition	Non-native condition
Hi, I am a native English speaker from the USA. I am now living in New York and studying for a BA degree at the City University of New York. I'd like to play a sentence comprehension game with you. I will give you a sentence and a yes-or-no question regarding the sentence. Please simply answer "Yes" or "No" to the question. For instance, if you read "The boy went to the park to fly a kite", and the question is "Did the boy go to school", you can answer "No". Next, read the following sentence and answer the question.	*Hi, I am a non-native speaker of English from China. I am now living in Shanghai and studying for a BA degree at the Shanghai Jiaotong University. I'd like to play a sentence comprehension game with you. I will give you a sentence and a yes-or-no question regarding the sentence. Please simply answer "Yes" or "No" to the question. For instance, if you read "The boy went to the park to fly a kite", and the question is "Did the boy go to school", you can answer "No". Next, read the following sentence and answer the question.*

2. Exemplar sentences and comprehension questions:

The girl kicked the ball. Did the girl kick something/someone?
The sister mailed the niece the letter. Did the niece receive something/someone?
The businessman benefited the tax law. Did the tax law benefit from anything?

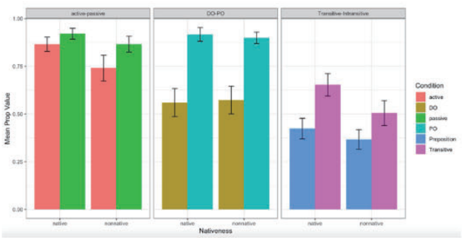
3. System prompt of Exp.2

I'd like to play a sentence comprehension game and a sentence completion game with you. I will first give you a sentence and a yes-or-no question regarding the sentence. Please simply answer "Yes" or "No" to the question. Then repeat the preamble and continue it into a full sentence. For instance, if you read "The boy went to the park to fly a kite", and the question is "Did the boy go to school", you can answer "No". If you then see the preamble "The man was sick …". You can repeat and continue the preamble as "The man was sick and did not go to work".

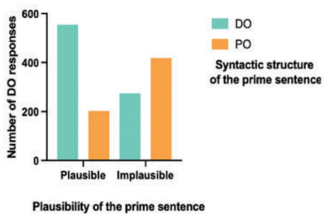
Next, I am going to give you a sentence and a comprehension question.
XXXX
XXX

Next, I am going to give you a preamble to repeat and continue.
XXX

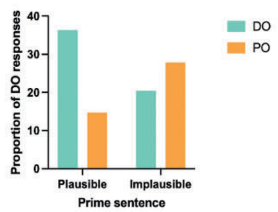
Appendix B



1



2



3

Figure 1. *Proportion of "No" answers (adopting a literal interpretation) when GPT read an implausible sentence in different structures*; Figure 2. *The results of preamble completions;* Figure 3. *The results of preamble completion comparisons after literal interpretations of plausible and implausible prime sentences*

References

[1] Cai, Z. G. et al (2023). PsyArXiv. [2] Cai, Z. G. (2022). *Psychonomic Bulletin & Review*, *29*(3), 1026–1034. [3] Cai, Z. G. et al (2022). *Cognition*, *225*, 105101. [4] Gibson, E. et al (2013). *Proceedings of the National Academy of Sciences*, *110*(20), 8051–8056. [5] Gibson, E. et al (2017). *Psychological Science*, *28*(6), 703–712. [6] Hu, J. et al (2023). arXiv. [7] Mahowald, K. et al (2023). arXiv.

PS.2.03

A perspective from Bayesian learning on unaccusativity

Kaiying Kevin Lin (Institute of Linguistics, Academia Sinica, Taiwan)

Unaccusativity, or split intransitivity refers to the binary distinction of unergative and unaccusative verbs of intransitive verbs (Permultter, 1978), while later studies found that the division of intransitive verbs in languages like Mandarin (e.g., Laws & Yuan, 2010) exhibits semantic gradients similar to the proposal by Sorace (2000). This raises the controversy over whether the universal syntactic subclasses should be binary (e.g., Fukuda, 2017) or inherently multicategorical with each semantic subclass mapping to a syntactic subclass (cf. Huang & Snedeker, 2020). We conduct a language-learning simulation using the Bayesian inference model because the model contains a prior belief term and can examine these assumptions (Shin, 2023). The results show that when the prior belief is solely based on two syntactic subclasses of a few verbs, the ultimate posterior distribution is the most similar to the gradient of Sorace (2000). This supports the existence of universal syntactic subclasses of unergative and unaccusative, but different language environments may shape representations that appear similar to semantic gradients.

Experiment: We trained Bayesian inference models with two different priors: (1) binary verb subclasses of unergative and unaccusative verbs (prior=0.5), and (2) two verb subclasses plus an additional neutral category (three categories, prior=0.33). We computed the word occurrences in sentences from the CHILDES Mandarin corpus (MacWhinney, 2000) to obtain transitional probabilities (i.e., the likelihood of one word following another, based on their co-occurrence in text). The three most typical verbs from each category (or class) were selected and their transitional probabilities in sentences where these verbs occurred were computed in each condition. These formed the typical unergative, unaccusative (and neutral) transitional probabilities. The likelihood of a sentence belonging to each verb category was calculated by summing log transitional probabilities using each typical transitional probability, and posterior probabilities of a sentence in each category were thus computed and normalized. We calculated the posteriors of verbs in the postverbal-subject diagnostic sentence (Huang, 1987), as well as the posteriors of new verbs in certain constructions (Table 2). The former aims to simulate likely rating speakers assign to verbs within the diagnostic, while the latter simulates the effect of transitional cues on new verb categorization.

Results: The posterior probabilities from verb categories were converted into four sets of numerical values for comparison with two previous studies. We found that using two verb classes as the prior yields the most similar distribution (Table 1, Student t-test, $t(23)=0.7162$, $p=0.48$), while the model using three categories yields a significantly different distribution from Laws & Yuan (2010) (Student t-test, $t(23)=3.479$, $p<.01$**). Moreover, the effect of transitional cues aligns better with previous research (Lin & Washington, 2023), when two verb classes of priors were used (Table 2, $t(8)=1.2011$, $p=0.1302$ vs $t(8)= -1.4587$, $p=.09$.). These findings indicate that even in a language that exhibit semantic gradients, the underlying system might consist of only two syntactic subclasses of unergative and unaccusative verbs. However, exposure to language-specific factors, such as transitional probabilities, may contribute to a semantic gradient representation beyond a simple binary distinction.

(wordcount: 496)

Table 1 Posteriors in our study converted into a range of -2 (unerg) to 2 (unacc) for comparison with Laws and Yuan (2010)

Semantic Subclasses	verb	Laws & Yuan (2010)	Posteriors (two class)	Posteriors (three categories)
change-of-location	come	1.97	1.75	1.749
	fall_1	1.87	1.32	0.5445
	escape	1.67	-0.17	-0.002
	arrive	1.57	0.03	0.023
	fall_2	0.63	0.71	0.2423
change-of-state	die	1.7	1.62	1.614
	born	1.57	1.95	0.152
	appear	1.7	1.51	0.8324
	happen	1.73	1.94	0.6218
Internally caused change	decay/rot	1.5	1.96	0.8317
	grow	1.57	0.9	-0.5544
	wilt	0.53	1.15	0.9362
change of directed motion	descend	1.87	1.92	1.17
	rise	1.83	0.36	0.0138
continuation of condition (inanimate)	continue	-0.3	1.15	0.5057
	persist	-0.3	-0.15	-0.096
	remain	-0.27	1.6	0.145
continuation of condition (animate)	remain	-0.57	0.029	0.003
	survive	0.8	-0.15	-0.018
	exist, have	1.8	-0.15	-0.083
existence of posture (inanimate)	sit	-0.07	-0.17	-0.07
	stand	0.17	-0.15	-0.1
existence of posture (animate)	sit	0.63	-0.29	-0.18
	stand	0.53	1.97	1.05
t-test		N/A		*

Selected Reference:

Fukuda, S. (2017). Split intransitivity in Japanese is syntactic: Evidence for the Unaccusative Hypothesis from sentence acceptability and truth value judgment experiments. *Glossa: a journal of general linguistics* 2(1): 83. 1–41, DOI: https://doi.org/10.5334/gjgl.268

Huang, Y., & Snedeker, J. (2020). Evidence from the visual world paradigm raises questions about unaccusativity and growth curve analyses. *Cognition* Jul;200:104251. doi: 10.1016/j.cognition.2020.104251.

Laws, J., & Yuan, B. (2010). Is the core-peripheral distinction for unaccusative verbs cross-linguistically consistent? Empirical evidence from Mandarin. *Chinese Language and Discourse*, 1(2), 220–263. DOI: 10.1075/cld.1.2.03law.

Lin, K., & Washington, P. Y. (2023). The categorization of intransitive verbs: Evidence from word2vec modeling and a behavioral experiment. In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), Proceedings of the 45th Annual Meeting of the Cognitive Science Society (pp.1990-1997).

Shin, G.-H., & Mun, S. (2023). Korean-speaking children's constructional knowledge about a transitive event: Corpus analysis and Bayesian modelling. *Journal of Child Language*, *50*(2), 311–337. doi:10.1017/S030500092100088XSo

race, A. (2000). Gradients in auxiliary selection with intransitive verbs. Language, 76(4), 859–890. DOI: doi.org/10.2307/417202

Table 2, Posteriors converted into a range of -1 to 1 for comparison with Lin and Washington (2023)

Construction	ba+V	V1+V	V+perf	hui	LVS
Lin & Washington (2023)	0.5	0.4444	0.53	0.4444	0.5
Posteriors (two classes)	0.59	0.5738	0.4959	0.4844	0.4617
Posteriors (three categories)	0.5	0.5	0.5	0.49	0.475
Construction	want+V	V+dur	V+complement	imperf+V	t-test
Lin & Washington (2023)	0.3111	0.4333	0.5	0.33	
Posteriors (two classes)	0.45	0.442	0.4145	0.36	0.13
Posteriors(three categories)	0.49	0.4601	0.4464	0.49	0.09.

PS.2.04

Bridging the Gap: Assessing GPT-4's Capacity for Causal Inference

Xingyuan Zhao (xiz103@ucsd.edu), Wenjun Ma & Seana Coulson, UC San Diego

Language comprehension frequently requires readers to infer the causal relationship between events in a text. Here we compare human comprehenders' sensitivity to causal information in text to that of the large language model GPT-4.

Dataset The dataset (Davenport, 2014) contains 160 written scenarios, each followed by a probe word that was either causally related to scenario, lexically related to the final word in the scenario, or unrelated. Scenarios consisted of two sentences; the first set the scene and the second described a subsequent event that resulted from an unstated cause. The three categories of probe words were based on their relationship to the narrative (see Table 1).

Expt.1 GPT-4 classified probe words for all scenarios and provides explanations to its answers. Binary logistic regression on classification results indicated exceptional performance on causal probes (97.5% accuracy) and unrelated probes (92.1%, $\beta = -1.20$, 95% CI [-2.27, -0.12], $p < .05$), with significant drops for lexical (30.1%, $\beta = -4.48$, 95% CI [-5.53, -3.43], $p < .001$). This suggests that LLMs like GPT-4 'appreciate' causal relationships presented in text. **Expt.2** 174 native English speakers classified the relationship between 40 randomly assigned narrative scenarios and associated probe words as either causal, lexical, or unrelated. Binary mixed-effects logistic regression analyses showed highest accuracy for causal probes (87.3%), with significantly lower accuracy for lexical (75.2%, $\beta = -1.14$, 95% CI [-1.62, -0.65], $p < .001$) and unrelated probes (79.8%, $\beta = -0.54$, 95% CI [-0.97, -0.10], $p < .05$). This confirms that GPT-4 performs equally, or even outperforms, human, on causality, but failed to detect lexical relation to a human-level. Then they rated the adequacy of GPT-4's explanations for its probe classifications in those scenarios on a scale from 1 (not convincing at all) to 5 (extremely convincing). Linear mixed-effects regression showed a significant positive effect on GPT-4 explanations for correct causal probes (rating: 4.58; Figure 1), indicating that explanations accompanying correct causal classifications were particularly well-received.

The results showed that GPT-4 appears to have a robust understanding of causality, as indicated by its near-perfect performance in identifying causally related probes and the high endorsement by humans of its explanations for the choices made. Notably, GPT-4's performance on the Lexical probes was below the chance level. Optimized for generating coherent and contextually relevant text, GPT-4 may lack the ability to explicitly identify lexical relationships. This focus on broader context and coherence in text generation might have limited its ability to perform tasks that may require ignoring higher levels of context.

Figures and Tables

True Relation	Scenario	Probe	GPT4 Answer	GPT4 Explanation
Causal	Anna stayed in bed all morning and afternoon. She was drinking hot soup and resting.	Cold	Causal ✓	The probe word "cold" is likely related to the reason why Anna was staying in bed, drinking hot soup, and resting, as these are common responses to having a cold.
Lexical	Alphonse made the grave mistake of insulting a gentleman. The next morning, the bullet pierced his heart from a distance of 10 paces.	Steps	Lexical ✓	The probe word "steps" is lexically related to the narrative because "paces" is a synonym for "steps," but it does not indicate a cause for the event.
Unrelated	The matador stepped into the arena. He had to be taken away in an ambulance.	Crab	Unrelated ✓	The word "crab" has no causal or lexical relation to the narrative about a matador being taken away in an ambulance.
Causal	Robert was flirting with an attractive woman he'd met in the bar. but when she saw his hands, she left him with a stinging slap.	Ring	Unrelated ×	The probe word "ring" does not have a causal relationship with the event of the woman slapping Robert, nor is it lexically related to the last word "slap." The narrative does not provide any information about a ring being involved in the interaction.
Lexical	The burglar broke the window and entered the house. Moments later, he was running for his life.	Death	Unrelated ×	The probe word "death" does not have a causal relationship with the burglar running for his life, nor is it lexically related to any words in the second sentence.
Unrelated	The clouds gathered over the dry, dead field. The next morning, the field was turning green.	Repair	Causal ×	The gathering of clouds likely indicates that it rained, which would be the cause of the dry, dead field turning green.

Table 1: Sample vignettes with probe words and its' relation(left) & GPT-4's Classifications and Explanations of Scenarios(right)

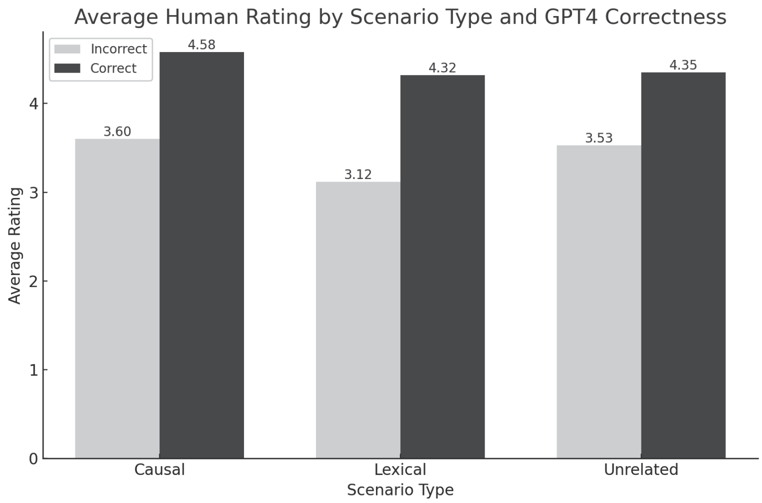


Figure 1: Average Human Ratings by Scenario Separated by GPT-4 Correctness

References

Davenport, T. S. (2014). Causal inference and language comprehension: Event-related potential investigations (Doctoral dissertation). UC San Diego.

PS.2.05

The source of pragmatic knowledge: A computational investigation

Wenjun Ma (w7ma@ucsd.edu), Xingyuan Zhao & Robert Kluender, UC San Diego

Ambiguity is a pervasive feature of human language, but is usually resolved through the use of discourse context and/or pragmatic knowledge (Piantadosi et al., 2012). The excellent performance of recently developed large language models (LLMs) raises the question of whether they are able to resolve ambiguity without direct access to such knowledge. Here we test whether contextual plausibility, informed by real-world knowledge, can be reduced to the frequency of linguistic co-occurrence. The use of the Mandarin Chinese coverb '给' (GEI) offers a useful test case for this question. Traditionally used to mark datives (recipients/benefactives), GEI has since been extended to mark agents in passive constructions (= coverb BEI) as well as patients in object pre-posing constructions (= coverb BA; Simon, 1958). 吴一给李四打了 [Wuyi-GEI-Lisi-hit-PFV] could thus mean either 'Wuyi hit Lisi' or 'Lisi hit Wuyi.' However, in real-life scenarios, pragmatic knowledge often renders one interpretation impossible. In this study, we investigated whether LLMs can accurately discern such nuances in usage, based on pragmatic knowledge, of the Mandarin coverb '给' to mark both agents and patients. We also explore whether this ability can be reduced to differences in surprisal, the unpredictability of words given their context, with higher surprisal indicating lower probability of occurrence in the training data for LLMs.

In a classification task, we presented GPT-4 with 52 sets of '给' sentences (104 sentences total; Table 1), in which '给' was used as either an agent or a patient marker. In half the sets, sentences were constructed with one animate and one inanimate argument, rendering one interpretation 0% possible (Type 1). In the other half, sentences had two animate arguments, with one interpretation more likely (Type 2). Compared to human classification, GPT-4 achieved a remarkable overall accuracy rate of 98.1%, with 100% accuracy for Type 1 sentences and 96.2% for Type 2 sentences, suggesting that it registers pragmatic information like animacy or asymmetry of relations (cat-mouse) quite accurately.

Pragmatic plausibility is derived from familiarity with an event, and this may be reflected in linguistic co-occurrence in a corpus, which surprisal should be able to evaluate. We conducted a further experiment testing the power of surprisal to function as an indicator of pragmatic knowledge. The more plausible usage of GEI in the training data may also be the more frequent one (frequency of the linguistic co-occurrence), with lower surprisal values. Due to the unavailability of a surprisal API in GPT-4, we instead used surprisal values from Davinci-002, a GPT-3 model, which is only trained on linguistic data without reinforcement learning from human feedback. After changing GEI to BEI or BA in the stimuli, and calculating surprisal for each variant (see formula 1), the correctly switched variant had lower surprisal 76% of the time, significantly better than chance ($Z = 7.5$ $p < 0.001$). This suggests that mere linguistic co-occurrence may play a key role in acquiring pragmatic knowledge.

Figures and Tables

Type	Sentence	Sentence Translation	Marker
1	兔子给萝卜吃光了。	rabbit-GEI-radish-eat-PFV	Patient
1	萝卜给兔子吃光了。	radish-GEI-rabbit-eat-PFV	Agent
2	警察给小偷抓住了。	police-GEI-thief-catch-PFV	Patient
2	小偷给警察抓住了。	thief-GEI-police-catch-PFV	Agent

Table 1: Sample sentences used in Expt. 1

Coverb	Plausibility	Sentence	Sentence Translation
BEI	Yes	萝卜被兔子吃光了。	radish-BEI-rabbit-eat-PFV
BEI	No	兔子被萝卜吃光了。	rabbit-BEI-radish-eat-PFV
BA	Yes	兔子把萝卜吃光了。	rabbit-BA-radish-eat-PFV
BA	No	萝卜把兔子吃光了。	radish-BA-rabbit-eat-PFV

Table 2: Sample sentences used for the surprisal calculation in Expt. 2

Formula 1: Sentence Surprisal Calculation for Sentence 'rabbit-BA-radish-eat-PFV'

$$Surprisal(rabbit) = -\ log(P(rabbit))$$
$$Surprisal(BA) = -\ log(P(BA \mid rabbit))$$
$$Surprisal(radish) = -\ log(P(radish \mid rabbit\ BA))$$
$$Surprisal(eat) = -\ log(P(eat \mid rabbit\ BA\ radish))$$
$$Surprisal(PFV) = -\ log(P(PFV \mid rabbit\ BA\ radish\ eat))$$

$$Sentence\ Surprisal = \frac{Surprisal(rabbit) + Surprisal(BA) + Surprisal(radish) + Surprisal(eat) + Surprisal(PFV)}{5}$$

References

Michaelov, J. A., Coulson, S., & Bergen, B. K. (2023). Can Peanuts Fall in Love with Distributional Semantics?. arXiv preprint arXiv:2301.08731.

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. Cognition, 122(3), 280-291.

Simon, H. F. (1958). Some remarks on the structure of the verb complex in Standard Chinese. *Bulletin of the School of Oriental and African Studie*s, 21(3), 553-557.

Typical active transitive sentences in Mandarin are in SVO word order

警察 抓住 了 小偷
jingcha zhazhu le xiaotou
police catch PFV thief
‘The police caught the thief.’ (PFV = perfective aspectual marker)

Coverbs in Mandarin Chinese

Coverbs in Mandarin Chinese typically occur after the subject {e.g. NP coverb NP…} and mark the thematic relation of the following argument NP to the verb

- BEI (被) marks **agents** in passive constructions
  小偷 被 警察 抓住 了
  xiaotou BEI jingcha zhazhu le
  thief BEI police catch PFV
  ‘The thief was caught by the police.’

- BA (把) marks **patients** in object-preposing constructions
  警察 把 小偷 抓住 了
  jingcha BA xiaotou zhazhu le
  police BA thief catch PFV
  ‘The police caught the thief.’ (~‘It was the thief that the police caught.’)

- GEI (给) (‘give’) marks **recipients & benefactives** in indirect object position
  小王 给 我 买 了 一 本 书
  XiaoWang gei wo mai le yi ben shu
  XiaoWang GEI me buy PFV one CL book
  ‘Xiao Wang bought a book for me’

More recently, GEI has been extended as a substitute for both BEI and BA coverbs

- GEI can mark **agents** (earlier development in all Mandarin dialects)
  车 给 小王 修 好 了
  che GEI XiaoWang xiu hao le
  car GEI XiaoWang repair good-RSLT PFV
  ‘The car was repaired by Xiao Wang.’ / #The car repaired XiaoWang.

- GEI can also mark **patients** (more recent extension in certain contexts and dialects)
  小王 给 车 修 好 了
  XiaoWang GEI che xiu hao le
  XiaoWang GEI car repair good-RSLT PFV
  ‘XiaoWang repaired the car.’ / #XiaoWang was repaired by the car.

PS.2.06

Distinguishing ChatGPT-Generated Translation from Neural Machine Translation and Human Translation: A Linguistic and Statistical Approach

Zhaokun Jiang and Qianxi Lv, Shanghai Jiao Tong University | E-mail: zhaokun_jiang@sjtu.edu.cn

There is growing interest in the interplay between AI-powered translation tools and human translators [1, 2]. Particularly, ChatGPT, a prominent large language model (LLM), has demonstrated strong performance in translation tasks, rivaling mainstream neural machine translation (NMT) systems like Google Translate [3]. While existing literature has compared ChatGPT's translation quality to NMT [4, 5], few studies have delved into the nuanced linguistic and stylistic features that distinguish ChatGPT-generated translations from both NMT and human translation (HT).

The present study aims to bridge this gap by exploring the separability of ChatGPT-generated translations from NMT and HT, and the distinctive linguistic patterns across the three translation varieties. We employed a mixed-method approach combining computational techniques and detailed textual analysis. First, we created a corpus of Chinese Spokesperson's Remarks and their official English translations. The same source texts were also translated by Google Translate and ChatGPT respectively at the textual level. We then extracted 112 linguistic features to conduct three-way classification tasks using five classifiers, perform multidimensional analysis [6], and compute Euclidean distances.

The classification tasks revealed a clear distinction among the three translation varieties, with high accuracy across the five classifiers. However, there were instances where NMT and ChatGPT translations were misclassified, while HT was consistently classified correctly (see Table 1). The multidimensional analysis revealed that, across the identified dimensions, ChatGPT's translations were more similar with NMT in terms of the patterns of linguistic features (see Fig. 1). Notably, colloquial and conversational expressions were more prevalent in ChatGPT-generated translations compared to NMT and HT (see Example 1). The computed distances further confirmed that ChatGPT-generated translations were closer to NMT, while HT was more distant from the two (see Fig. 2). This may be attributed to the superiority of human translators in idiomaticity, the use of translation strategies, and conveying register [4].

Table 1 Results of machine learning classifiers

Classifiers	Accuracy	Class	Precision	Recall
		ChatGPT	0.91	1.00
Linear SVM	0.97	HT	1.00	1.00
		NMT	1.00	0.91
		ChatGPT	1.00	1.00
Random Forest	1.00	HT	1.00	1.00
		NMT	1.00	1.00
		ChatGPT	1.00	1.00
MLP	1.00	HT	1.00	1.00
		NMT	1.00	1.00
		ChatGPT	1.00	1.00
AdaBoost	0.97	HT	1.00	0.89
		NMT	0.92	1.00
		ChatGPT	0.77	1.00
Naïve Bayes	0.90	HT	1.00	1.00
		NMT	1.00	0.73

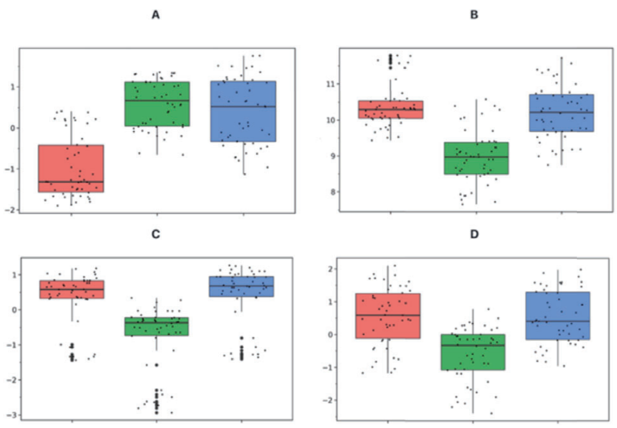


Figure 1 z-transformed scores of translations by ChatGPT (red), human translation (green), and NMT (blue) along four dimensions identified in the multidimensional analysis

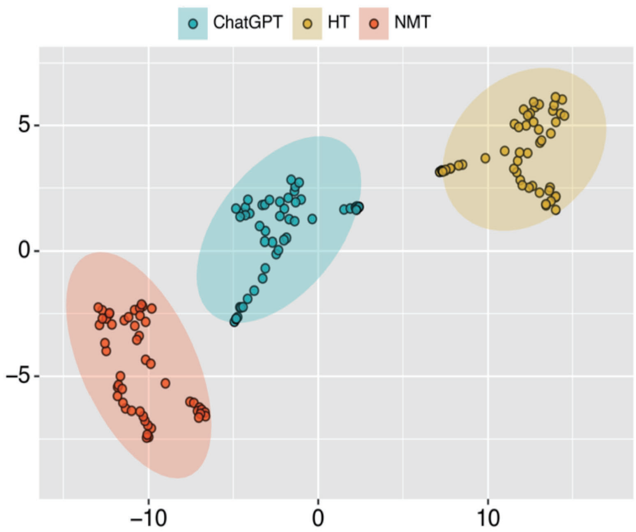


Figure 2 t-SNE visualization of distance distributions

Example 1:

HT: It is reported that Indian Prime Minister Narendra Modi visited the so-called "Arunachal Pradesh" on February 9th. […] It has been verified **that** there were eight Chinese citizens on board, including one from the Hong Kong SAR.

NMT: It is hoped that all parties in Myanmar will proceed from the fundamental and long-term interests of the country and the nation, resolve emerging problems by peaceful means under the constitutional and legal framework, and continue to advance the process of democratic transition in the country in an orderly manner.

ChatGPT: Really afraid of chaos in the world! We can't help but ask these lawmakers, are you "legislators" or "lawbreakers"? […] You better mind your own business. Hong Kong doesn't need you to worry about it.

References:

[1] Hu, K., Li, X. (2023) 'The creativity and limitations of AI neural machine translation A corpus-based study of DeepL's English-to-Chinese translation of Shakespeare's plays', *Babel*, 69 (4): 546-563.

[2] Lee, T. K. (2023) 'Artificial intelligence and posthumanist translation: ChatGPT versus the translator', *Applied Linguistics Review.*

[3] Gaspari F, Almaghout H, and Doherty S. (2015) 'A survey of machine translation competences: Insights for translation technology educators and practitioners', *Perspectives*, 23(3): 333–358.

[4] Karpinska, M., and Iyyer, M. (2023) 'Large language models effectively leverage document-level context for literary translation, but critical errors persist', In *Proceedings of the Eighth Conference on Machine Translation (WMT23)*, Singapore, 6-7 December 2019, pp.419–451.

[5] Hendy, A., Abdelrehim, M.G., Sharaf, A., Raunak, V., Gabr, M., Matsushita, H., Kim, Y., Afify, M. and Awadalla, H.H. (2023) 'How good are GPT models at machine translation? A comprehensive evaluation', *arXiv*: 2302.09210[Cs].

[6] Biber, D. (1988). *Variation across Speech and Writing*. Cambridge: Cambridge University Press.

PS.2.07

To see is to know but what if we just look?

Serene Siow[1], Rachel Dudley[2] and Nick Huang[1]   [1] National University of Singapore; [2] Central European University

English *see* and Mandarin *kan* are verbs of visual perception that also can be used to describe beliefs (1-2). When used to describe beliefs, *see* (and other visual verbs like *notice*/*observe*) is factive, presupposing the truth of its complement – "Meg doesn't see that Sally likes apples" presupposes that Sally indeed likes apples (2). How might learners learn such an abstract semantic property? It has been suggested that belief uses of *see* may be learnt through a metaphor where seeing an event is equivalent to knowing an event happened [1-3]. It is not surprising, then, that *see* is factive like *know*. However, unlike English *see*, Mandarin *kan* is non-factive (*e.g.*, [4]). One important difference between *see* and *kan* is that *kan* in the visual sense can be paraphrased as *look* (3). *See* is an achievement verb, while *look* is an exploration verb [5]. A person can look but not see, but cannot see without looking. A question then is whether learners associate all visual verbs with knowing (and therefore being factive), or if the link is specific to *see*-like verbs and not *look*-like verbs. <u>Methods.</u> We designed a Human Simulation Paradigm study to test whether metaphorical links between vision and knowing would interfere with the acquisition of non-factive *kan*. We did this by testing English *look*, which is not used to describe beliefs in English. We artificially coerced *look* into a frame with a sentential complement (*e.g.*, "Meg looks that Sally likes…") which is typical for *see* and other belief verbs including Mandarin *kan*, to investigate whether *look* would be interpreted as factive in this frame. We tested a total of 80 adult participants on sentences containing negated forms of *see*, *look*, *know*, *think*, or a nonce word (yed/dap/pem/mip/yud). Participants completed a two-alternative forced choice task where they were asked to pick which of two items the mentioned character liked more. As target sentences had a "Meg doesn't VERB that Sally likes A more than B" structure, a high number of A responses would indicate a factive interpretation (like *know*), and a high number of B responses would indicate a non-factive interpretation (like *think*). <u>Results.</u> We found significant differences between *see* and *look* trials (z=9.00, p<.001). *See* followed the expected pattern of responses under a factive interpretation, similar to *know* (Figure 1a). However, *look* patterned with neither *know* nor *think*. Post-hoc analyses revealed distinct subgroups in the data for *look*, some participants showing a factive-like interpretation and some participants showing a non-factive-like interpretation (Figure 1b). This pattern was shared by the nonce words, with correlation of r=.61 between *look* and nonce responses. Nevertheless, the results suggest that participants treat *see* and *look* differently. Learners may infer the factive interpretation of *see* via a seeing is knowing metaphor, but this metaphor seems unlikely to interfere with acquisition of Mandarin *kan* and other *look*-like verbs.

References

[1] Johnson, 1999. *Cultural, typological and psychological issues in cognitive linguistics*. [2] Lakoff & Johnson, 1980. *The Journal of Philosophy*. [3] Sweetser (1990). *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. [4] Li, 2022. *BUCLD46*. [5] Fisher, Gleitman & Gleitman, 1991. *Cognitive Psychology*.

Example sentences for *see* and *kan*

(1) Visual uses of *see* and *kan*
  a. I see an apple.
  b. Wo kan pingguo.
     I KAN apple "I see/look at (the/an) apple."

(2) Belief uses of *see* and *kan*
  a. She sees that they like apples.
     (factive; contradictory to follow up with "... but they don't like apples")
  b. Ta kan tamen xihuan pingguo.
     3sg KAN they like apple Idiomatically: "S/he thinks they like apples."
     (non-factive; not contradictory to follow up with "... but they don't like apples")

(3) Ni zai kan shenme?
   you PROG KAN what
   "What are you looking at?" (*kan* used as "look")

Example trial

"Meg doesn't look that Sally likes apples more than oranges."
Which does Sally like more?
apples oranges

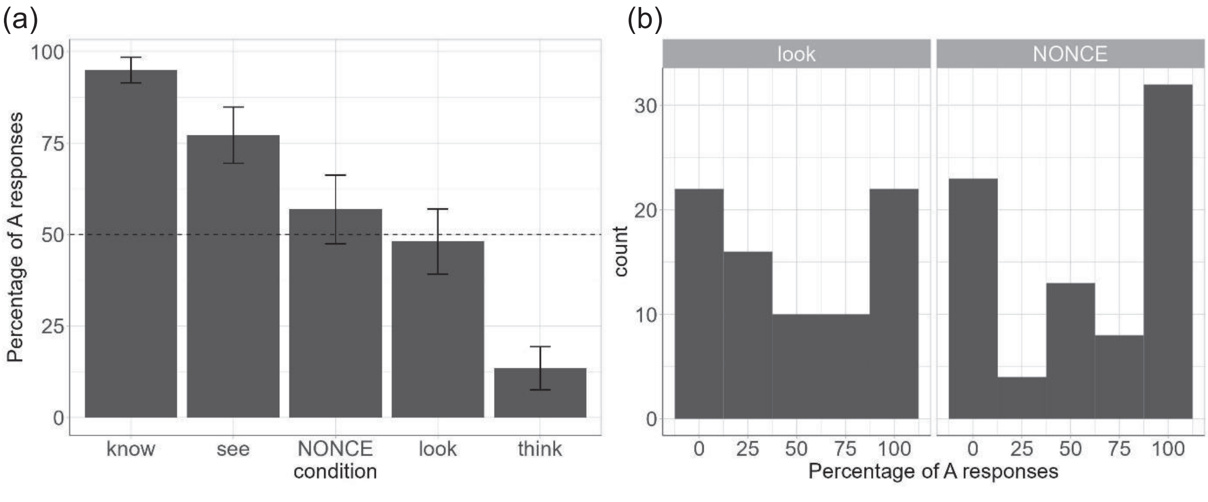


Figure 1. (a) Barplot showing mean percentage of A responses on trials where the verb was negated. High number of A answers indicates that the verb is interpreted as factive; low number of A answers indicates that it is non-factive. (b) Histogram showing distribution of A response percentages for *look* and NONCE negated trials.

PS.2.08

The role of language comprehension ability on anxiety in autistic and typically-developing children

Matthew Ganquan Shi[1,2], Hanlin Wu[1], Gary Yu-Hin Lam[1], Xin Kang[3], Patrick CM Wong[1], Xiujuan Geng[1]; [1]The Chinese University of Hong Kong, [2]The University of Maryland, College Park, [3]Chongqing University, China

Introduction Autistic children have a higher prevalence of anxiety than typically developing (TD) children[1]. Limited social communication (SC) skills among autistic children may lead to social rejection and subsequent anxiety[2], while language deficits can also contribute to heightened anxiety[3]. Given the prevalence of language delay or deficits in autistic children[4], we investigate the role of language comprehension (LC) skills in autism severity (AS) and anxiety levels (AL), potentially shedding light on the underlying neuronal mechanism. We hypothesize that (i) higher AS, particularly reflected in SC skill differences, predicts poorer LC skills, which subsequently predicts heightened anxiety level, and (ii) higher AS predicts higher AL. However, the results showed that there is not prediction from language to anxiety. That is, the correlation between language deficiency and anxiety may in fact be derived from their autistic traits. Path analysis is employed to test these hypotheses.

Methods Ninety Cantonese-speaking children aged 7;8-12;10 (y;m) (Mean = 10.0, SD = 1.34), diagnosed as autistic (*N*=61, M:F=52:9) or not (*N*=29, M:F=22:7), were recruited from local mainstream schools. All participants had normal hearing and nonverbal IQs > 80 and were administered the Autism Diagnostic Observation Schedule, Second Edition, Module 3 (ADOS-2)[5] and Textual Comprehension subscale of the Hong Kong Cantonese Oral Language Assessment Scale (HKCOLAS-TC)[6]. Their parents filled out the Anxiety Scale for Children with [Autism Spectrum Disorder] (ASC-ASD)[7] and the Spence Children's Anxiety Scale (SCAS)[8,9]. ADOS-2 total scores reflect AS, while the Social Affects (SA) subtotals suggest SC skills difficulties. HKCOLAS-TC assesses LC skills. ASC-ASD and SCAS each measures AL, while the former is designed for better detection of autistic children's anxiety. Correlation analysis was conducted among (i) LC, AS, and AL (AL_A for the ASC-ASD, and AL_S for the SCAS), and (ii) LC, SC, and AL. Using R's "lavaan" package, the path analysis tested our hypotheses with the following settings (abbreviations represent variables):

(1) a. $AL = \alpha_1 + \beta_1{*}LC + \beta_2{*}AS + e_1$ b. $AL = \alpha_1 + \beta_1{*}LC + \beta_2{*}SC + e_1$

$AS = \alpha_2 + \beta_3{*}LC + e_2$ $AS = \alpha_2 + \beta_3{*}LC + e_2$

Results LC predicted AS, which subsequently predicted AL. However, LC did not directly affect AL, differing from the original hypothesis. Parallel results were obtained when using SA subtotals to represent SC skill deficits. See Results (supplementary) for the details of correlation and the path analysis.

Discussion and conclusion The path modeling established a total mediation model among language comprehension skills, autistic features, and anxiety in school-age autistic and typically-developing children.[1] The indirect effect of language on anxiety differs from Rodas et al. (2017)[10], possibly due to different assessment criteria (grammar vs. integrative comprehension) and age range. Impaired comprehension may lead to avoidance in social interaction and ineffective communication, resulting in increased anxiety. The results call for specific treatment targeting language comprehension. Moreover, as the neuronal mechanism behind autistic traits remains elusive, our behavioral research provides a possible model connecting them with language and anxiety.

Results (supplementary)

a. **Table 1**. Pearson's Correlation results (**p*<0.05, ***p*<0.01, ****p*<0.001. *N* = 53)

	LC	AS/SC	AL-ASC-ASD
LC	1		
AS/SC	-0.52***/-0.43***	1	
AL-ASC-ASD	-0.31**	0.44***/0.37***	1
	LC	AS/SC	AL-SCAS
LC	1		
AS/SC	-0.52***/-0.43***	1	
AL-SCAS	-0.24*	0.39***/0.33**	1

LC = Language comprehension; AS = Anxiety severity; SC = Social communication; AL = Anxiety level; ASC-ASD and SCAS are the two anxiety scales.

b. **Figure 1**. Path analysis for LC, AS/SC, and AL.

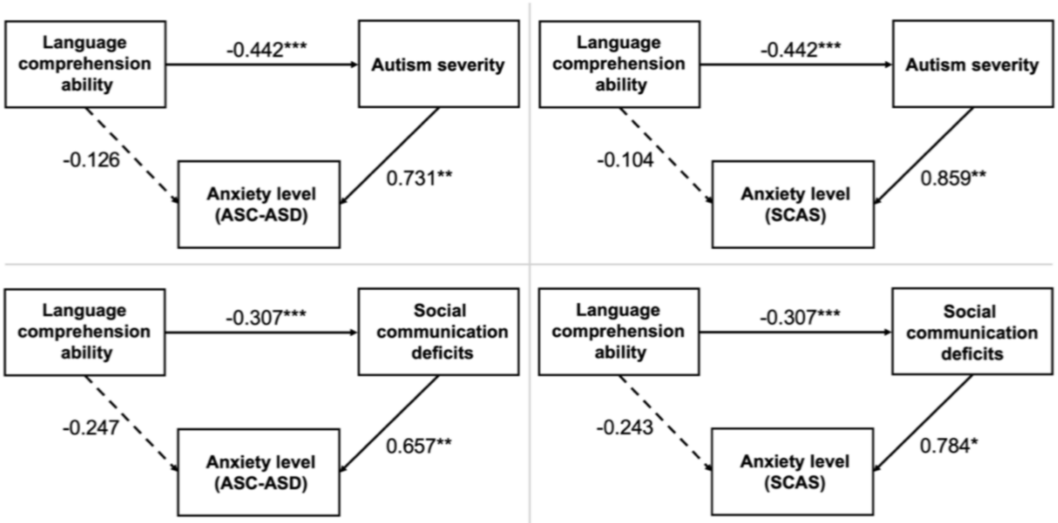


Fig. 1. Models predicting autism severity/social communication deficits and anxiety levels (ASC-ASD/SCAS). **p*<0.05, ***p*<0.01, ****p*<0.001. Dotted lines are non-significant.

References

1. van Steensel, F. J. A., & Heeman, E. J. (2017). *Journal of Child & Family Studies.*

2. Wood, J. J., & Gadow, K. D. (2010). *Clinical Psychology*.

3. Asher, S. R., & Gazelle, H. (1999). *Topics in Language Disorders*.

4. Tager-Flusberg, H. (2016). *Journal of Speech, Language, and Hearing Research*.

[1] If using the language comprehension scale LC to directly predict each of the anxiety measures, you would find significant results (LC predicts AL_A, β=-0.449, p=0.004, R^2=0.085; LC predicts AL_S, β=-0.483, p=0.018, R^2=0.059).

5. Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. L. (2012).
6. T'sou, B., Lee, T. H.-T., Tung, P., Man, Y., Chan, A., & To, C. K. S. (2006).
7. Rodgers, J., Wigham, S., McConachie, H., Freeston, M., Honey, E., & Parr, J. R. (2016). *Autism Research*.
8. Spence, S. H. (1997). *Journal of Abnormal Psychology*.
9. Spence, S. H. (1998). *Behaviour Research and Therapy*.
10. Rodas, N. V., Eisenhower, A., & Blacher, J. (2017). *Journall of Autism and Developmental Disorders*.

PS.2.09

Investigating prediction revision process in English-Chinese bilinguals: A visual-world eye-tracking study

Danning Sun (danning_sun@u.nus.edu), Aine Ito National University of Singapore

Prediction happens in both native and non-native language comprehension [1,2], and a previous study has shown that predictions in a native language can be revised rapidly when the initial prediction is disconfirmed [3]. However, it is unclear whether bilinguals can similarly revise their prediction in their less-dominant language when encountering a prediction error. To this end, we tested predictions and prediction-revision based on classifier constraints in Chinese and investigated whether English-Chinese bilinguals can update their predictions using prediction-inconsistent cues.

Forty-one English-Chinese bilinguals who speak Chinese as a less-dominant language listened to 36 sentences in Chinese while viewing four images (Fig.1). The scene contained a highly predictable object, a less predictable object, and two distractors. In the "less predictable noun, specific classifier" condition, the initial prediction (e.g. coffee) would be disconfirmed by the classifier (e.g. 'kuai') because it was only compatible with the less predictable object (e.g. cake). Thus, the specific classifier served as a prediction-inconsistent cue, which allowed participants to revise their predictions.

A linear mixed-effects model testing the fixation bias towards highly predictable vs. less predictable object 1000ms before classifier onset showed that participants fixated more on the highly predictable object than the less predictable object (β=.11, *SE*=.03, *t*=4.53), suggesting they indeed predicted the highly predictable object based on the context (Fig.2). The cluster-based permutation analysis testing the effect of classifier informativity (general vs. specific) 1000ms before target word onset revealed that participants started looking more at the highly predictable object when the specific classifier was only compatible with it (Fig.3, Highly predictable N, Specific CL). Moreover, participants looked at the less predictable object when the classifier disconfirmed initial prediction (Fig.3, Less predictable N, Specific CL). These findings suggest that informative cues were used to confirm initial prediction as well as to revise prediction. Our study highlights bilinguals' ability to revise and update their predictions based on prediction-inconsistent, informative cues in real-time sentence processing in a less dominant language.

Example of critical sentence and experiment design

Highly predictable noun, specific classifier 安妮在星巴克，买了一杯很美味的咖啡。

Highly predictable noun, general classifier 安妮在星巴克，买了一个很美味的咖啡。

Less predictable noun, specific classifier 安妮在星巴克，买了一块很美味的蛋糕。

Less predictable noun, general classifier 安妮在星巴克，买了一个很美味的蛋糕。

Figure 1 | An example of the visual stimuli. Objects in the two boxes indicate predictable and plausible items in the given context, and the distractors are unpredictable and implausible. 咖啡 (coffee) is the target in the highly predictable condition and 蛋糕 (cake) is the target in less predictable condition.

The general (i.e. uninformative) classifier was compatible with the two predictable nouns, whereas the specific (i.e. informative) classifier was only compatible with one of the two nouns. The target noun in the highly predictable condition was incompatible with the specific classifier in the less predictable condition.

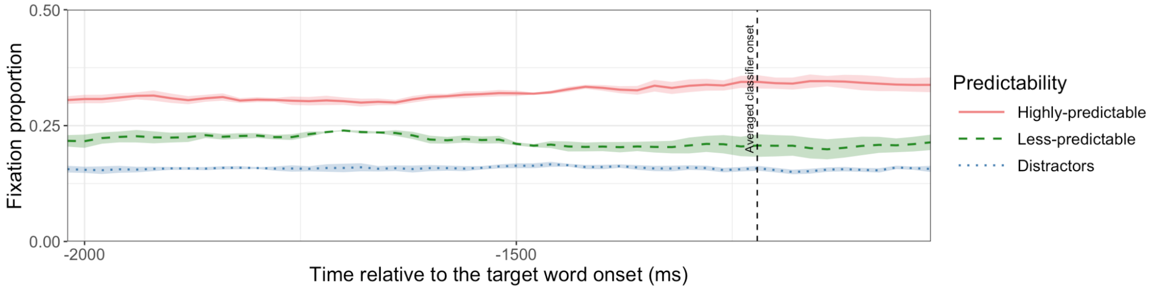


Figure 2 | Fixation proportion on each object with ±1 standard errors (shaded area around each line) before classifier onset.

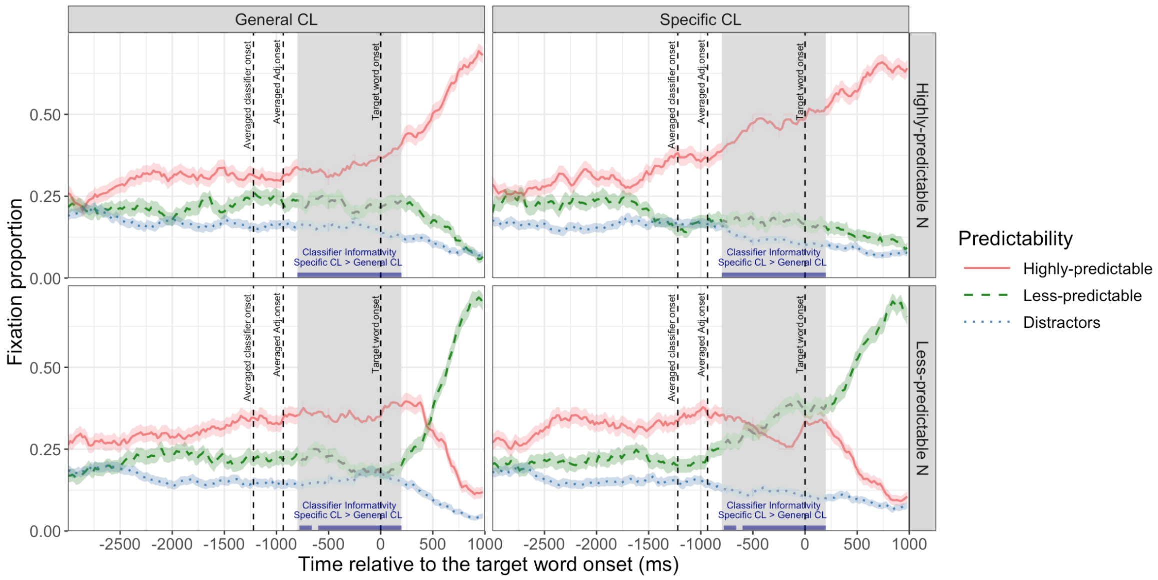


Figure 3 | Fixation proportion on each object with ±1 standard errors (shaded area around each line) before target word onset. The gray shaded area indicates the analyzed time window. The blue line at the bottom indicates a cluster in which the target noun attracted significantly more fixations in Specific CL than in General CL condition.

Reference

[1] Ito, Corley and Pickering, 2018; [2] Chun and Kaan, 2019; [3] Chow and Chen, 2020

Gloss translation

安妮 An ni/ 在 is in/ 星巴克 Starbucks/ 买了 buying/ 一杯 a cup of / 很美味的 very tasty/ 咖啡 coffee.
安妮 An ni/ 在 is in/ 星巴克 Starbucks/ 买了 buying/ 一个 a / 很美味的 very tasty/ 咖啡 coffee.
安妮 An ni/ 在 is in/ 星巴克 Starbucks/ 买了 buying/ 一块 a piece of / 很美味的 very tasty/ 蛋糕 cake.
安妮 An ni/ 在 is in/ 星巴克 Starbucks/ 买了 buying/ 一个 a / 很美味的 very tasty/ 蛋糕 cake.

PS.2.10

The markedness effect on the form-based predictions of sound and number information: a visual-world study

Daiwen GONG (daiwen_gong@u.nus.edu), Aine ITO National University of Singapore

Prediction plays an important role in comprehension, but it remains unclear to what extent listeners use word forms to predict detailed linguistic information, such as the sound and number of upcoming words (e.g., consonant-/vowel-initial; singular/plural). Previous studies have shown mixed results regarding the predictive use of word forms [1,2,3], and these studies did not consider the markedness of word forms—a variable that may affect prediction as the marked form tends to be processed faster than the unmarked form [4]. Markedness indicates an essential representation of a property [5]. In phonology, markedness creates a mutually exclusive relationship between marked and unmarked forms. For example, *an* precedes a vowel-initial word, while *a* does not. In contrast, the different morphological forms of the be-verb, *is* and *are*, are both marked for number. *Is* is marked for singular and *are* is marked for plural [2,6]. To investigate how the markedness of words influences prediction, we conducted a visual-world study on sound and number prediction based on marked or unmarked words in 40 dominant English speakers.

On each trial, participants viewed a scene and listened to a sentence with a noun that was highly predictable from a preceding word (see Figure 1 for examples). The scene contained two images: a target image depicting a noun that aligned with the preceding word and a competitor image showing an incompatible noun. On sound prediction trials, the unmarked (*a*) and marked (*an*) forms of the indefinite article cued predictions of consonant-/vowel-initial nouns. On number prediction trials, the be-verb (*is*/*are*) in there-be constructions cued predictions of a singular/plural noun. We calculated fixation proportion log-ratio between target and competitor images for each 20 ms bin from -800ms to 200ms relative to the target word onset. Additionally, we used the cluster-based permutation analysis to test whether there were more predictive eye movements in the marked vs. unmarked condition.

Participants predictively looked at the target over the competitor on sound prediction trials, regardless of the cue markedness. Besides, the bias towards the target was significantly larger in the marked condition (*an*) compared to the unmarked condition (*a*), suggesting that the marked form cues prediction to a greater extent than the unmarked form. However, on number prediction trials, predictions occurred only in the *are* condition although both *is* and *are* are marked. The lack of predictions in the *is* condition may be because a plural noun may follow *there is* (e.g., *there is a set of plates, there is two guys* [7]), so *there is* may be less informative than *there are*. Based on the findings from the sound and number prediction trials, we argue that the markedness of word forms influences form-based predictions, with the marked and informative cues generating stronger predictions.

Examples of trials

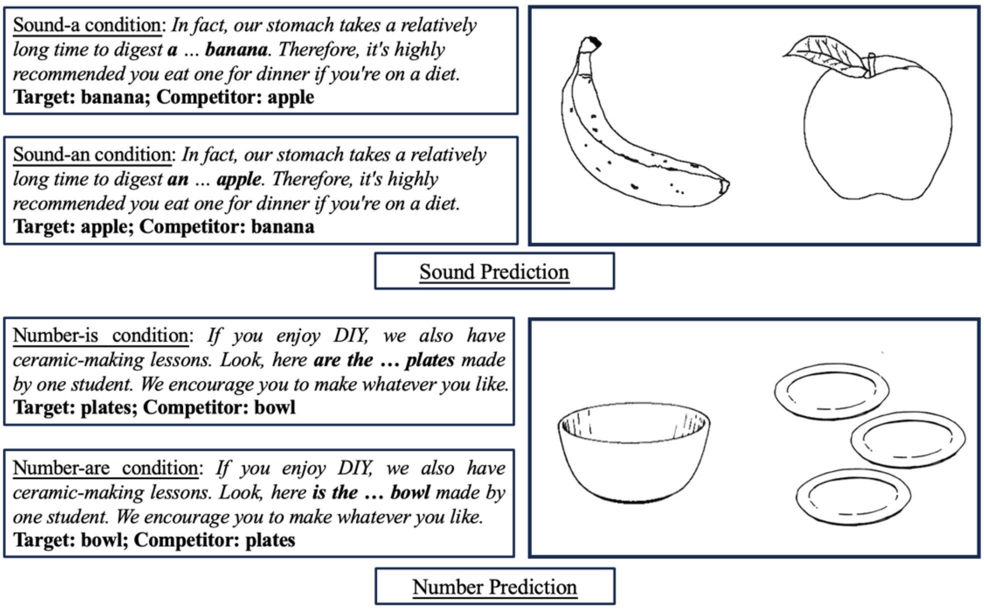


Figure 1 | Examples of trials in Number-are condition, Number-is condition, Sound-a condition, and Sound-an condition.

Results

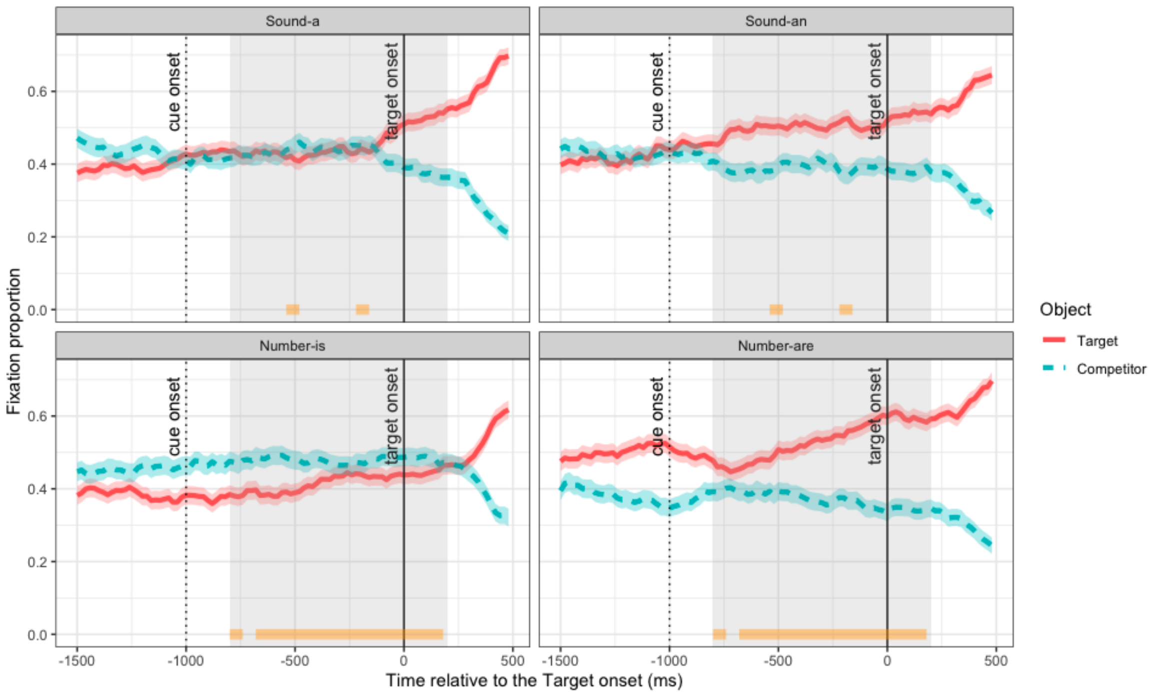


Figure 2 | Proportion of fixation to target and competitor images over time for each condition. The gray shaded area indicates the analyzed time window. Time 0 indicates the onset of the target word. The yellow lines at the bottom indicate clusters in which the target image attracted significantly more fixations in the *an/are* conditions than in the *a/is* conditions.

Reference

[1] Gambi et al., 2018; [2] Lukyanenko & Fisher, 2016; [3] Riordan et al., 2015; [4] Eberhard, 1997; [5] Andrews, 1990; [6] Fraser et al., 1963; [7] National Speech Corpus

PS.2.11

Capturing a Big Picture of a Simultaneous Japanese-English Bilingual: A Preliminary Study collecting both Behavioural and Neurolinguistic Data

TAURA, Hideyuki (Ritsumeikan University, Kyoto, JAPAN)

Abstract

Language proficiency assessment has conventionally targeted a language learner's performance based on such aspects as complexity, accuracy, fluency, and vocabulary. Meanwhile, the last 3 decades have witnessed a rapid advancement in technology, which enables linguists to obtain neuro-linguistic data using fMRI, ERP, and eye-tracking data alongside linguistic data. To date, however, there has been scant research reporting on both aspects of a language learners' performance, that is, the inclusion of both linguistic and neuro-linguistic data. The present study attempts to fill the gap by collecting linguistic and neuro-linguistic data to capture a big picture of bilingual performance.

An early Japanese-English bilingual female participated in this study. For linguistic analysis, spontaneous narrative and written data were collected, using 'Frog, where are you?' (Mayer, 1969) and the Test of Written Languages III (Hammill and Larsen, 1996), respectively. For the neuro-linguistic analysis, fNIRS (spatial resolution) and ERP (temporal resolution) data were collected while the participant engaged in three production tasks - i) a verbal fluency task, ii) a code-switching task, and iii) an oral and written consecutive interpretation. The preliminary analysis revealed that (1) her writing skills are on a par with her monolingual counterparts, (2) the IFG (Broca's area) seems to be efficiently functioning irrespective of the language used in the tasks – either Japanese (Figure 1) or English (Figure 2), and finally (3) the ERP data (Figure 3) indicate that Japanese-to-English and English-to-Japanese code-switching triggered significantly more N400 negativity. This was also observed while the participant repeated English words in English. However, the task of repeating Japanese words in Japanese induced a significantly higher P600 positivity. An overall conclusion will be presented at the conference based on thorough analyses of all the remaining data.

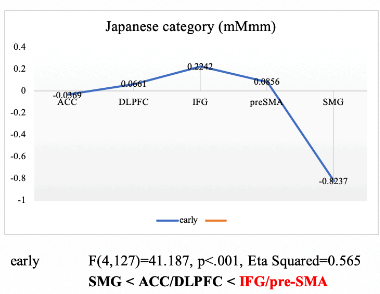


Figure 1: Japanese VFT

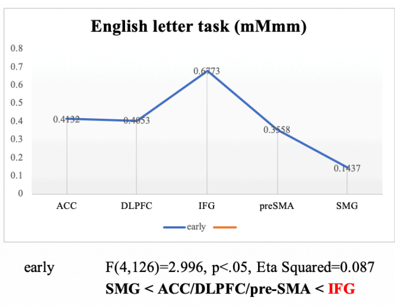


Figure 2: English VFT

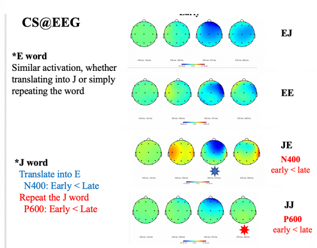


Figure 3: Code-switching

PS.2.12

Readers' Use of Simultaneous Pragmatic Cues in Pronoun Resolution

Tiana V. Simovic & Craig G. Chambers, University of Toronto (tiana.simovic@mail.utoronto.ca)

Identifying a pronoun's referent is key for establishing utterance meaning. Empirical accounts of this process have focused on biases like prioritizing first-mentioned antecedents [1,2] or lexical/discourse-based cues (implicit causality, coherence relations [3,4]). Less work has assessed the effects of pragmatic reasoning, despite theoretical claims about its importance (e.g., [5,6]). Recent work exploring the interpretation of **ambiguous pronouns** in sentences like (1a-b) has highlighted a key role for perspective-based computations [7]. Specifically, in (1a), a report of "remembering when the lecture starts" naturally reflects Madeline's intention to tell Anna something that Anna does not know (cf. to "be informative" [8,9]) and leads readers to overwhelmingly select the subject antecedent (Madeline; 99.7%). In (1b), it is now Madeline who lacks the relevant information, leading to a near-ceiling preference for the object antecedent (Anna; 99.8%). However, do these patterns reflect genuine perspective inferences, or could they reflect reliance on lexical cues stemming from *tell/ask*? Further, are the inferences influenced by additional information? We assessed these questions in two studies. **Study 1** (*n*=60) explored the effect of including **a context sentence** containing additional relevant information that, critically, could shift the above-described patterns ("shifting" condition, 2a-b). A "neutral context" condition, which should preserve the previous patterns, was also included (2c). In the shifting context passages, readers reversed their preferences, selecting **object** antecedents 77% of the time for *tell* and **subject** antecedents 68% of the time for *ask* (neutral: 5% for *tell*, 10% for *ask*), see Fig. 1. This provides compelling evidence that the original patterns are not solely due to *tell/ask*, and instead reflect a rich combination of pragmatic and perspective reasoning. **Study 2** (*n*=60) explored whether these interpretive patterns occur spontaneously **within a standalone sentence**, where readers cannot rely on a situation model generated from a previous sentence, but instead must incorporate different intrasentential cues on the fly. We kept the main verb constant (*ask*) but made simple changes that varied what readers knew about the object antecedent. For example, in (3a, "neutral"), Max is likely asking if the addressee (his son) understood, yet in (3b, "shifting"), it seems more likely that Max is asking the addressee (his tutor) about himself, as here a tutor would hold the relevant expertise. In neutral cases, readers selected **object** antecedents 87% of the time. However, in shifting cases, readers now preferred **subject** antecedents 86% of the time, showing a reversal from default preferences, see Fig. 2. This further shows the effects are not driven solely by the main verb, and that readers also make inferences using information from prior discourse or elsewhere in the sentence containing the antecedents and pronoun. Together, the findings show the robustness and richness of pragmatic reasoning in core aspects of referential interpretation.

Supplemental Example Sentences

(1) a. Madeline told Anna that she remembers when the lecture starts.
b. Madeline asked Anna if she remembers when the lecture starts.

(2) a. Molly, a tour guide, was talking to Hana, who is unfamiliar with Japanese currency. Molly told Hana that she had enough cash to buy a sandwich. ***[shifting-tell]***
c. Molly, who is unfamiliar with Japanese currency, was talking to her tour guide, Hana. Molly asked Hana if she had enough cash to buy a sandwich. ***[shifting-ask]***
c. Molly, who noticed it was almost 12:30 PM, was walking with her good friend Hana. Molly [told/asked] Hana [that/if] she had enough cash to buy a sandwich. ***[neutral]***

(3) a. Max asked his son Gerald if he understood the assignment correctly. ***[neutral]***
b. Max asked his tutor Gerald if he understood the assignment correctly. ***[shifting]***

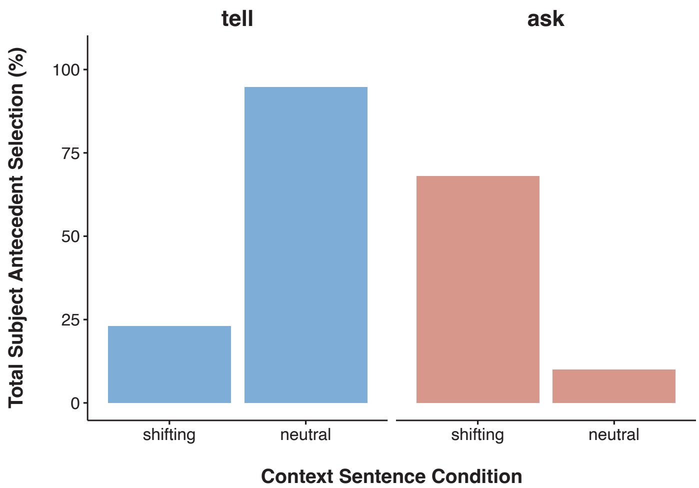


Figure 1*:* Subject antecedent selections (by matrix verb), showing how the **shifting** context sentences effectively "reverse" the preferred antecedent. GLME modelling confirmed this clear switch ($p < .001$). (Expt 1, n_{trials}=20)

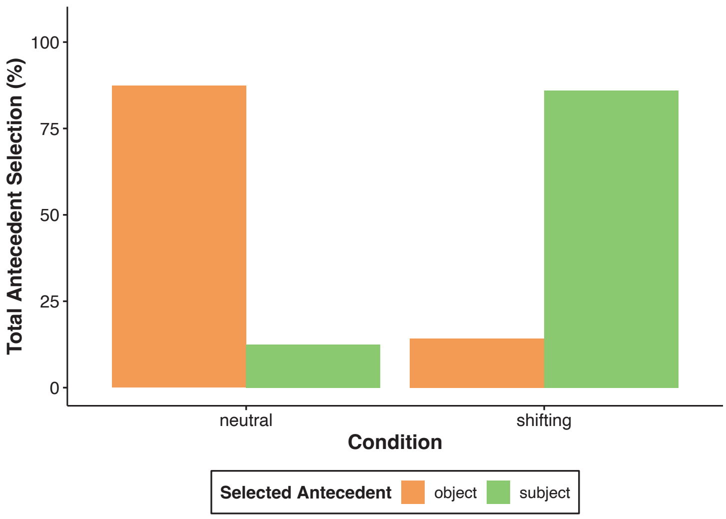


Figure 2*:* Antecedent selections for neutral and shifting sentences, showing how preferences switch in the shifting condition (confirmed with GLME modelling: $p < .001$). (Expt 2, n_{trials}=24)

References: [1] Arnold et al. (2000). *Cognition*. **[2]** Grosz et al. (1995). *Computational Linguistics.* **[3]** Kehler (2002). *Coherence, reference, and the theory of grammar.* **[4]** Garvey & Caramazza (1974). *Linguistic Inquiry*. **[5]** Hobbs (1979). *Cognitive Science.* **[6]** Winograd (1972). *Cognitive Psychology.* **[7]** Simovic & Chambers (2022). *28th AMLaP Meeting; York, UK.* **[8]** Grice (1975). In *Syntax & Semantics Vol. 3.* **[9]** Smyth (1995). *Journal of Child Language*.

PS.2.13

Rapid prediction updating driven by Mandarin classifiers in Maze Reading

Runyi Yao, Yang Fan, Matthew Husband (University of Oxford)

While comprehenders continuously make predictions during sentence processing, these predictions may sometimes be incorrect. An open question is whether such prediction failures are ever costly. A previous EEG study found Italian speakers failed to use local adjectives to update their predictions after encountering prediction-inconsistent gender markers, suggesting that the costs generated by prediction failures limit the overall usefulness of prediction [1]. However, in a visual world eye-tracking study, Mandarin listeners were found to use the information conveyed by the specific classifier (CL) to revise predictions immediately [2]. Similarly, a recent EEG study found that Mandarin readers used informative adjectives to rapidly update their predictions after encountering a prediction-inconsistent CL, even without the aid of visual options [3]. Given the mixed evidence and the different designs across studies (Table 1), the current study aimed to obtain more comparable data by importing the design used in [1] into Mandarin and manipulating CL-context consistency and adjectives to investigate the local effects of prediction failures.

Method. 40 native Chinese speakers participated in a word-by-word Maze reading task involving 32 experimental items and 64 fillers, where distractors were automatically generated using [4]. Following the design of [1], we manipulated global Context {congruent, incongruent} to change whether the CL is consistent with the initial context-based prediction, and manipulated the local Adjective {predictive, neutral} by changing the informativeness of local semantic cues.

Prediction. If prediction failures do not block the use of local semantic cues [2-3], we expect predictive adjectives to speed the processing of target nouns in incongruent conditions. Otherwise [1], no effect of adjective is expected on the target noun in the incongruent condition. Given the results from [1-3], predictive adjectives in congruent conditions should facilitate target nouns.

Results. RT analyses using linear mixed effects models showed a main effect of Context ($p < .01$) on CLs. On target nouns, we found main effects of Context ($p < .001$), Adjective ($p<.05$), and their interaction ($p <.05$). Pairwise comparisons revealed that target nouns were facilitated by predictive adjectives only in incongruent conditions ($p < .01$), but not in congruent conditions ($p = .97$).

Discussion. Our results suggest that Mandarin readers can use subsequent semantic information to update predictions after prediction failures signaled by prediction-inconsistent CLs, supporting the conclusions in [2-3]. The difference between our results and [1] might indicate that comprehenders are better at dealing with semantically-based error signals (e.g., CLs) rather than morphosyntactically-based error signals (e.g., gender markers). Interestingly, predictive adjectives did not further facilitate target nouns in congruent conditions, suggesting that Maze reading may have a lower bound on RTs. Future studies will examine these possibilities.

	Update	Language	Task	Error signal	Local cues
Husband & Bovolenta (2020)	✗	Italian	EEG while reading	Gender marker	ADJ
Chow & D. Chen (2020)	✓	Chinese	Visual world while listening	CL	CL
K. Chen et al. (2024)	✓	Chinese	EEG while reading	Specific CL or ADJ	ADJ
The current study	?	Chinese	Maze	Specific CL	ADJ

Table 1: A summary of three previous studies.

Context	Adjective	Sentences	Distractors
Congruent	Predictive	小明｜走进｜饮料店，｜买｜了｜一｜**杯**｜鲜榨｜的｜**果汁**｜来｜解渴。	X-x-x｜趋势｜笑嘻嘻，｜项｜说｜你｜**敢**｜语句｜说｜**复辟**｜所｜篇章。
Congruent	Neutral	小明｜走进｜饮料店，｜买｜了｜一｜**杯**｜打折｜的｜**果汁**｜来｜解渴。	X-x-x｜趋势｜笑嘻嘻，｜项｜说｜你｜**敢**｜语句｜说｜**复辟**｜年｜篇章。
Incongruent	Predictive	小明｜走进｜蛋糕店，｜买｜了｜一｜**杯**｜鲜榨｜的｜**果汁**｜来｜解渴。	X-x-x｜趋势｜笑嘻嘻，｜项｜说｜你｜**敢**｜语句｜说｜**复辟**｜所｜篇章。
Incongruent	Neutral	小明｜走进｜蛋糕店，｜买｜了｜一｜**杯**｜打折｜的｜**果汁**｜来｜解渴。	X-x-x｜趋势｜笑嘻嘻，｜项｜说｜你｜**敢**｜语句｜说｜**复辟**｜年｜篇章。

Table 2: Example sentences and distractors generated by A-maze, with vertical bars representing word cuttings. CLs and target nouns are bolded, as are their paired distractor words. See the next page for English translations and additional information about Mandarin classifiers.

CL				
	Est	SE	t	p
Adjective	0.98	23.4	0.04	.97
Context	-112.1	34.6	-3.24	.003**
Interaction	-25.5	46.9	-0.54	.59
Target noun				
	Est	SE	t	p
Adjective	-44.8	19.7	-2.28	.029*
Context	-100.3	22.5	-4.46	<.001***
Interaction	87.5	35.3	2.48	.016*
Contrast: Congruent - Incongruent				
Congruent	1.02	23.1	0.04	.96
Incongruent	88.5	30.0	2.95	.007**

Table 3: Output of LMEM and post-hoc pairwise of RTs on CLs and target nouns.

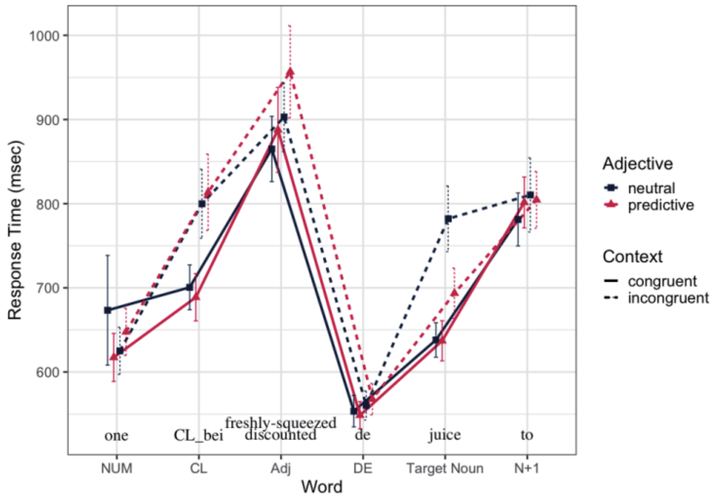


Figure 1: Estimated RTs in four conditions.

Reference
[1] Husband, E. M., & Bovolenta, G. (2020). Prediction failure blocks the use of local semantic context. Language, Cognition and Neuroscience, 35(3), 273-291. [2] Chow, W. Y., & Chen, D. (2020). Predicting (in) correctly: listeners rapidly use unexpected information to revise their predictions. Language, cognition and neuroscience, 35(9), 1149-1161. [3] Chen, K., Xia, F., Wang, S., Chow, W. Y. (2024). Incremental prediction updating through consecutive cues: Evidence from ERPs *HSP 2024*. [4] Levinson, L., Tang, Y., Chiang, L. Y.-C., Zhou, W.-J., & Chung, S. (2023). Multilingual A-maze: Generating Maze Experiments in Mandarin and Beyond. *HSP 2023*.

English translations for example sentences

Context	Adjective	Sentences
Congruent	Predictive	Xiaoming walked into a beverage shop and bought one **CL_bei** freshly-squeezed **juice** to quench his thirst.
Congruent	Neutral	Xiaoming walked into a beverage shop and bought one **CL_bei** discounted **juice** to quench his thirst.
Incongruent	Predictive	Xiaoming walked into a bakery shop and bought one **CL_bei** freshly-squeezed **juice** to quench his thirst.
Incongruent	Neutral	Xiaoming walked into a bakery shop and bought one **CL_bei** discounted **juice** to quench his thirst.

CLs and target nouns are bolded

Additional information about classifiers in Mandarin

Mandarin Chinese is a classifier language where a noun must be accompanied by a classifier when it is quantified or modified by a numeral. Classifiers typically indicate the category of the noun and are chosen based on characteristics such as shape, size, or type. For example, *bei* (杯) is the classifier used for liquids in a cup (e.g., *one CL_bei juice* in the above sentences), and *kuai* (块) is the classifier for cubic item (e.g., bread and cake). Thus, classifiers can serve as an error signal in predictive processing. For example, in incongruent conditions, the global context should raise an expectation for bread or cake, which should be modified by *kuai* rather than *bei*.

PS.2.14

Evaluating the Influence of Background Music on L2 Learners' Syntactic Comprehension —A view of shared structural integration resource hypothesis

Qianxi Lv & Yiwei Zhang Shanghai Jiao Tong University lvqianxi@sjtu.edu.cn

Background music (BGM), accompanying the dialogue or action of the radio or drama, is commonly used by learners when study[1,2]. Whereas the linguistic and musical knowledge systems may be independent, the online structural integration system of which may be shared (SSIRH)[3], esp. managing their hierarchical structures where discrete units combine to form larger entities. This leads to two key predictions: First, concurrent music and language processing may compete for cognitive resources, challenging syntactic comprehension alongside music(Hypothesis1). Secondly, the domain-general resource networks may facilitate online syntactic integration of elements retrieved from representational networks, so that the concurrent processing of music can enhance the integration of syntax (Hypothesis 2,Figure 1). The present study investigated the effect of BGM listening on simultaneous online syntactic processing of L2 learners with a moving-window self-paced reading task (Figure 2). A 3 (background sound: natural quietness; white noise; BGM) ×2 (syntactic complexity: high vs. low) ×2 (music listening habits while learning) mixed-subject design was adopted. 30 sets of object-extracted (high complexity) and subject-extracted (low complexity) RC were used according to the dependency locality hypothesis[4]. 33 Chinese native speakers learning English were divided into two groups according to music listening behavior: Group-MW (N=15) listens to music while working; Group-MLL (N=18) listens to music during language learning.

It was found that both groups comprehended better for lower complexity sentences than higher ones in silence, but both performed better for higher than lower complexity ones with BGM. Notably, Group-MLL achieved higher accuracy for complex sentences with BGM(M=86.84%, SD=33.91%) than without (M=85.16%, SD=35.66%), and it was the opposite for Group-MW. In conclusion, BGM listening can facilitate integrating in reading complex syntax, and thus testified the SSIRH. This enhancement could point to the domain-general resources in syntactic processing. It also revealed the advantage of those who were more accustomed to the musical setting. Despite the increased cognitive load, L2 learners who have the habit of listening to music while learning English are more accustomed to the setting with sounds and thus suffer less from physiological overarousal [5,6].

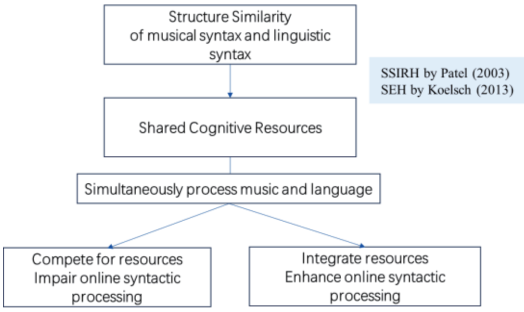


Fig. 1. Hypothesized model of simultaneous processing of language and music of L2 learners

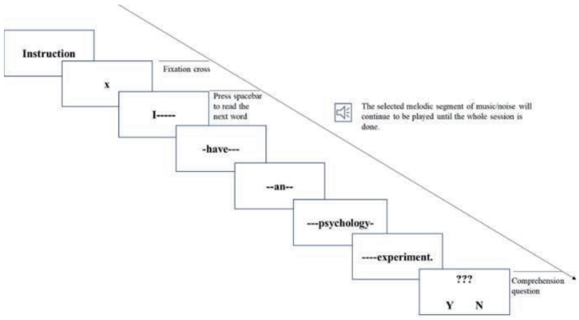


Fig. 2. Example of self-paced reading task

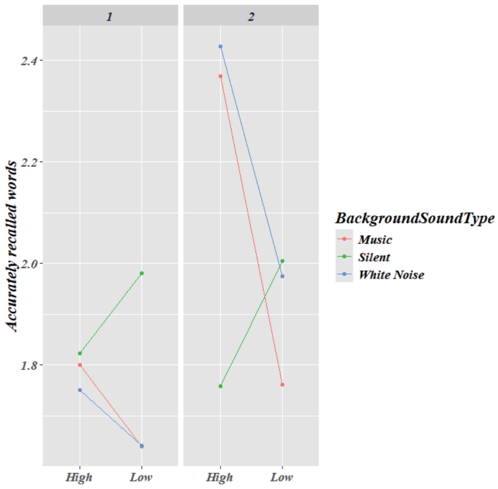


Fig. 3. Predicted comprehension accuracy performance

1: Group-MW; 2: Group-MLL

Table 3. Mean RT of the Critical Region [Mean (Standard Deviations)]

	Group-MW		Group-MLL	
	High complexity	Low complexity	High complexity	Low complexity
Silent	3560.79 (1454.89)	3429.78 (1610.74)	3101.60 (1451.42)	2939.43 (1361.67)
White Noise	3497.83 (1795.15)	3675.39 (1562.10)	2964.69 (1529.55)	2909.25 (1971.96)
Background Music	3561.92 (1630.19)	3376.52 (1356.72)	2931.99 (1528.88)	2787.37 (1414.10)

Table 1. Examples of Language Materials		
Syntactic Complexity	Sentence Type	Material
Low	subject-extracted relative clause	Ex 1. The/ president/ ignored/ the/ reporter/ ***who/ the/ senator/ attacked/*** on/ Tuesday. **Critical region**
High	object-extracted relative clause	Ex 2. The/ president/ ignored/ the/ reporter/ ***who/ attacked/ the/ senator/*** on/ Tuesday. **Critical region**
Comprehension Question		Did the reporter attack the senator on Tuesday?

Selected references:

[1] Hu, X., Jing, C., & Wang, Y. (2021). University students' use of music for learning and well-being: A qualitative study and design implications. *Information Processing and Management*, 58(1), 102409.

[2] Calderwood, C., Ackerman, P. L., & Conklin, E. M. (2014). What else do college students "do" while studying? An investigation of multitasking. *Computers & Education*, 75, 19–29.

[3] Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, 6(7), 674–681.

[4] Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, *68*(1), 1–76.

[5] Etaugh, C., & Ptasnik, P. (1982). Effects of studying to music and Post-Study relaxation on reading comprehension. *Perceptual and Motor Skills*, *55*(1), 141–142.

[6] Crawford, H. J., & Strapp, C. M. (1994). Effects of vocal and instrumental music on visuospatial and verbal performance as moderated by studying preference and personality. *Personality and Individual Differences*, *16*(2), 237–245.

PS.2.15

Processing Speed and Neural Activity Change during Second Language Comprehension Effected by Vocabulary Knowledge

Takara Kenza Allal-Sumoto (*Kobe Design University, Kyoto University*) and Hiroaki Mizuhara (*Nara Women's University, Kyoto University*)
sumoto-k@kobe-du.ac.jp

Enhancing efficient listening comprehension, particularly quick processing, is crucial for foreign language learners. Rapid understanding is needed to match the flow of spoken information. Vocabulary understanding is essential, but may not always suffice for real-time comprehension. Learners often prioritize learning word meanings over effective use in communication, though productive vocabulary knowledge (using words in speech and writing) may positively influence comprehension. This study aims to investigate how productive vocabulary knowledge affects sentence listening comprehension.

We examine (1) comprehension processing speed and (2) neural activity changes using electroencephalography (EEG), focusing on vocabulary production knowledge or the lack thereof. We adopted formulaic sequences as task targets, selecting three- to four-word phrases placed at sentence ends. The experiments were divided into two parts: 1) a screening task, which included both a production and comprehension task to assess each participant's knowledge of the phrases, and 2) the main task, which consisted of a comprehension task for behavioral data collection and both production and comprehension tasks for EEG data. In the main task, participants listened to sentences containing the target phrases. During the comprehension task, participants pressed a key as soon as they understood the sentence's meaning. In the production task, the final word of the sentence—also the last word of the phrase—was omitted, and participants pressed a key upon recalling the missing word, then vocalized their response. For EEG analysis, we performed TFR of EEG power. Additionally, a decoding model was used to predict vocabulary knowledge levels from neural activity using multivariate pattern analysis (MVPA).

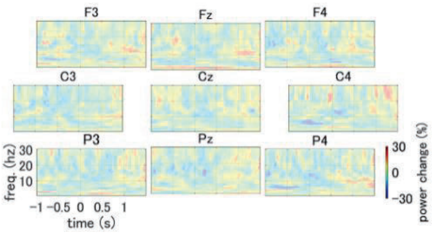
Behavioral results showed that sentences with producible phrases were comprehended significantly faster than those with only comprehensible or non-comprehensive vocabulary (Table 1). As for EEG experiments, although the TFR analysis for the comprehension task did not reveal significant differences between sentences with and without producible phrases (Fig. 1), the decoding results showed that EEG activity could distinguish between sentences with and without productive knowledge (Fig. 2). This indicates that the positive impact of productive vocabulary knowledge on sentence comprehension, driven by distinct neural processing, was evident.

Table 1. Association between knowledge and comprehension processing time based on Generalized Linear Mixed Model Multiple - Comparison with Tukey test

	Estimate Std. Error	z value	Pr(>|z|)
Comprehensible – non comprehensive	0 0.004293	0.004652	0.923 0.624756
Producible – non comprehensive	0 0.019006	0.005256	3.616 0.000824***
Producible - comprehensible	0 0.014713	0.004690	3.137 0.004783**

Fig. 1. Comparison of producible vs non-producible trials: t-test

Comprehension task



Production task

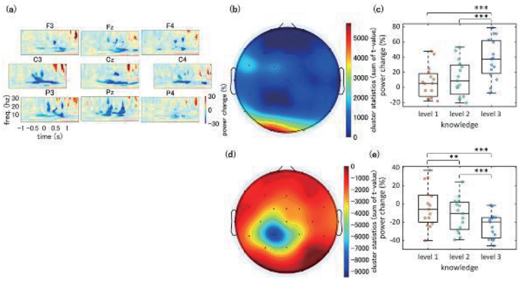
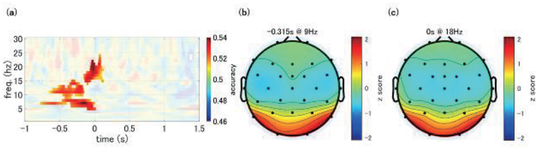
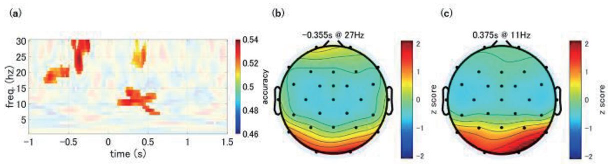


Fig. 2. Decoding model: MVPA

Comprehension task



Production task



PS.3.01

Predicting locative construction in Mandarin Chinese

Li Yuting National University of Singapore (liyuting@nus.edu.sg)

Introduction The question of whether human grammatical knowledge is better explained by categorical or probabilistic grammar has long engaged linguists, psychologists, and cognitive scientists (Bresnan 2007, Bresnan et al. 2007, Lau et al. 2017, Sprouse et al. 2018). This study provides new experimental evidence for the idea that human grammatical knowledge is probabilistic in nature. Using corpus-based probabilistic models of locative inversion in Mandarin Chinese, I conducted two acceptability judgment experiments to explore: (1) Do speakers' acceptability judgments mirror probabilities derived from corpus frequencies? (2) Are the weights of importance of factors in the experiment consistent with those in the probabilistic model?

Experiment 1 examined whether corpus probabilities of locative inversion can predict sentence acceptability ratings. Based on a quantitative analysis of roughly 900 instances, including 562 locative inversion constructions (LI) and 337 uninverted constructions (non-LI), which were annotated for various factors such as discourse function and semantic components, I used logistic regression model to compute the probability of each form occurring as locative inversion in the corpus. 26 native speakers of Mandarin Chinese participated in this experiment. They were provided with 20 small texts (along with 10 fillers) sampled from the corpus data and embedded with both LI and non-LI (see Appendix). Participants were asked to (i) choose the more natural construction and (ii) rate each on a naturalness scale from 0 to 100. The results showed a positive correlation between the computed corpus probabilities and the acceptability ratings, suggesting that participants are sensitive to the likelihood of certain linguistic patterns.

Experiment 2 investigated the relative importance for individual factors used in the corpus. I designed a 2x2x2 experiment analysing the effect of newness of theme (new, given), definiteness of theme (indefinite, definite), and heaviness (long, short). I selected 8 texts from the corpus dataset—4 with locative inversion and 4 with uninverted constructions—and adapted them accordingly. 96 native Mandarin speakers were then presented with 20 items (including 12 fillers) and asked to complete the rating task. Statistical analysis showed that newness had a more pronounced effect compared to definiteness and heaviness, which were similarly weighted. These findings are consistent with the corpus model predictions (newness > definiteness > heaviness).

Conclusion This research shows that corpus probabilities of LI predict acceptability ratings and align with the corpus model, where newness is more significant than definiteness and heaviness. These findings are encouraging as a response to what linguistic generalisations can be learned from a model, an issue addressed in Francis (2021) and many others.

Appendix

One experimental item in **Experiment 1**

(1) *从此小河村结束了没有学校的历史, 半山坡上,* _____。

'Since then, Xiaohe Village has ended the history of having no schools, on the hillside, _____.'

a. Original sentence from the corpus data

一眼旧窑洞里传出了朗朗的读书声. (LI)

[$_{LocP}$ *Yi-yan jiu yaodong-li*] *chuan-chu-le* [$_{NP}$ *langlangde dushu sheng*].

one-CLF old cave.dwelling-in spread-exit-PFV loud reading sound

'From an old cave dwelling came a sound of reading aloud.'

b. Constructed sentence as the alternative variant

朗朗的读书声从一眼旧窑洞里传出(来)了. (non-LI)

[$_{NP}$ *Langlangde dushu sheng*] [$_{LocP}$ *cong yi-yan jiu yaodong-li*] *chuan-chu(-lai) le.*

loud reading sound from one-CLF old cave.dwelling-in spread-exit-hither PST

'The sound of reading aloud came from an old cave dwelling.'

References

Bresnan, J. (2007). Is syntactic knowledge probabilistic? Experiments with the English dative alternation. *Roots: Linguistics in search of its evidential base*, *96*, 77-96.

Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation. In *Cognitive foundations of interpretation* (pp. 69-94). KNAW.

Francis, E. (2022). *Gradient acceptability and linguistic theory* (Vol. 11). Oxford University Press.

Lau, J. H., Clark, A., & Lappin, S. (2017). Grammaticality, acceptability, and probability: A probabilistic view of linguistic knowledge. *Cognitive science*, *41*(5), 1202-1241.

Sprouse, J., Yankama, B., Indurkhya, S., Fong, S., & Berwick, R. C. (2018). Colorless green ideas do sleep furiously: gradient acceptability and the nature of the grammar. *The Linguistic Review*, *35*(3), 575-599.

PS.3.02

Focus restrictions in specificational clauses are language-specific: Experimental evidence from Russian

Anastasia Gerasimova, Lomonosov Moscow State University,
anastasiagerasimova432@gmail.com

The talk addresses the information structure of specificational copular clauses in comparison with predicational ones with regard to the experimental evidence from Russian. The distinction between predicational and specificational readings is based on the order of a non-referential and a referential DP, with DP1 being referential in predicational clauses and DP2 — in specificational clauses [1, 2]. Investigators suggest that specificational sentences have a fixed information structure with the referential DP2 in focus [1, 3]. The restrictions on focus position are treated as universal, although the unavailability of focus on a non-referential DP1 has been experimentally examined only for English [4].

This talk suggests and further tests an alternative explanation for the observed asymmetry. Specificational clauses are usually modelled as an inverted small clause with a left-dislocated non-referential DP1 [1, 5, 6]. The unavailability of focus on this DP1 can follow from the obligatoriness of thematization when fronting. Therefore, languages with different communicative structure might lack such restrictions on specificational clauses. I present experimental data on specificational and predicational copular clauses in Russian, which has another set of constraints on fronting in comparison with Germanic languages, in which specificational clauses have been studied previously.

I conducted two auditory acceptability judgment experiments, which included three variables in their design: the type of copular clause, the focus position and the agreement pattern (2*2*2 design, Latin square, fillers 1:1; lab.js on JATOS3). The two experiments both included a context that licensed the focus marking and differed in the pause position either before or after the copula. 100 Russian native speakers participated in the study and provided their judgments on a Likert scale from 1 to 7 (after excluding outliers; 43 F, 57 M; age: mean 43, SD 12, min 18, max 72). The hypothesis that the focus acceptability in specificational clauses depends on language-specific information structure constraints is borne out. Unlike Germanic languages Russian allows the rhematic accent (which I treat as focus) on the non-referential DP1 if it is licensed by the previous context. Nevertheless, the rhematic accent on a referential DP2 turns out to be the most acceptable experimental condition.

I conclude that the restrictions on focus placement in copular clauses are indeed language-specific. I explain the distinctive theme-rheme structure of specificational clauses by a particular contribution of attributive status of a non-referential DP1 to the meaning. The corresponding feature in the DP structure is considered to be the motivation for the small clause inversion. Importantly, the referential features are found throughout other formal models of Russian and other languages which favorably distinguishes the suggested analysis from previous models of copular clauses.

Figure 1. Z-transformed acceptability scores for experimental conditions and fillers.

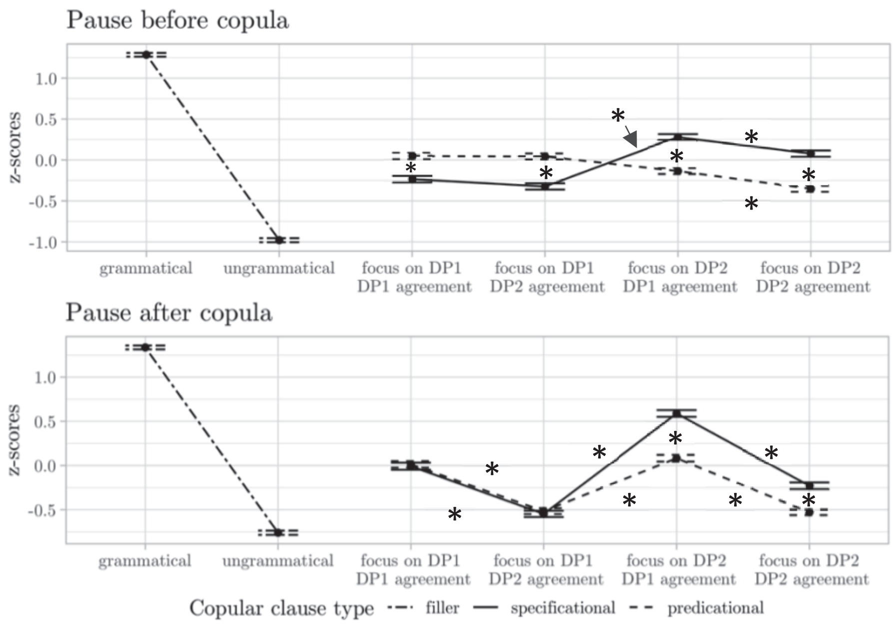


Table 1. The results of the statistical analysis

The optimal LMM after the iterative reduction of model complexity (identical for both experiments)	
z-score ~ BC TYPE + ACCENT SCHEME + AGREEMENT + BC TYPE : ACCENT SCHEME + ACCENT SCHEME : AGREEMENT + (1 + BC TYPE + ACCENT SCHEME + AGREEMENT + BC TYPE : ACCENT SCHEME + BC TYPE : AGREEMENT | id) + (1 | stimul)	
The significance of fixed effects computed using lmerTest	
Pause after copula	BC TYPE (β = -0.32, SE = 0.07, t = -4.55, p-value << 0.0001) ACCENT SCHEME (β = -0.21, SE = 0.08, t = -2.64, p-value = 0.009) BC TYPE : ACCENT SCHEME (β = 0.75, SE = 0.1, t = 7.53, p-value << 0.0001)
Pause before copula	AGREEMENT (β = -0.53, SE = 0.07, t = -7.74, p-value << 0.0001) AGREEMENT : ACCENT SCHEME (β = -0.18, SE = 0.09, t = -2.07, p-value = 0.04) BC TYPE : ACCENT SCHEME (β = 0.43, SE = 0.1, t = 4.47, p-value << 0.0001)

Appendix. Additional information on Russian copular clauses

In the experiments we observe low or marginal acceptability of copular clauses with two nominative DPs (1a). This result is expected as Russian has an alternative way of conveying the same meaning with a non-referential DP1 marked instrumental (1b), which is a more frequent strategy than binominative clause [7].

(1)	a. [*Pričina*	*avarii*]$_{F.SG}$	*byla*$_{F.SG}$	/*byli*$_{PL}$	[*neispravnye*	*tormoza*]$_{PL}$.
	reason.NOM	of failure	was.F.SG	/were.PL	broken.NOM	brakes.NOM

‘The reason of failure were the broken brakes.’

b. [*Pričinoj* *avarii*]$_{F.SG}$ *byli* $_{PL}$ [*neispravnye* *tormoza*]$_{PL}$.
reason.INST of failure were.PL broken.NOM brakes.NOM
'The reason of failure were the broken brakes.'

References: [1] den Dikken, M. (2006) Specificational copular sentences and pseudoclefts. The Blackwell Companion to Syntax. 2006. **[2] Mikkelsen, L.** (2011) Copular clauses. In K. von Heusinger, C. Maienborn, P. Portner (Eds.), Semantics: An international handbook of natural language meaning, Vol. 2 (pp. 1805–1829). Berlin: De Gruyter. **[3] Heycock, C., Kroch, A.** (2002). Topic, focus, and syntactic representation. In L. Mikkelsen, C. Potts (Eds.), Proceedings of WCCFL 21 (pp. 141–165). Somerville, MA: Cascadilla Press. **[4] Hartmann, J. M.** (2019). Focus and prosody in nominal copular clauses. In S. Featherston, R. Hörnig, S. von Wietersheim, S. Winkler (Eds.), Information structure and semantic processing (pp. 71–104). De Gruyter, 2019. P. 71–104. **[5] Hartmann, J. M., Heycock, C.** (2020) (Morpho) syntactic variation in agreement: specificational copular clauses across Germanic. Frontiers in Psychology, 10, 2994. **[6] Béjar, S., Kahnemuyipour, A.** (2017). Non-canonical agreement in copular sentences. Journal of Linguistics, 53(3), 463–499. **[7] Pereltsvaig, A.** (2007). Copular sentences in Russian: A theory of intra-clausal relations. Dordrecht: Springer Science & Business Media.

PS.3.03

Dynamic Mental Representations of Motion Paths in Incremental Sentence Processing

Menghan Wang (minnie.wang@unimelb.edu.au), Helen Zhao, University of Melbourne

Language speakers construct dynamic mental representations of scenes in sentence comprehension (Zwaan, 2004). Yet few studies have investigated the dynamicity of mental representations in language processing (Sato et al., 2013) and little attention is paid to the onset and offset of constructing mental representations in sentence processing (Zwaan, 2008). The study investigates the spatial construal of motion paths in incremental sentence processing. We selected a satellite language Mandarin, in which motion is typically implied by resultative verb compounds composed of a displacement verb (e.g., *zǒu* 'walk', indicating the manner of motion) and a directional verb (e.g., *jìn* 'enter', indicating the path of motion).

We recruited 215 adult Mandarin native speakers to read Chinese motion path sentences segment by segment with interleaved diagrams (see Example 1). Participants were randomly assigned to one of two tasks: a self-paced reading task (SPRT) (n = 91) to passively read diagrams, or a self-paced verification task (SPVT) (n = 124) to judge whether the diagram depicted the motion path implied in the sentence. We set three timepoints by presenting a diagram after the displacement verb (Timepoint 1), the directional verb (Timepoint 2), or the landmark (LM) (Timepoint 3). Congruency was manipulated on the diagram (Example 2) by contrasting a matching (e.g., *into*) diagram and mismatching (e.g., *out of*) diagram, with a neutral baseline diagram represented by a series of hashtags (#####). The reaction times (RTs) for processing diagrams in each task were analysed using linear mixed-effects models.

The SPRT results revealed interference effects at Timepoint 1 (Figure 1) and Timepoint 2 (Figure 2), evidenced by longer RTs for mismatching diagrams compared to the neutral baseline. However, RTs for matching diagrams did not differ significantly from the neutral condition. No effect was observed at Timepoint 3 (Figure 3). Conversely, the SPVT results indicated compatibility effects across all timepoints (Figures 4, 5, and 6), with RTs for the matching condition shorter than the mismatching condition but longer than the neutral condition. The SPRT findings suggest mental representations are routinely activated upon processing information from the displacement verb, which triggers predictions about the upcoming motion direction and subsides once the direction is specified. It supports that mental simulation is a dynamic process (Zwaan, 2004). Meanwhile, the compatibility effects in the SPVT imply that conscious deliberation on the connection between visual and sentential meaning enhances the activation of mental representations of motion paths. This active engagement is irrespective of the information availability and distinctive from routine activation when integrating ongoing spatial configuration in meaning construction.

Example 1

早上	九点， /	一位 /	勤奋的 /	员工 /	
zǎo-shàng	*jiǔ-diǎn*	*yí-wèi*	*qín-fèn-de*	*yuán-gōng*	
[Adverbial phrase]		DET	ADJ	SUBJ	
morning	nine o'clock	a	diligent	employee	
走 /	进 /	办公室 /	然后 /	开始 /	工作。
zǒu	*jìn*	*bàn-gōng-shì*	*rán-hòu*	*kāi-shǐ*	*gong-zuò*
Displacement V	Directional V	OBJ	Conjunction		[Verb phrase]
walk	into	office	then	start	work

'At 9 am, a diligent employee walked into the office and started working.'

Example 2

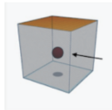
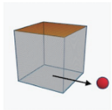
Matching	Mismatching	Neutral
		#####
into	*out of*	

Figure 1. SPRT: RTs at Timepoint 1 (Diagram - Directional Verb - LM Sequence)

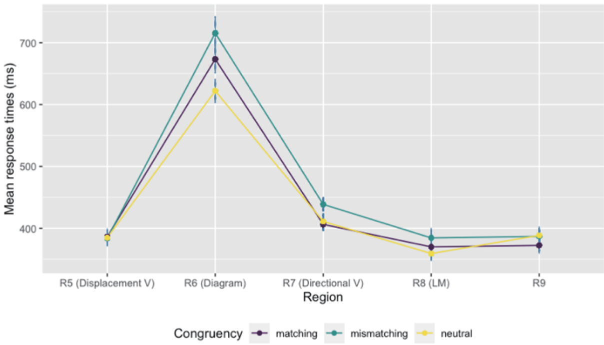


Figure 4. SPVT: RTs at Timepoint 1 (Diagram - Directional Verb - LM Sequence)

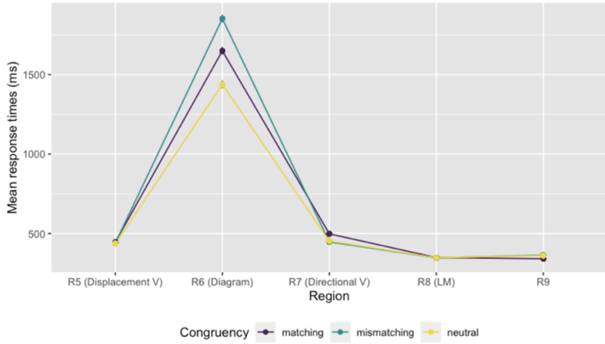


Figure 2. SPRT: RTs at Timepoint 2 (Directional Verb - Diagram - LM Sequence)

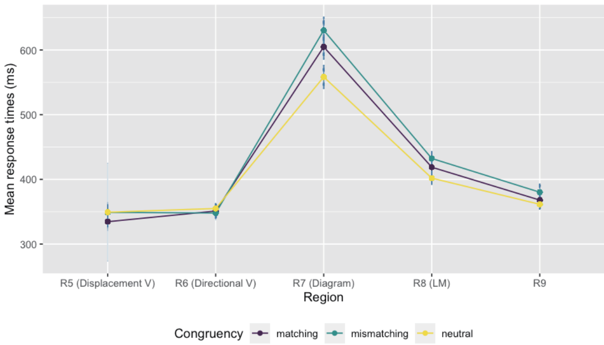


Figure 5. SPVT: RTs at Timepoint 2 (Directional Verb - Diagram - LM Sequence)

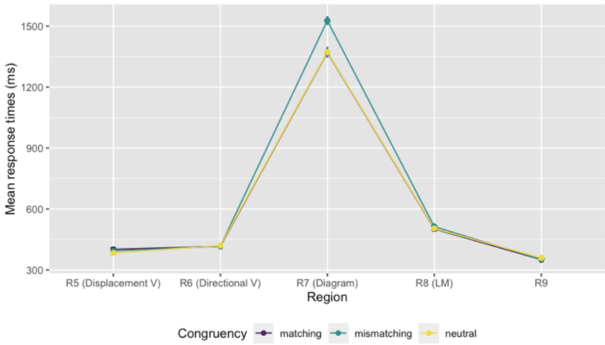


Figure 3. SPRT: RTs at Timepoint 3 (Directional Verb - LM - Diagram Sequence)

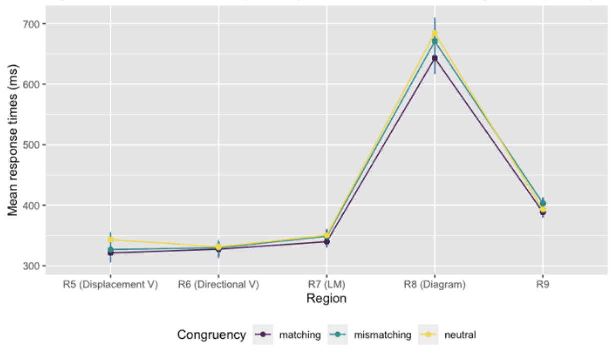
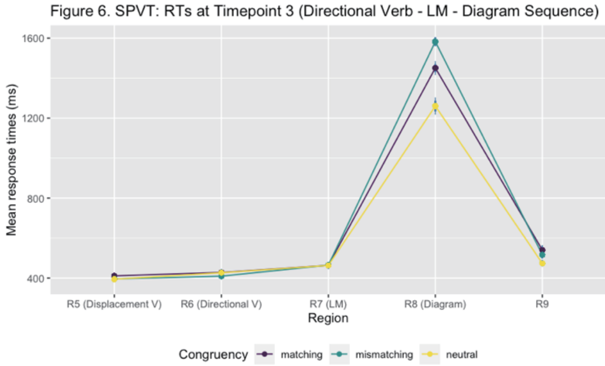


Figure 6. SPVT: RTs at Timepoint 3 (Directional Verb - LM - Diagram Sequence)



References

Sato, M., Schafer, A. J., & Bergen, B. K. (2013). One word at a time: Mental representations of object shape change incrementally during sentence processing. *Language and Cognition, 5*(4), 345–373. https://doi.org/10.1515/langcog-2013-0022

Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and* theory (Vol. 44, pp. 35–62). Elsevier Science.

Zwaan, R. A. (2008). Experiential traces and mental simulations in language comprehension. In M. de Vega, A. Glenberg, & A. Graesser (Eds.), *Symbols and embodiment: Debates on meaning and cognition* (pp. 165–180). Oxford University Press.

PS.3.04

Online measures of factivity in Chinese: An eye-tracking experiment

Darcy Sperlich, Hong Liu, Stephen Jeaco, Xi'an Jiaotong – Liverpool University,
darcy.sperlich@xjtlu.edu.cn

Factivity concerns verbs like 'know' (factive) that presuppose the truth of their complement, and 'believe' (non-factive), that do not. For instance, in 'John knows that Susan is late', it is presupposed that 'Susan is late', unlike with 'believe'. Previous studies suggest that syntactically, factive and non-factive verbs select for different structures, with Kiprasky and Kiprasky (1970) suggesting that factive verbs select for a simpler structure, whilst de Cuba (2007) hypothesized the opposite. However, psycholinguistic evidence for such selectional properties remains scant. Moreover, there are verbs such as Chinese 记 得 *jìdé* 'remember' which peculiarly alternate between factive and non-factive readings. The variable status of the verb *jìdé* prompts further investigation into processing dynamics of factivity.

This study investigates the processing of factivity, specifically focusing on the cognitive effort required for different syntactic structures, and the implications for understanding how presuppositions are processed in language cognition. We conducted an eye-tracking experiment with 241 participants in a between-participant design to explore how *jìdé*, the factive verb *zhīdào* ('know'), and the non-factive *xiāngxìn* ('believe'), are processed within contexts containing contradictions, as in (1). Contradictions challenge the presupposition introduced by factive verbs, but not for non-factive verbs, which have no presupposition.

We expect to observe different reading patterns when the embedding verb is *zhīdào* compared to *xiāngxìn*. We hypothesized in the beginning, *jìdé* will be processed in the spirit of cognitive economy, selecting for the simpler structure, similar to *zhīdào* if factive readings are easier, or akin to *xiāngxìn* if non-factive readings prevail. However, as *jìdé* is lexically ambiguous and may activate both readings, we expect it to show unique behaviors later in sentences containing contradictions. Readers might resolve contradictions more swiftly with *jìdé* than with *zhīdào* due to the accessible non-factive reading, or experience processing delays compared to *xiāngxìn*, influenced by the concurrent activation of the factive reading.

Overall, the data suggests that as compared to the factive *zhīdào*, the non-factive *xiāngxìn* incurred longer regression paths, reflecting increased cognitive effort in processing the non-factive verb. Additionally, *jìdé* selects for a more complex structure (non-factive) in the first sentence construction. As the sentence progresses, more differences are found between *jìdé* and *zhīdào*, supporting a continued non-factive interpretation. Moreover, a non-factive reading profile still prevails in second-pass reading times for *jìdé*, indicating that an alternative factive interpretation was not maintained. These findings are discussed in relation to the semantics and pragmatics of 'remember', while investigating speaker commitment to propositions.

(1) 老孔[知道/相信/记得]小余尖叫了，对面的小余。其实小余没尖叫。

Lǎokǒng [zhīdào/xiāngxìn/jìdé] Xiǎoyú jiānjiàole, duìmiàn de Xiǎoyú.

Laokong know/believe/remember Xiaoyu scream.ASP opposite POS

Xiaoyu Qíshí Xiǎoyú méi jiānjiào.

Actually Xiaoyu NEG scream

‘Laokong knows/believes/remembers that Xiaoyu screamed, Xiaoyu who lives opposite. Actually, Xiaoyu didn’t scream.’

References

de Cuba, C. F. (2007). On (Non)Factivity, Clausal Complementation and the CP-Field. PhD Dissertation. NY: Stony Brook University.

Kiparsky, P. & Kiparsky, C. (1970). FACT. In Progress in Linguistics: A Collection of Papers (pp. 143-173). Berlin, Boston: De Gruyter Mouton.

PS.3.05

The Impact of Auditory and Visual Exposure on the Imageability of Chinese Verbs

Huanyu Zhu[1], Yuzhe Tang[1], Johannes Gerwien[2]

[1]Institute for General and Typological Linguistics, University of Munich, Munich, Germany

[2]Institute for German as a Foreign Language Philology, Heidelberg University, Heidelberg, Germany

zhu.huanyu@outlook.com

This study investigates how auditory and visual modalities affect the imageability and processing speed of Chinese verbs. *Imageability*—the ease with which a word evokes a mental image—is crucial in language comprehension and memory (Paivio, 1991). While modality effects have been studied in alphabetic languages, less is known about logographic languages like Chinese, where characters provide direct visual semantic meaning cues (Liu et al., 2007). We hypothesized that visual presentation of Chinese verbs would lead to higher imageability ratings and faster reaction times than auditory presentation.

An internet-based experiment with 100 native Chinese speakers aged 18–23 was conducted. Sixty Chinese verbs of varying imageability levels were selected from an established database (Su et al., 2023). In a within-subjects design, participants rated the imageability of each verb presented in both auditory and visual modalities, with reaction times recorded.

Bayesian linear regression analyses showed a trend of higher imageability ratings for visually presented verbs, but the difference was not statistically significant (Estimate = 0.92, 95% CI [-0.77, 2.59]). Reaction times were significantly shorter for visual presentation (Estimate = -25.07, 95% CI [-44.14, -6.21]), indicating faster mental image formation. We attribute the faster processing to the simultaneous activation of phonological and semantic codes in the visual modality, unlike the sequential processing in the auditory modality (Perfetti et al., 2005; Zhou & Marslen-Wilson, 2000).

These findings suggest that while the intensity of imageability is similar across modalities, the visual modality enhances the speed of mental image formation for Chinese verbs. This has implications for theories of dual coding and sensory processing, indicating that visual stimuli may facilitate quicker semantic access in logographic languages.

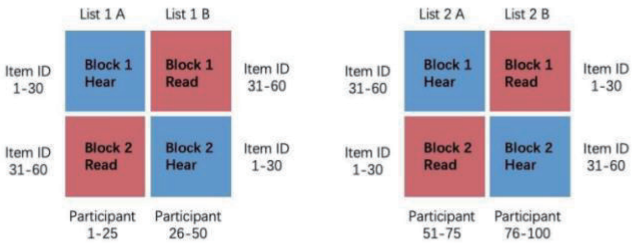


Fig.1 Experiment Design

Item ID 1-25 contain low, medium, high imageability words

Item ID 26-50 contain low, medium, high imageability words

Every experimental session contains 2 blocks

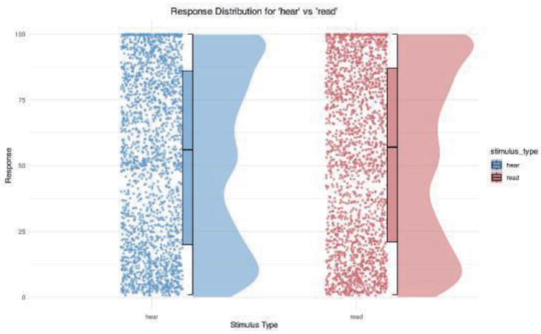


Fig.2 Raincloud Plot for Ratings

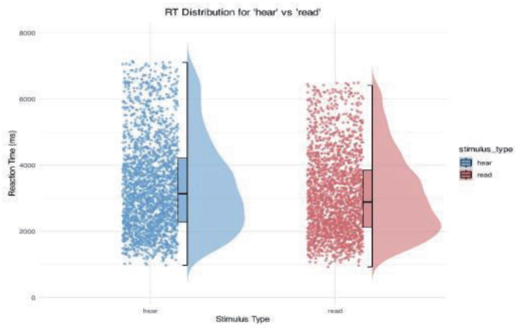


Fig.3 Raincloud Plot for RTs.

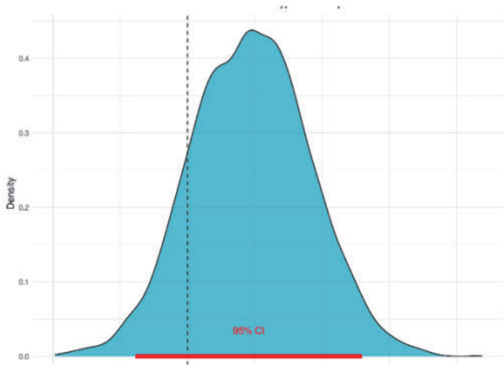


Fig.4 Bayesian linear regression-Ratings

Estimate: 0.92

Standard Error (Est.Error): 0.86

95% Confidence Interval (CI): [-0.77, 2.59]

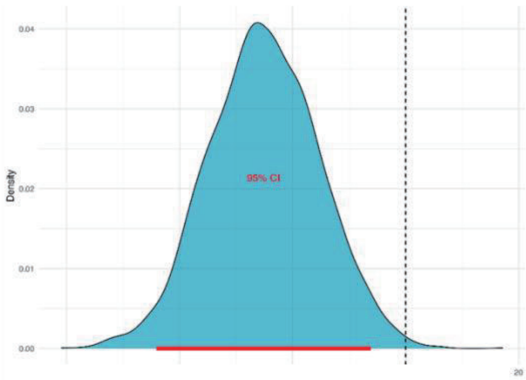


Fig.5 Bayesian linear regression-RTs

Estimate: -25.07

Standard Error (Est.Error): 9.74

95% Confidence Interval (CI): [-44.14, -6.21]

References

Paivio, A. (1991). *Dual coding theory: Retrospect and current status*. Canadian Journal of Psychology, 45(3), 255–287.

Liu, Y., Perfetti, C. A., & Wang, H. (2007). *Visual analysis and lexical access of Chinese characters by Chinese as second language learners*. Language and Linguistics, 8(4), 979–992.

Perfetti, C. A., Liu, Y., & Tan, L. H. (2005). *The lexical constituency model: Some implications of research on Chinese for general theories of reading*. Psychological Review, 112(1), 43–59.

Zhou, X., & Marslen-Wilson, W. (2000). *The relative time course of semantic and phonological activation in reading Chinese*. Journal of Experimental Psychology: Learning, Memory, and Cognition, 26(5), 1245–1265.

Su, Y., Li, Y., & Li, H. (2023). *Imageability ratings for 10,426 Chinese two-character words and their contribution to lexical processing*. Current Psychology, 42, 23265–23276.

PS.3.06

Emotion on the Scale: Testing EMOTIONS ARE WEIGHT Metaphor

Yi-Chi Hsiao, Yao-Ying Lai (Graduate Institute of Linguistics, National Chengchi University)

112555002@g.nccu.edu.tw

Metaphors can be regarded as a cognitive process through which people access an abstract concept (target domain) via a more tangible one (source domain), according to the Conceptual Metaphor Theory (e.g., Lakoff & Johnson, 1980; Lakoff & Kövecses, 1987). Among various metaphorical mappings, the EMOTIONS ARE WEIGHT metaphor remains relatively underexplored. The study examined the metaphorical connection between EMOTIONS and WEIGHT, where positive emotions are linked with lightness, exemplified by metaphorical expressions such as *qingkuai* 'cheerful', and negative emotions with heaviness, exemplified by expressions such as *chénzhòng* 'heavy-hearted', in Mandarin.

We conducted an Implicit Association Test (IAT) experiment consisting of five stages (Fig. 1; Greenwald & Farnham, 2000), as adopted in previous studies on metaphors (e.g., Zhu et al., 2024). In each trial, the participants categorized a target word (e.g., *excited*) as related to certain emotional valence (positive, negative) and weight (light, heavy), combined parallelly as two answer options in the Congruent and Incongruent conditions (Fig. 2). Congruency refers to whether or not the emotion valence and weight matched as hypothesized. In the Congruent condition, the positive emotion combined with lightness, and the negative emotion combined with heaviness; the combinations were reversed in the Incongruent condition. We recruited 19 native speakers of Taiwan Mandarin (female=14). Results of one-way repeated-measures ANOVA showed a significant effect of Condition (F (1,18) = 14.61, p = 0.001), such that the Congruent condition were responded faster than Incongruent (Fig. 3). This suggests a conceptual link between positive emotion and lightness, as well as between negative emotion and heaviness. The findings manifest the EMOTIONS ARE WEIGHT metaphor in online processing, supporting an embodied metaphorical understanding of emotional experiences through weight.

Figure 1. Five stages of IAT (revised from Greenwald & Farnham, 2000, Fig. 1)

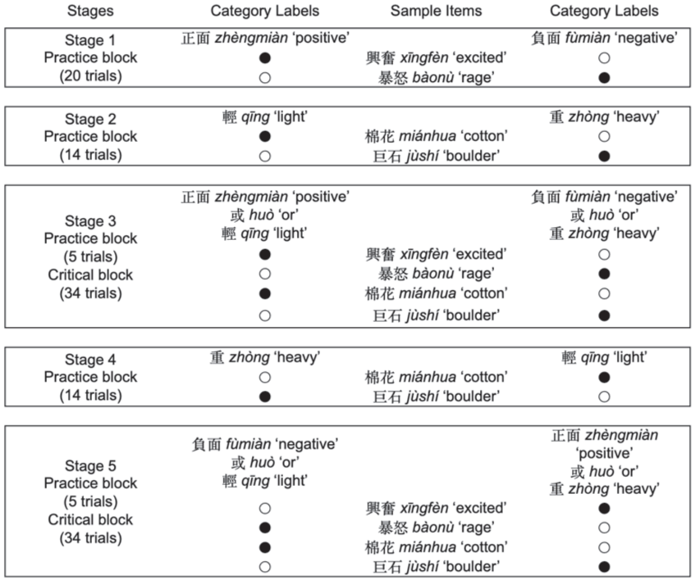


Figure 2. Trial procedure of stage 3 (Congruent) and stage 5 (Incongruent)

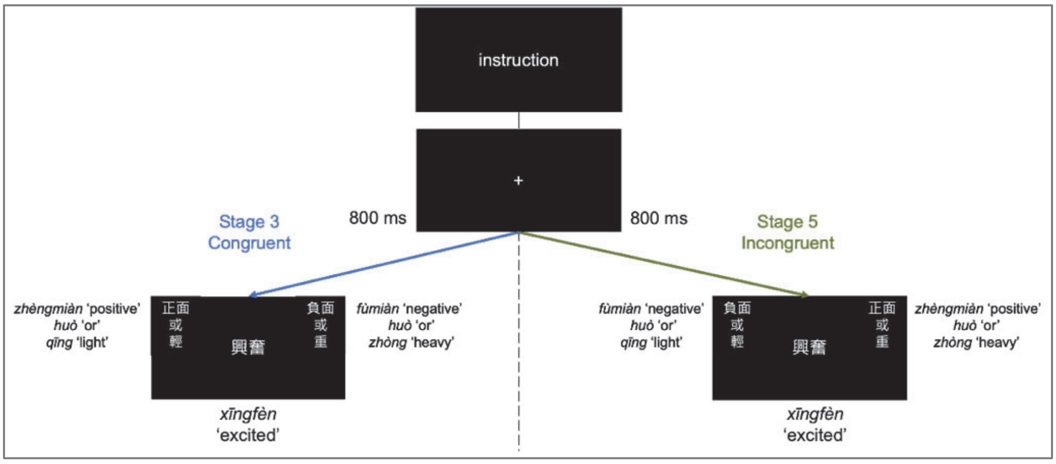
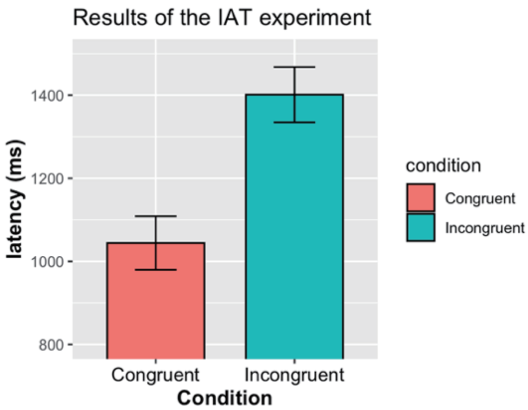


Figure 3. Results of the IAT experiment: mean reaction latency (ms) ± 1 se.



References: **[1]** Greenwald, A. G., & Farnham, S. D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology*, *79*(6), 1022–1038.
[2] Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By* (M. Johnson, Ed.). University of Chicago Press.
[3] Lakoff, G., & Kövecses, Z. (1987). The cognitive model of anger inherent in American English. In D. Holland & N. Quinn (Eds.), *Cultural Models in Language and Thought* (pp. 195–221). Cambridge University Press.
[4] Zhu, T. O., Chen, P., & Durgin, F. H. (2024). The Ups and Downs of Black and White: Do Sensorimotor Metaphors Reflect an Evolved Perceptual Interface? *Metaphor and Symbol*, *39*(3), 169–182.

PS.3.07

Contextual coherence and expectancy influence semantic integration: Evidence from ERP and neural oscillation

Mengyue Peng, Xiaoming Jiang, Xiaolin Zhou; Institute of Linguistics, Shanghai International Studies University; pengmengyue1016@163.com

Sentence processing involves a dynamic and incremental construction of contextual representation, and different contextual cues may affect subsequent semantic integration. Using electrophysiological measures, we examined how contextual expectancy and coherence influence semantic integration interactively during sentence reading. We used Chinese sentences with the structure of “subject + verb + AP (adjective phrase, including an adjective and a structural auxiliary *de*) + noun”, in which the object noun could be semantically constrained by both the local AP and the distant verb. The strength of contextual expectancy towards the noun was manipulated according to the cloze probability of the mostly used word in the coherent sentence context, while the coherence of context was manipulated according to whether the AP could fulfill the selection restriction of the verb. The high-expectancy sentence elicited a reduced N400 than the low-expectancy sentence. Compared with the correct, baseline sentence, an N400 effect was observed for the noun violating semantic constraints from the verb and the AP. Importantly, the N400 effect and the synchronic low gamma (35-45 Hz) activity was larger for the noun violating the high-expectancy context than that violating the low-expectancy context. The above results indicate the costs of disconfirmed predictions and the mismatch between predicted and perceived words. Moreover, a left-lateralized P600 effect was also observed for the noun violating preceding context, but this effect disappeared when the context was incoherent, indicating that coherent context is necessary for re-analysis. Furthermore, the synchronic high gamma (60-80 Hz) activity, which relates to the predictability of an upcoming word based on the preceding sentence context, was larger for the noun violating the coherent context with high expectancy than that with low-expectancy. These findings suggest that both contextual expectancy and coherence affect semantic integration of incoming words, reflected in different ERP and oscillatory signatures.

PS.3.08

From Morphosyntax to Semantics: L1 Korean Learners' Acquisition of Japanese 'ga'

Chika Okada / The University of Auckland
coka031@aucklanduni.ac.nz, chikaokadaling@gmail.com

Background: Both Korean and Japanese are pro-drop languages that possess nominative particles (Akiyama 1991). The Japanese nominative marker ga 'NOM' (***JP-ga***) is an independent particle that requires an exhaustive-listing reading (ELR) only when it is the subject of an individual-level predicate, which denotes the individuals' permanent properties (Kratzer 1989, Kuroda 1965, Kuno 1973). Korean has nominative allomorphy i/ka 'NOM' (***KOR-i/ka***) in the case of matrix clauses has ELR when it is also the subject of an individual-level predicate (e.g., Han 1992, Ha 2007, Kim 1967). Given these similarities between ***KOR-i/ka*** and ***JP-ga***, to what extent do L1-Korean learners acquire the ELR of ***JP-ga***, considering its interpretive characteristics, the influence of L1 knowledge, and differences in proficiency levels? **Experiment:** A picture selection task was administered to 24 L1-Korean learners of Japanese (elementary (N4) and intermediate (N2) levels) and 24 native speakers of Japanese as a control group. The test sentences (1) require ELR, as it includes ***JP-ga*** with an individual-level predicate in a matrix clause, fulfilling the condition for ELR in Korean as well (2). Stimuli included 9 target sentences and 49 fillers, with 9 sentences targeting the Japanese topic marker 'wa' as a baseline, due to its morpho-syntactic and semantic similarities to the Korean topic marker 'nun-/neun-' (3,4) (Lee 2002). Participants were presented with a context and a target sentence (5 seconds each) with audio stimuli, then had 15 seconds to select as many pictures as they liked based on how they interpreted (Figure 1). **Results**: For the experiment on ***JP-ga***'s ELR, L1 Korean learners showed a significant difference compared to the control group. Beginners accepted the ELR 11.6% of the time, intermediates 21.2% of the time, while the control group accepted it 89.4% of the time. A linear regression with the response variable as the obtained response and the predictor variable as participants' group revealed statistical significance ($F(2, 645) = 278.7$, $R^2 = 0.66$, $p < 2.2e-16$) (Figure 2). For 'wa' type fillers, single linear regression did not present statistical significance, indicating that participants performed like the control group (Figure 3). **Proposal**: The results indicate that participants struggle to acquire the ELR of ***JP-ga*** even at intermediate levels, as shown by their behavior compared to filler results. Considering the similarity of ***KOR-i/ka*** and ***JP-ga***, it is unlikely that L1-Korean learners do not transfer their knowledge of ***KOR-i/ka*** to ***JP-ga***. This study suggests that the acquisition of semantic domains may occur after the morpho-syntactic level (e.g., Schwartz & Sprouse 1996, White 2011). Given that ***KOR-i/ka*** can be seen as a semantic subset of ***JP-ga*** as it requires more conditions for ELR, it should be acquirable for L2 learners if positive evidence is provided (e.g., Grüter et al. 2010). This study considers that the acquisition difficulty observed here has been caused by ELR itself being a subtle semantic property, and despite the presence of positive evidence, learners still face challenges in mastering this nuanced aspect of ***JP-ga***.

PS.3.09

Children's derivation of scalar inference from *or*-sentences: Evidence from varying the degree of relevance

Children's difficulty in deriving scalar implicature (SI) from *or*-sentences has been attributed to a variety of factors, the most prevalent of which is the difficulty in accessing the alternative (e.g., Chierchia et al., 2001; Tieu et al., 2016; Gotzner et al., 2020). However, considering the recent discussion about the constraint of relevance (e.g., Skordos and Papafragou, 2016), the present study explores whether children's pragmatic difficulty with *or*-sentences stems from their problem accessing the alternative or from discerning the relevance of SI.

Using a 4 between-subject design, Bengali-speaking preschool children (N= 40) and adults (N=40) were assigned to 4 different conditions: C1: [+alternative, high-relevance]; C2: [–alternative, high-relevance]; C3: [+alternative, low-relevance]; and C4: [–alternative, low-relevance]. In +alternative conditions, the alternative was primed in the participants' recent memory. In –alternative conditions, such priming was absent. The degree of relevance was manipulated by altering the question under discussion. Relevance was deemed high when only two objects (A, B) were under question and low when four objects (A, B, C, D) were under question. While in high-relevance conditions, SI was relevant (A or B, and not A and B), in low-relevance conditions, exhaustivity implicature held a greater relevance (A or B, and not C and D). The prediction was that if challenge lies in finding relevance, performance should improve in high-relevance conditions than in low-relevance conditions, regardless of alternative priming.

Children observed two puppets — Mili and Doraemon — interacting while looking at some picture cards (N=15), each featuring four objects. In test trials, Mili asked Doraemon if he would bring 2 objects (high-relevance)/4 objects (low-relevance) to her, to which Doraemon replied with the promise, "I will bring you an A or a B". In +alternative conditions, the alternative was primed by embedding $A \wedge B$ in Mili's question. Children had to give Doraemon color pencils based on the number of objects he promised to bring (2 objects: big pencil; 1 object: small pencil). The prediction was that children would give a small pencil to Doraemon if they derived the SI.

Results indicate that children in high-relevance conditions derived implicature significantly more than those in low-relevance conditions (p = 0.0424 (C1vs.C3); 0.0089 (C2vs.C3); 0.0123 (C1vs.C4); 0.002 (C2vs.C4)) and there was no significant effect of alternative priming (p = 0.665 (C1vs.C2); 0.807 (C3vs.C4)) [Fisher's Exact Test]. Nevertheless, children's performance did not surpass the chance level even in high-relevance conditions (p = 0.214 (C1); 0.0675 (C2)) [Binomial Probability test]. Adults' success rates were at the ceiling for all the conditions.

Findings here contribute to the current debate regarding children's difficulty deriving SI from *or*-sentences suggesting that their difficulty lies in discerning relevance. Additionally, they bring crosslinguistic support for the nativist account of children's disjunction comprehension from a population that is understudied in acquisition research, thereby contributing to the ongoing nature-nurture debate (e.g., Crain and Khlentzos, 2010; Morris, 2008; Tomasello, 2005).

Figures

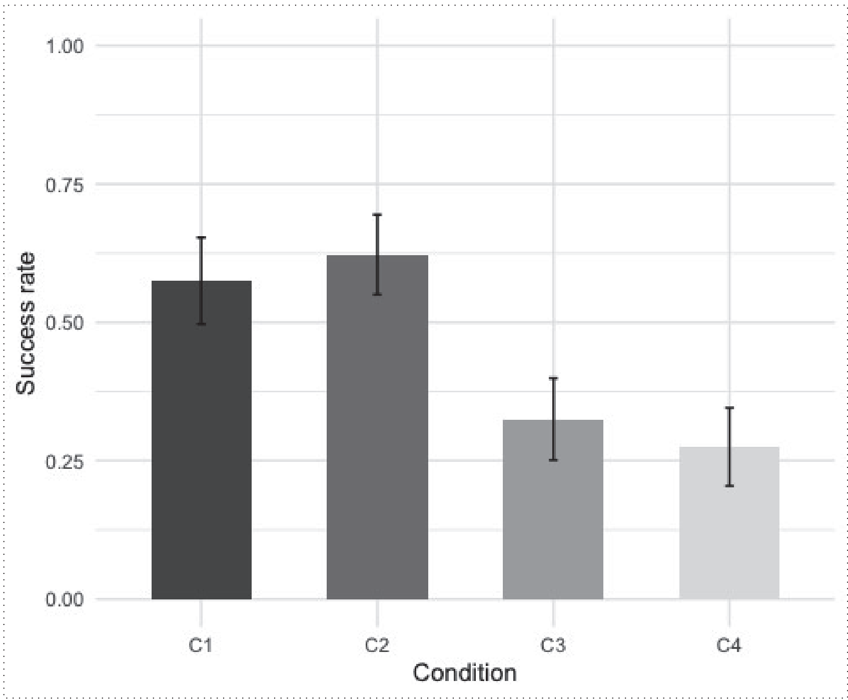


Figure 1: Children's success rates in four conditions

References

Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A., Meroni, L., et al. (2001). The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. In *Proceedings of the 25th Boston University conference on language development*, volume 25, pages 157–168.

Crain, S. and Khlentzos, D. (2010). The logic instinct. *Mind and Language*, 25(1):30–65.

Gotzner, N., Barner, D., and Crain, S. (2020). Disjunction triggers exhaustivity implicatures in 4-to 5-year-olds: Investigating the role of access to alternatives. *Journal of Semantics*, 37(2):219–245.

Morris, B. J. (2008). Logically speaking: Evidence for item-based acquisition of the connectives and & or. *Journal of Cognition and Development*, 9:67 – 88.

Skordos, D. and Papafragou, A. (2016). Children's derivation of scalar implicatures: Alternatives and relevance. *Cognition*, 153:6–18.

Tieu, L., Romoli, J., Zhou, P., and Crain, S. (2016). Children's knowledge of free choice inferences and scalar implicatures. *Journal of semantics*, 33(2):269–298.

Tomasello, M. (2005). *Constructing a language: A usage-based theory of language acquisition*. Harvard university press.

PS.3.10

The Interpretational Preferences of Null and Overt Pronouns in L2 Chinese

Xiaoyu Liu & Nayoung Kwon, University of Oregon. xliu12@uoregon.edu; nkwon@uoregon.edu

Languages display systematic differences in their pronominal system. For example, while Chinese has both null and overt pronouns, English only has overt pronouns in finite clauses. In this study, we aim to examine the acquisition of linguistic features absent in one's language by examining pronominal resolution in Chinese by English-Chinese bilinguals.

It has been suggested that the interpretation of a pronominal element is not random in a given language. For example, the Accessibility Theory (Ariel, 1990, 1994) argues that the use and interpretation of anaphoric expressions are influenced by cognitive motivations. In this theory, the choice of referential forms depends on the mental accessibility of their referents. *Pro*, the highest marker of accessibility, typically refers to the most salient antecedent, such as the syntactic subject. Overt pronouns, positioned lower on the Accessibility scale than *pro*, are more likely to refer to relatively less salient antecedents, like non-subject NPs (e.g., objects). Referential NPs are expected to refer to even less salient antecedents compared to overt pronouns. Evidence was found both in English and Chinese. In English, a language without *pro,* sentences with overt pronoun were read faster than a full NP when the antecedent occurs in the subject position (Garrod & Sanford,1982). In Chinese, studies have shown that both *pro* and overt pronouns are subject-biased (Yang et al., 1999) but the preference for the subject antecedent is stronger for *pro* than for overt pronoun (Zhang & Kwon, 2022). These results suggest that successful interpretation of sentences requires the understanding of different interpretational biases of referential expressions. Accordingly, interpretation of a pronominal element could pose a challenge for language learners especially when their target language differs from their mother tongue in the pronominal inventory, as is the case of English and Chinese. On the other hand, if the interpretational biases are indeed motivated by cognitive mechanism as the Accessibility Theory argues, language learners will not face challenges when mastering different interpretational biases of pronominal elements in their target language. In fact, previous studies suggest that English learners of Chinese can successfully interpret *pro* and overt pronouns (Zhao, 2012, 2014; Slabakova et al., 2023). However, these studies examined the sentences with only one potential referent. Thus, it is essential to investigate whether Chinese learners exhibit sensitivity to the relative saliency of arguments when faced with multiple potential antecedents. This study aims to address this question and understand the interpretational biases in adult L2 Chinese speakers.

We ran a questionnaire study with 21 L2 learners of Chinese and 21 native Mandarin speakers. The study involved globally ambiguous sentences in which the subject and object (e.g., Miss Zhu and Miss Wan) could potentially serve as antecedents for *pro*, overt pronouns, and full NPs (n=30, see Table 1). Participants completed a web-based questionnaire in a controlled lab setting to identify the antecedent for pronominal elements within the sentences. The results showed that both *pro* and overt pronouns were subject-biased, while full NPs were object-biased for both native speakers and L2 learners (Fig.1&2). Furthermore, as found in Zhang and Kwon (2022), *pro* displayed a stronger preference for the subject antecedent compared to overt pronouns. Importantly, this preference was weaker for L2 learners in comparison with native speakers. Additional analysis among L2 learners of different proficiency levels showed that advanced learners exhibited a stronger subject preference for *pro* compared to overt pronouns than L2 intermediate learners. Overall, our findings are consistent with the Accessibility Theory, suggesting that the interpretational biases are motivated by cognitive mechanisms, even when specific linguistic features are absent in one's native language. In addition, the results suggested that a higher level of fluency is associated with more native-like pronoun resolution.

Table 1. Experimental Sentences

<table>
<tr><th>Condition</th><th>Test sentence</th></tr>
<tr><td>pro</td><td>朱小姐陪万小姐逛时，Ø 遇到了一个朋友。
Miss Zhu/accompany/Miss Wan/ shop/when, Ø meet ASP one friend
'When Miss Zhu was shopping with Miss Wan, Ø met a friend. '</td></tr>
<tr><td>overt pronoun</td><td>朱小姐陪万小姐逛时，她遇到了一个朋友。
Miss Zhu/accompany/Miss Wan/ shop/when, she meet ASP one friend
'When Miss Zhu was shopping with Miss Wan, she met a friend.'</td></tr>
<tr><td>NP</td><td>朱小姐陪万小姐逛时，这位小姐遇到了一个朋友。
Miss Zhu/accompany/Miss Wan/ shop/when, this lady meet ASP one friend
'When Miss Zhu was shopping with Miss Wan, this lady met a friend.'</td></tr>
<tr><td colspan="2">Comprehension Question:
根据这个句子，谁遇到了一个朋友?
According to the sentence, who met a friend?
A. 万小姐 Miss Wan
B. 朱小姐 Miss Zhu</td></tr>
</table>

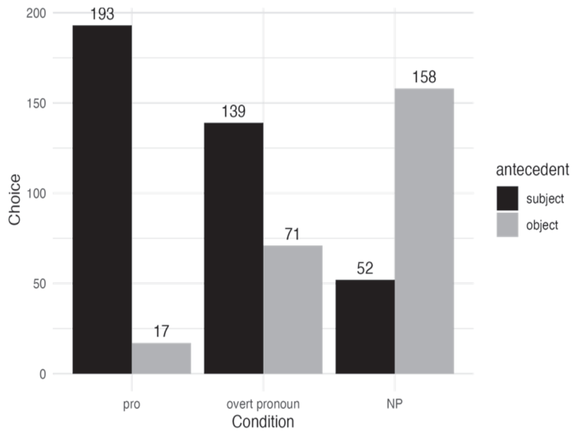


Fig.1 Antecedent Choices of native speakers Chinese

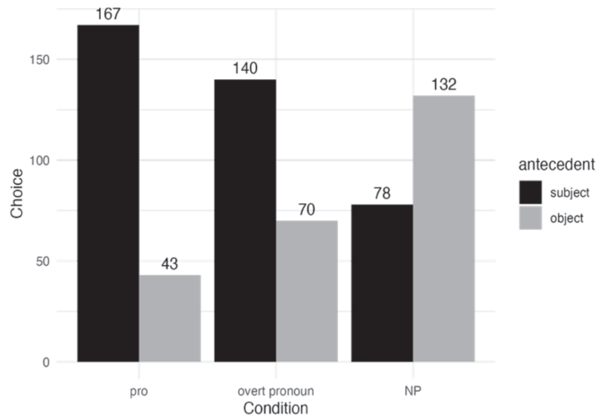


Fig.2 Antecedent Choices for L2 learners of Chinese

References
[1] Zhang, A., & Kwon, N. (2022). The interpretational preferences of null and overt pronouns in Chinese. *Journal of Linguistics*, *58*(3), 649-676.
[2] Ariel, M. 1990. Accessing Noun Phrase Antecedents. London: Routledge.
[3] Ariel, M. 1994. Interpreting anaphoric expressions: a cognitive versus a pragmatic approach. Journal of linguistics 30,3-42.
[4] Garrod, S & A.J. Sanford, 1982. The mental representation of discourse in a focused memory system: Implications for interpretation for the anaphoric noun-phrases. Journal of Semantics 1, 21-41.
[5] Slabakova, R., Zhao, L., Baker, L., Turner, J., & Tuniyan, E. (2023). Null and overt pronoun interpretation in L2 Mandarin resultative constructions. *Linguistic Approaches to Bilingualism*.
[6] Zhao, L. X. (2012). Interpretation of Chinese overt and null embedded arguments by English-speaking learners. *Second Language Research*, *28*(2), 169-190.
[7] Zhao, L. X. (2014). Ultimate attainment of anaphora resolution in L2 Chinese. *Second Language Research*, *30*(3), 381-407.

PS.3.11

Attraction Effects in the Processing of Long-distance Chinese Classifiers: An Eye-tracking Study

Xiaoyu Liu & Nayoung Kwon, University of Oregon. xliu12@uoregon.edu; nkwon@uoregon.edu

This study investigated memory retrieval processes during the online processing of the classifier-noun agreement in Chinese, using the eye-tracking method. Previous studies suggested that the retrieval processes involved in the processing of an agreement are affected by an illicit antecedent (hereafter *distractor*) as well as a licit antecedent (Dillon et al., 2013; Kwon & Sturt, 2016). These results were taken to suggest that memory retrieval processes are parallel and content-addressable such that a distractor with matching features to the retrieval cues could be mis-retrieved (McElree et al., 2003; Lewis and Vasishth, 2005; Van Dyke and McElree, 2006) as a last resort to save the sentences particularly when they are ungrammatical (Wagers et al., 2009). However, these studies were based on binary features, and thus do not provide information on the effect of the degree of semantic overlap between the distractor and a retrieval cue.

Native speakers of Chinese (n=40) read the sentences (n=160) consisting of four conditions. Four conditions are created to manipulate the semantic distance between the classifier and the distractor (Table 1). Semantic relatedness is determined by animacy, and lexical effects are counterbalanced. The Grammatical condition included a classifier matching its licit noun. The Distractor-Matching condition included a classifier matching the distractor. The Distractor-Related condition included a classifier related to the distractor. The Distractor-Unrelated condition employed a classifier unrelated to the distractor. Note that all these conditions are ungrammatical except for the first condition. Analysis of linear-mixed effects regression was done within all regions and two regions of interest are reported: R1(distractor “flower”) and R9 (classifier spill-over “is”).

Based on previous findings mentioned in the preceding section, we predict that a distractor will be mis-retrieved when it shares features with a retrieval cue. Thus, there will be a stronger distractor effect for the Distractor-Matching condition than for the Distractor-Related condition. On the other hand, the Distractor-Unrelated condition will not show any distractor related effect, as the distractor does not share any feature with a retrieval cue. The results found a significant grammaticality effect at R9: the Grammatical condition was processed faster than the remaining three conditions for two reading measures for first pass multi fixation and total reading times. Additionally, the results indicated the distractor-related effects (Fig.1): At R9, Distractor-Unrelated condition elicited significantly longer reading times than the Grammatical Condition. Both Distractor-Matching and Distractor-Related conditions did not differ from the Grammatical condition. The distractor-related effects were also found at R1 (Fig.2), both Distractor-Matching and Distractor-Related conditions did not differ from the Grammatical condition. To summarize, these results suggested that (i) the parser is sensitive to the grammatical constraints of the classifier-noun agreement, and (ii) the retrieval processes during the processing of the agreement are also sensitive to semantic features, given that processing difficulty was reduced as the parser is sensitive to the overlapping features or even partial between the retrieval cue and the distractor. Our study provides crosslinguistic evidence that memory retrieval process is parallel and content-addressable.

Table 1 Sample Experimental Sentences

Region	**R1**	**R2**	**R3**	**R4**	**R5**	**R6**	**R7**	**R8**	**R9**	**R10**
Grammatical condition	花儿 flower	吸引	了	很多	小狗	其中	一	只	是	金毛
Distractor-Matching	花儿 flower	吸引	了	很多	小狗	其中	一	朵	是	金毛
Distractor-Related	小草 grass	吸引	了	很多	小狗	其中	一	朵	是	金毛
Distractor-Unrelated	食物 food	吸引	了	很多	小狗	其中	一	朵	是	金毛
Gloss	Flower/flower /grass/food	attrac t	PFV	many	dog	among	one	CL	is	Golden retrieve r
Trans.	'Flower/flower/grass/food attracts many dogs, among which one is a golden retriever'									
Question	"Is there a dog?" Yes/No									

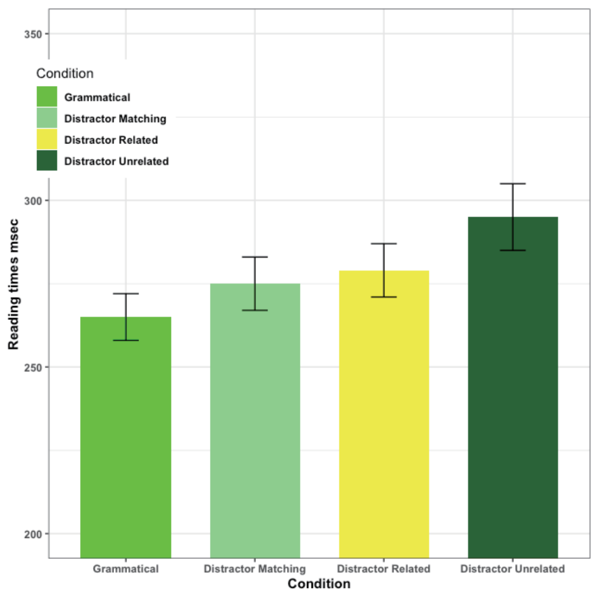


Fig.1 Gaze duration at classifier-spillover (R9)
(Error bars shows standard errors)

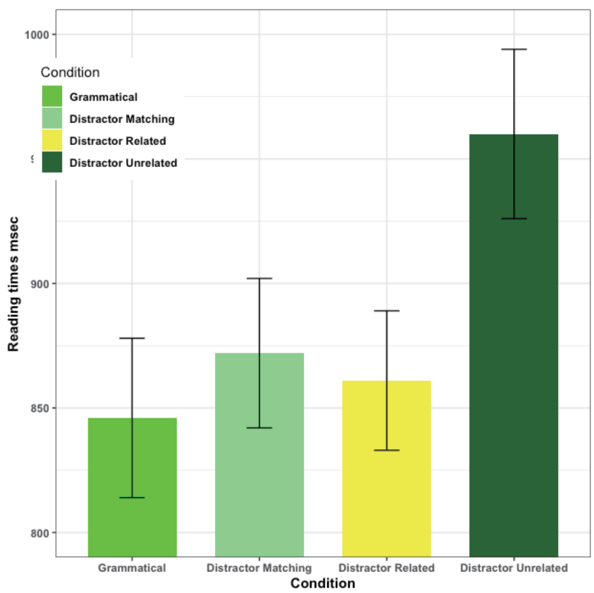


Fig. 2 Total reading times at distractor (R1)
(Error bars shows standard errors)

References

[1] Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. Journal of Memory and Language, 69(2), 85-103.
[2] Kwon, N., & Sturt, P. (2016). Attraction effects in honorific agreement in Korean. Frontiers in psychology, 7, 1302.
[3] McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. Journal of memory and language, 48(1), 67-91.
[4] Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. Journal of memory and language, 55(2), 157-166.
[5] Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. Journal of memory and language, 61(2), 206-237.

PS.3.12

Exploring Hindi: The Lexical Initiative

1. Vivek Singh Sikarwar- Indian Institute of Technology Kanpur , Kanpur, U.P., India

2. Ark Verma- Indian Institute of Technology Kanpur , Kanpur, U.P., India

Introduction: Understanding how Hindi speakers process written words is crucial for developing more effective educational tools and cognitive models. The Hindi Lexicon Project is an innovative research initiative aimed at mapping the cognitive processing of Hindi words through reaction time studies, providing foundational data for psycholinguistic research in Hindi. This abstract outline the methodology, findings, and implications of this ongoing extensive study.

Participant: We recruited 31 participants were native Hindi speakers and had normal or corrected to normal vision. Participants aged between 21 and 39 years (M=25.24, SD=5.62) and rated their proficiency (5-point scale) in Hindi as M=3.55, SD=0.67.

Method: Stimuli consisted of 11500 words selected from Shabd corpus (https://link.springer.com/article/10.3758/s13428-021-01625-2) , divided in 46 blocks drawn from a Hindi psycholinguistic corpus. Additionally, 11500 non-words were created by single substitution method. Participants responded by pressing 'm' for words and 'z' for non-words.

Result: Analysis was carried out only on data of 21 participants out of 31. 10 participants were removed due to partial and inconsistent data across various blocks. Validation of the collected data was done by testing effects of following factors. Significant effects of Word Frequency ($F(1,11222) = 1587.23$, $p < 0.001$), Contextual Diversity ($F(434, 10789) = 4.56$, $p < 0.001$), Akshar Count ($F(1, 11222) = 4015.46$, $p < 0.001$), Maatraa Count ($F(1, 11222) = 304.98$, $p < 0.001$), Total Length ($F(1, 11222) = 2519.53$, $p < 0.001$), Syllable Count ($F(1, 11200) = 2567.35$, $p < 0.001$), Phoneme Count ($F(1, 11200) = 3201.46$, $p < 0.001$) and Neighbourhood Density ($F(1, 11222) = 3494.09$, $p < 0.001$) were found on mean reaction time.

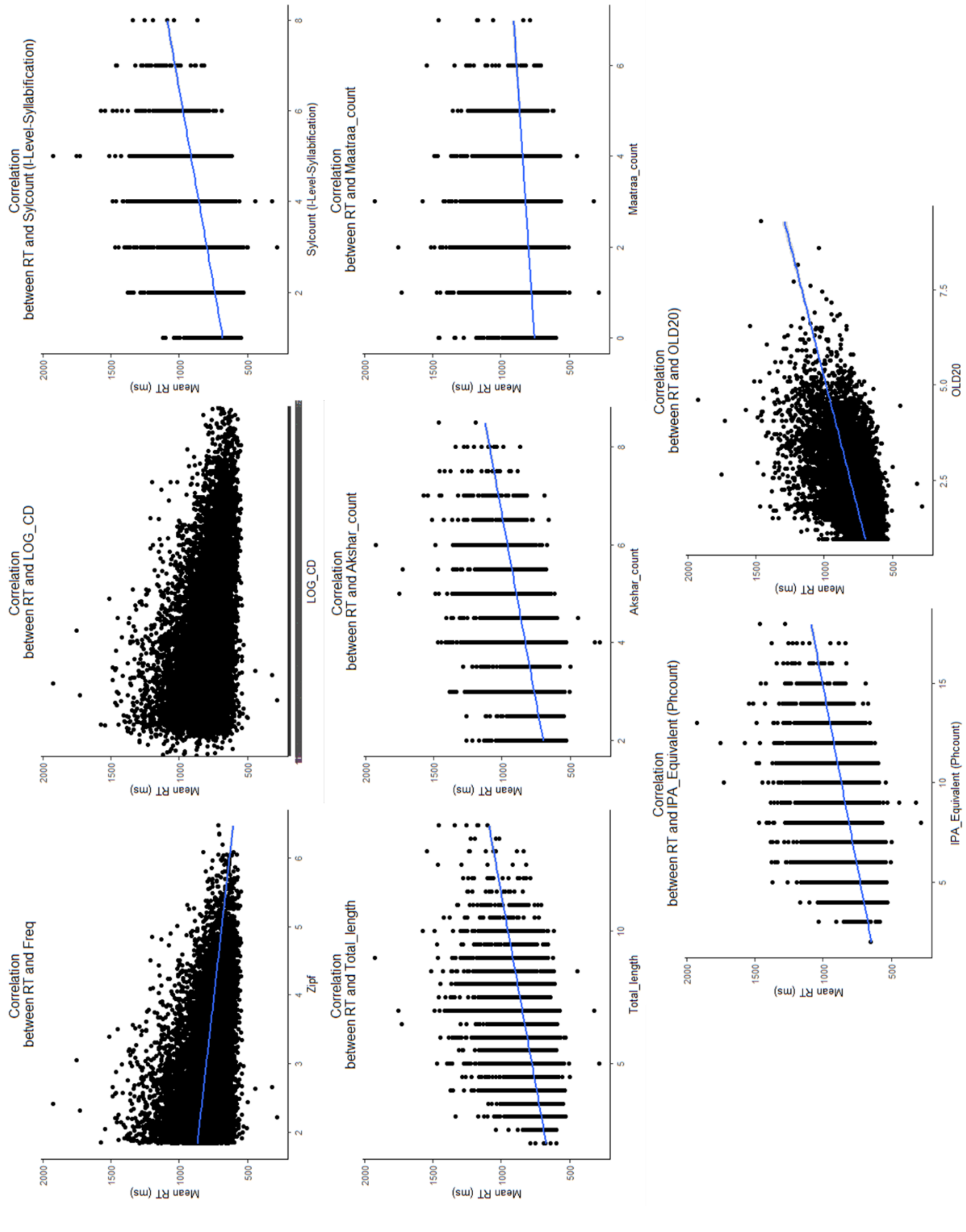
Word Frequency, Contextual Diversity, Akshar Count, Maatraa Count, Total Length, Syllable Count, Phoneme Count and Neighbourhood Density were found to be accountable for 12.4%, 15.5%, 26.4%, 2.7%, 18.3%, 18.7%, 22.2% and 23.7% respectively of variance of meant RT. Since, the data collection is still ongoing, complete data will give better insights.

Discussion: This project contributes to the broader understanding of visual word recognition in non-Western languages, providing empirical data that challenge and extend existing cognitive models primarily developed through research on Indo-European languages like English, Dutch, French etc. Our future work will focus on extending this dataset with eye-tracking measures to examine the fine-grained temporal dynamics of Hindi word recognition, and applying machine learning techniques to predict reaction times based on linguistic features.

Sample stimuli:

Words: रचाया, नौकरशाह, सुनार, कारीगर
Non-words: संशमी, सर्रावसान, नूफ, सुचेगी

Figures:



Additional information about Hindi Language:

Hindi differs from other languages in terms of how syllables and vowels are defined in the language. Consonants in the language have hidden vowels at their end, also called as schwa, for example, क /ka/ is just akshara in Hindi but can be considered as a whole syllable in some cases. However, in most cases, diacritic markers i.e. vowels are added to such aksharas or consonants to add to its phonology. Such markers are called matras and can be added to the aksharas in both linear and non-linear fashion, example (क+ा = का /kaa/ and क+ि = कि /ki/). Such markers (matras) can also be written to top or bottom of the consonant (akshara), each changing the pronunciation in different ways. Thus, we defined sub syllabic unit in Hindi as a combination of akshara (consonant) + matra (diacritic marker or vowel). (if matra is present), or just akshara (consonant) (if matra is absent).

References:

1. Keuleers, E. , Lacey, P. , Rastle, K. ,& Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words Behavior Research Methods 44 (1): 287–304. https://doi.org/10.3758/s13428-011-0118-4.
2. Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... & Treiman, R. (2007). The English lexicon project. Behavior Research Methods, 39(3), 445-459.
3. Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). The impact of word prevalence on lexical decision times: Evidence from the Dutch Lexicon Project 2. Journal of Experimental Psychology: Human Perception and Performance, 42, 441-458.
4. Sun, C. C., Hendrix, P., Ma, J., & Baayen, R. H. (2018). Chinese lexical database (CLD). Behavior resEarch Methods, 50(6), 2606-2629.
5. Verma, A. , Sikarwar, V. Yadav, H. Jaganathan, R., Kumar, P. (2022). Shabd: A psycholinguistic database for Hindi. Behavior Research Methods (2022) 54:830–844. https://doi.org/10.3758/s13428-021-01625-2

PS.3.13

Do Emoticons represent Face? An Experimental Study

Koyel Mukherjee (Centre for Linguistic Science and Technology, Indian Institute of Technology Guwahati), Email: koyel.mukherjee@iitg.ac.in.
Bidisha Som (Department of Humanities and Social Sciences, Indian Institute of Technology Guwahati),
Abhishek Shrivastava (Department of Design, Indian Institute of Technology Guwahati).

In Computer Mediated Communications (CMSs), emojis play a major role while providing pragmatic context to text messages. However, emojis are quite different from its precursor, ‘emoticons’ in terms of structure and design. This study explores the visual morphology that is present in emoticons, and analyses whether they are processed more like human faces or as lexical items in languages. 336 emoticons were generated randomly from a combination of 13 round structures (resembling face), 6 eye-structures, 7 mouth structures, and 8 up-fix structures (up-fix refers to drawings above the heads of comic characters which contribute to understanding of the emotive states or moods of a character). The stimuli were pre-processed where 20 anonymous volunteers labelled each generated emoticons as ‘congruent’ (meaningful), ‘incongruent’ (not meaningful due to mismatch or usage of opposing factors), and ‘ambiguous’ (where decisions could not be taken). A Lexical Decision Task equivalent was conducted using E-Prime 3.0 and data was collected from 92 participants (Age group: 18-35). Statistical analysis using MATLAB reveals a statistically significant difference between ‘congruent’ (M=1117.34; SD=704.18) and ‘incongruent’ emoticons (M=1158.64; SD= 754.02). Mann-Whitney U test results indicate p-value=0.03. The result mimics results in Lexical Decision Tasks in natural languages, indicating some grammatical constraints and a mental lexicon for visual items and their ‘lexicalization’. To further this finding, this experiment was followed by an Eye-Tracking study to check whether emoticons are processed similar to real faces (holistically) or similar to lexical items (in a componential manner). Areas of Interest (AOI) was marked on emoticons (‘eyes’, ‘mouth’, and ‘up-fix’) and on real faces (‘eyes’, and ‘mouth’). 200 face images were selected from AKDEF (an open-source database) that had a front profile as stimuli apart from the previously generated emoticons. 30 Participants (Age group: 18-25) were asked to categorize emoticons as ‘congruent’ and ‘incongruent’ and real faces as ‘positive expressions’ (happy, joyous, excited) and ‘negative expressions’ (sad, angry, disgusted). ‘Emoticons’ (M= 191.76; SD: 95.08) has a lower fixation duration than ‘Real face’ (M= 212.69; SD: 110.45). Mann-Whitney test across fixation duration reveals a statistically significant difference (p-value=0.000). Furthermore, the AOI ‘eye’ had the largest number of fixations in emoticons (28,460) followed by the ‘up-fix’ (13,675). This was indicative of a head driven structure mimicking a hierarchy-based word formation process. The results show that while emoticons are supposed to mimic facial expressions in CMCs, they have become lexicalized to become a part of the Visual Language used in multimodal communications. The results support contemporary claims on visual morphology as a part of the Visual Language. Supporting Jackandoff’s claim on Parallel Architectures, the emoticons are getting ‘lexicalized’ in Visual Language processing.

Figures:

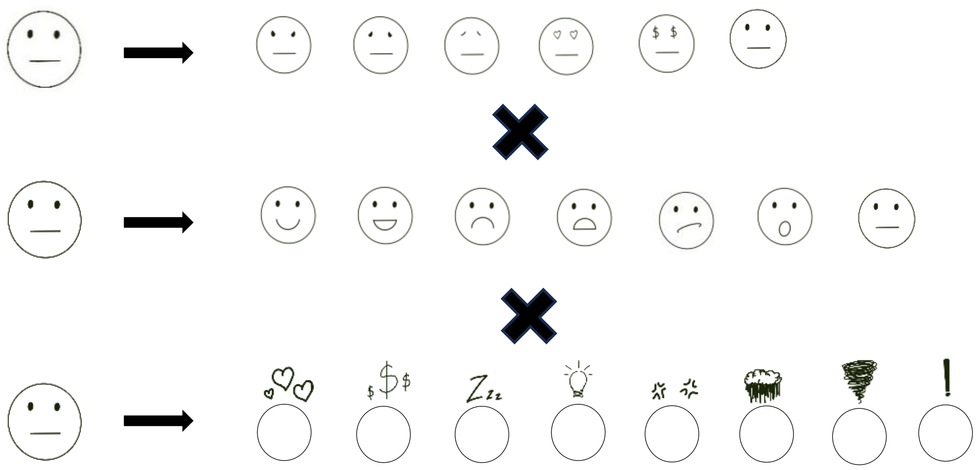


Fig. 1. Stimuli Generation for Emoticons

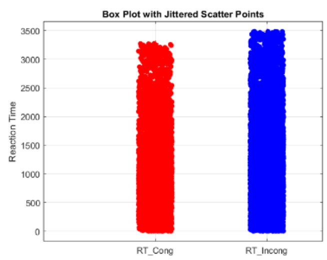


Fig. 2. Bar-plot (Emoticons Response Time).

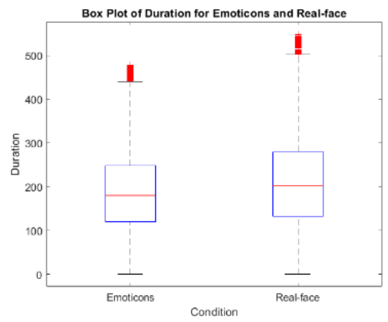


Fig. 3. Box-plot (Emoticons and real-face)

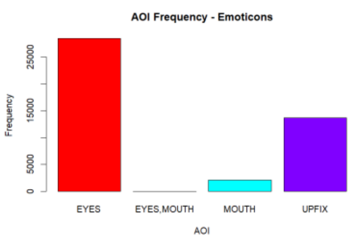


Fig. 4. AOI in Emoticons

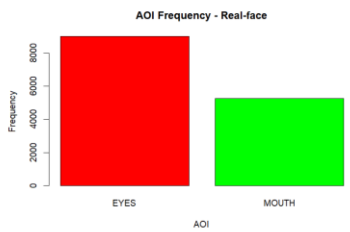


Fig. 5. AOI in Real-Face

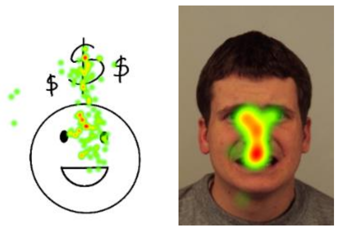


Fig. 6. Heatmaps of Eye-Tracking data on fixation points.

PS.3.14

Investigating speech segmentation: The language rhythm of Tamil

Poorani Vijayakumar, Laurence Bruggeman, Mark Antoniou *The MARCS Institute for Brain, Behaviour and Development* poorani.v@westernsydney.edu.au

Listeners rely on their native language's rhythm to segment speech (Cutler, 2012). Languages can generally be categorised into one of three language rhythm classes: stress-, syllable- or mora-timed[1]. Most speech segmentation research on language rhythm has investigated a restricted set of languages, such as stress-timed English (Cutler & Butterfield, 1992), syllable-timed French (Dumay, Frauenfelder & Content, 2002) and mora-timed Japanese (Otake, Hatano, & Yoneyama, 1996; Otake, Yoneyama, Cutler, & van der Lugt, 1996). Subsequent studies categorised more languages into either stress-timed (Dutch (Vroomen, van Zon, & de Gelder, 1996)) or syllable timed category (Spanish, Catalan (Pallier, Sebastián-Gallés, Felguera, Christophe, & Mehler, 1993 and Korean (Bond & Stockmal, 2002)). This was done by testing the rhythmic similarity hypothesis, according to which listeners can successfully segment speech in an unfamiliar language if that language has the same language rhythm as their native language rhythm (Cutler, 2012). However, Japanese has remained in the mora-timed category by itself, with rarely any language patterning similarly to it in segmentation experiments.

We propose investigating Tamil, a language from the family of Dravidian languages purported to possess similar rhythmic (i.e., moraic) properties like Japanese (e.g., Dutt, 1992; Ohno, 1981). There is some evidence that supports this: firstly, the syllable length of Tamil is calculated with a measure known as the maattirai, which generally corresponds to the concept of a mora (Zvelebil, 1989, pp.6). Secondly, Tamil was purported to neither syllable- nor stress-timed but showed indication that its syllable duration is correlated with a moraic structure (Balasubramaniam, 1980). Thirdly, Telugu, a Dravidian language, was found to be mora-timed as native Telugu and Japanese listeners segmented both languages successfully and similarly (Murty, Otake, & Cutler, 2007). Consequently, Tamil listeners may also successfully extend their segmentation strategies to Japanese (Keane, 2006).

Native Tamil listeners will complete a fragment detection task in stress-timed English, syllable-timed Korean, mora-timed Japanese, and their own native language of Tamil. Participants will be asked to detect a target fragment that is either aligned with or violates the segmentation boundaries of the target words presented to them. We hypothesise that Tamil listeners will segment Tamil and Japanese similarly: they will detect target fragments that align with the moraic boundaries of the target words presented more effectively that those that do not, thereby indicating that Tamil has mora-timed language rhythm. They will not segment English or Korean similarly to native listeners of either language. Preliminary results from this study will be presented. If the results confirm that Tamil is mora-timed, this will contribute to understanding language-specific aspects of speech segmentation. Additionally, investigating understudied languages, like Tamil, is important in determining universal aspects of speech segmentation of mora-timed languages.

[1] Initially, the importance of the rhythmic properties of a language for the strategies its listeners use during segmentation led to the assumption that all languages could be categorised into one of three rhythmic classes. Although this classification is now generally agreed to oversimplify a much more complex reality and is therefore considered problematic (White & Malisz, 2020), we opted to use these categorical terms in the interest of brevity.

References

Balasubramaniam, T. (1980). Timing in Tamil. *Journal of Phonetics,* 9, 151-161.

Bond, Z. S., & Stockmal, V. (2002). Distinguishing samples of spoken Korean from rhythmic and regional competitors. *Language Sciences, 24*(2), 175 – 185. https://doi.org/10.1016/S0388-0001(01)00013-4

Cutler, A. (2012). Native listening: Language experience and the recognition of spoken words. The MIT Press. https://doi.org/10.7551/mitpress/9012.001.0001

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language, 31*(2), 218 – 236. https://doi.org/10.1016/0749-596X(92)90012-M

Dumay, N., Frauenfelder, U. H., & Content, A. (2002). The role of the syllable in lexical segmentation in French: Word-spotting data. *Brain and Language, 81*(1-3), 144 – 161. https://doi.org/10.1006/brln.2001.2513

Dutt, N. V. B. S. (1992). *Japanese and Telugu easy learner*. Vishakapatnam.

Keane, E. (2006). Rhythmic characteristics of colloquial and formal Tamil. *Language and Speech*, 49(3), 299–332. https://doi.org/10.1177/00238309060490030101

Murty, L., Otake, T., & Cutler, A. (2007). Perceptual tests of rhythmic similarity: I. Mora rhythm. *Language and Speech*, *50*(1), 79-99. https://doi.org/10.1177/00238309070500010401

Ohno, S. (1981). *Nihongo to Tamiirugo.* (Japanese and Tamil). Tokyo: Shinchoosya.

Otake, T., Hatano, G., & Yoneyama, K. (1996). Speech segmentation by Japanese listeners. In T. Otake & A. Cutler (Eds.), *Phonological structure and language processing: Cross-linguistic studies* (pp.183 – 201). Berlin: Mouton. https://doi.org/10.1515/9783110815825.183

Otake, T., Yoneyama, K., Cutler, A., & van der Lugt, A. (1996). The representation of Japanese moraic nasals. *Journal of the Acoustical Society of America*, *100*(6), 3831–3842. https://doi.org/10.1121/1.417239

Pallier, C., Sebastián-Gallés, N., Felguera, T., Christophe, A., & Mehler, J. (1993). Attentional allocation within the syllabic structure of spoken words. *Journal of Memory and Language, 32,* 373 – 389. https://doi.org/10.1006/jmla.1993.1020

White, L., & Malisz, Z. (2020). Speech rhythm and timing. In C. Gussenhoven & A. Chen (Eds.), *Oxford Handbook of Language Prosody* (pp. 167-182). Oxford: Oxford University Press.

Zvelebil, K. V. (1989). *Classical Tamil Prosody: An Introduction.* Madras: New Era.

Background on Tamil

Tamil has 53 million speakers in Tamil Nadu, India as well as the diaspora including, Singapore, Sri Lanka, and Malaysia. Tamil is known to be highly diglossic, and the most dialectally variable of the Dravidian languages (Keane, 2004, 2006). The colloquial register is characterised by regional and societal variation, differing from the formal register in several ways. For example, all colloquial words are vowel-final; word-final nasals do not occur due to deleting word-final consonants and adding "enunciative" vowels to other final consonants, thereby resulting in less complex consonant clusters than in formal Tamil (Keane, 2006). The formal register, on the other hand, has not evolved much from its thirteenth century literary standards, making it mostly standardised amongst all Tamil speakers.

Attempts to categorise Tamil according to language rhythm has yielded mixed findings with descriptions including, syllable-timed (Ravisankar, 1994, pp.337, as cited in Keane, 2006), stress-timed (Marthandan, 1983, pp.308, as cited in Keane, 2006), not stress- nor syllable timed (Balasubramaniam, 1980, pp.466), and mora-timed (Ramus, Nespor & Mehler, 1999, pp. 266). Studies using phonetic analyses to classify Tamil have also produced mixed results with no clear consensus (Grabe, 2002; Grabe & Low, 2002).

References

Balasubramaniam, T. (1980). Timing in Tamil. *Journal of Phonetics, 8,* 449-467.

Grabe, E. (2002). Variation adds to prosodic typology. In B. Bel & I. Marlin (Eds.), *Proceedings of Speech Prosody 2002* (pp.127 – 132). Aix-en-Provence: Laboratoire Parole et Langage.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp.515 – 546). Berlin: Mouton.

Keane, E. (2004). Illustrations of the IPA: Tamil. *Journal of the International Phonetic Association, 34*(1), 111 – 116. https://doi.org/10.1017/S0025100304001549

Keane, E. (2006). Rhythmic characteristics of colloquial and formal Tamil. *Language and Speech, 49*(3), 299–332. https://doi.org/10.1177/00238309060490030101

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73*(3), 265 – 292. https://doi.org/10.1016/S0010-0277(00)00101-3

PS.3.15

Processing Expletiveness: An eye-tracking study on negation

Anna Teresa Porrini, Veronica D'Alesio, Matteo Greco

Previous literature on the psycholinguistics of negation showed that standard negation (SN), under certain circumstances, leads to inflated processing costs (Kaup & Dudschig, 2020). However, less attention has been devoted to expletive negation (EN), referring to those negative forms that do not provide any truth-conditional contribution to interpretation (Delfitto, 2020), such as in the Italian sentence "Starò qui finchè **non** arriva Gianni" (lit. stay.Fut.1stsg here until **neg** arrives John, 'I'll stay here until John arrives'). Greco et al. (2020) utilized the visual world paradigm to examine a specific type of EN, i.e., Surprise Negation sentences, observing fixation patterns similar to those of affirmative sentences. However, to our knowledge, no study has addressed the processing of expletive negation in reading.

This study aims to assess the processing cost associated with the comprehension of EN following the conjunction "finché" in Italian (see the example below). It follows up two previous experiments using different methodologies, an acceptability judgment task (80 participants), and a separate self-paced reading task (80 participants). While results from the former paradigm unexpectedly suggested that EN sentences are perceived as more natural and understandable than their affirmative counterparts, the latter showed that reading times increase significantly in the presence of a negative particle and on the chunk immediately following it, regardless of whether the negation is expletive. This outcome suggests that while sounding perfectly natural to native speakers, EN still entails higher processing costs.

One limit of the masked self-paced reading paradigm is that it doesn't allow for rereading, thus addressing processing solely in its earlier stages. Therefore, we implemented an eye-tracking experiment within the reading paradigm, aiming to investigate the later stage of processing as well, considered to be reflective of semantic and syntactic integration (Clifton et al. 2007). Specifically, we seek to address two main research questions: i) Are there any differences between the processing of SN and EN? ii) How do different degrees of naturalness influence the processing of EN compared to their affirmative counterparts? The experiment comprises 28 critical trials and 68 filler trials, and sentences are presented on a single line of text. The items were adapted from our two previous experiments and are shown in 4 different conditions: EN, SN, and their respective Affirmative counterparts (A_EN, A_SN). An example of experimental items can be found below. Crucially, sentences in the EN and A_EN condition have the same meaning, as the negation is expletive.

The data collection is currently in progress. The data will be analysed with linear mixed models and results will be discussed in light of previous literature on the psycholinguistics of negation, along with theoretical and methodological implications for future research.

Example sentences:

EN: Chiara è rimasta in casa finché Marco non ha chiamato la pizzeria.

Chiara remained in the house until Marco (NOT) called the pizzeria.

A_EN: Chiara è rimasta in casa finché Marcello ha chiamato la pizzeria.

Chiara remained in the house until Marcello called the pizzeria.

SN: Chiara è rimasta in casa perché Marco non ha chiamato la pizzeria.

Chiara remained in the house because Marco did NOT call the pizzeria.

A_SN: Chiara è rimasta in casa perché Marcello ha chiamato la pizzeria.

Chiara remained in the house because Marcello called the pizzeria.

References:

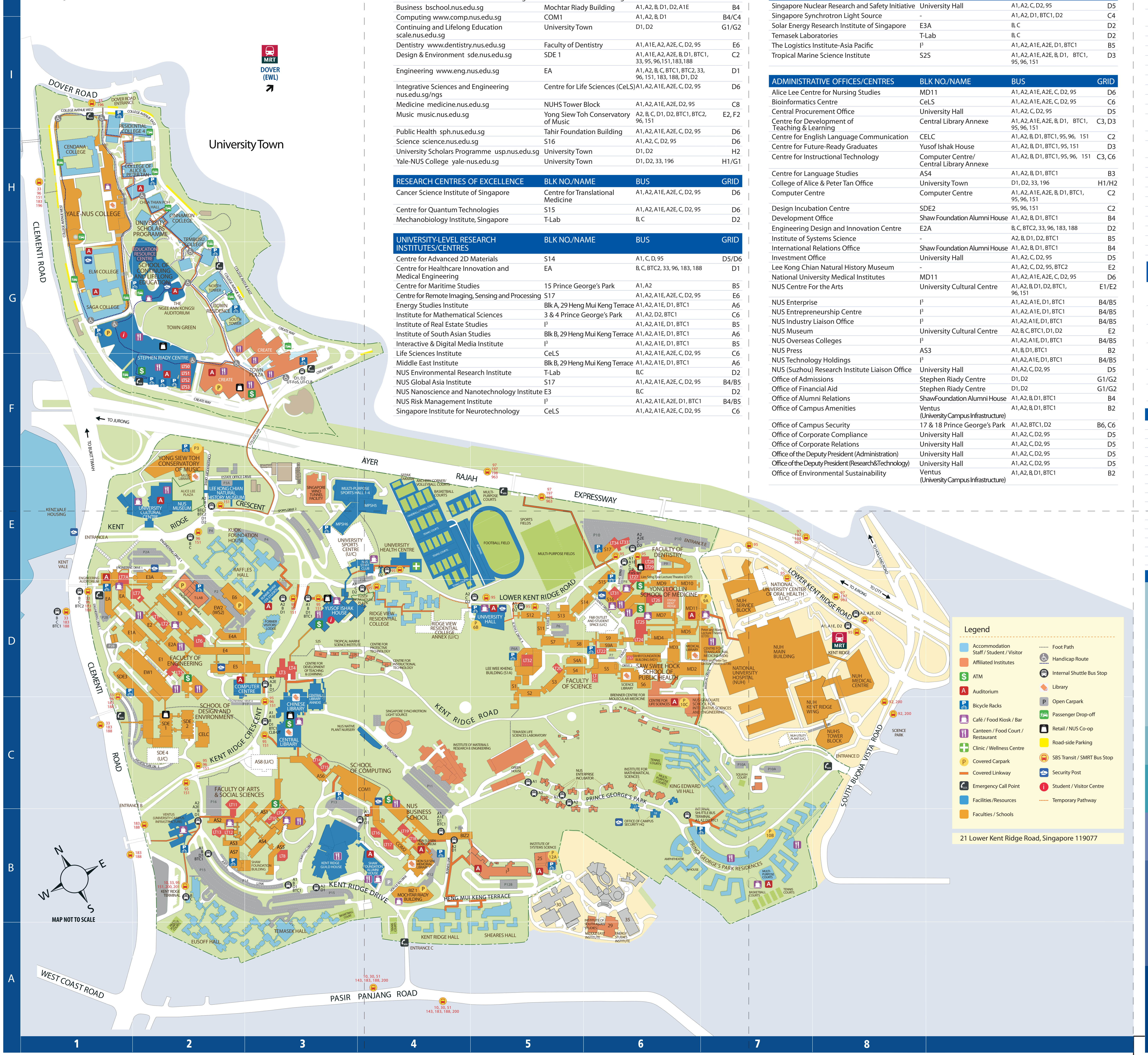
Clifton , C. Jr , Staub , A. , & Rayner , K . (2007). Eye movements in reading words and sentences . In: *Eye movement research: Insights into mind and brain* (eds. Gompel, R.V., Fisherand, M., Murray, W., & Hill, R.L.).

Delfitto, D. (2020). Expletive negation. In: *The Oxford Handbook of Negation* (eds. Deprez, V., & Espinal, M. T.).

Greco M., Canal, P., Bambini, V., & Moro, A. (2020). Modulating “Surprise” with syntax: A study on negative sentences and eye-movement recording. *Journal of Psycholinguistic Research, 49*.

Kaup, B., & Dudschig, C. (2020). Understanding negation: Issues in the processing of negation. In: *The Oxford Handbook of Negation* (eds. Deprez, V., & Espinal, M. T.).

Kent Ridge Campus



KENT RIDGE CAMPUS LISTING

SCHOOLS	BLK NO./NAME	BUS	GRID
Arts & Social Sciences www.fas.nus.edu.sg	Shaw Foundation Building	A1, A2, B, D1, BTC1	B2
Business bschool.nus.edu.sg	Mochtar Riady Building	A1, A2, B, D1, D2, A1E	B4
Computing www.comp.nus.edu.sg	COM1	A1, A2, B, D1	B4/C4
Continuing and Lifelong Education scale.nus.edu.sg	University Town	D1, D2	G1/G2
Dentistry www.dentistry.nus.edu.sg	Faculty of Dentistry	A1, A1E, A2, A2E, C, D2, 95	E6
Design & Environment sde.nus.edu.sg	SDE 1	A1, A1E, A2, A2E, B, D1, BTC1, 33, 95, 96,151,183,188	C2
Engineering www.eng.nus.edu.sg	EA	A1, A2, B, C, BTC1, BTC2, 33, 96, 151, 183, 188, D1, D2	D1
Integrative Sciences and Engineering nus.edu.sg/ngs	Centre for Life Sciences (CeLS)	A1, A2, A1E, A2E, C, D2, 95	D6
Medicine medicine.nus.edu.sg	NUHS Tower Block	A1, A2, A1E, A2E, D2, 95	C8
Music music.nus.edu.sg	Yong Siew Toh Conservatory of Music	A2, B, C, D1, D2, BTC1, BTC2, 96, 151	E2, F2
Public Health sph.nus.edu.sg	Tahir Foundation Building	A1, A2, A1E, A2E, C, D2, 95	D6
Science science.nus.edu.sg	S16	A1, A2, C, D2, 95	D6
University Scholars Programme usp.nus.edu.sg	University Town	D1, D2	H2
Yale-NUS College yale-nus.edu.sg	University Town	D1, D2, 33, 196	H1/G1

RESEARCH CENTRES OF EXCELLENCE	BLK NO./NAME	BUS	GRID
Cancer Science Institute of Singapore	Centre for Translational Medicine	A1, A2, A1E, A2E, C, D2, 95	D6
Centre for Quantum Technologies	S15	A1, A2, A1E, A2E, C, D2, 95	D6
Mechanobiology Institute, Singapore	T-Lab	B, C	D2

UNIVERSITY-LEVEL RESEARCH INSTITUTES/CENTRES	BLK NO./NAME	BUS	GRID
Centre for Advanced 2D Materials	S14	A1, C, D, 95	D5/D6
Centre for Healthcare Innovation and Medical Engineering	EA	B, C, BTC2, 33, 96, 183, 188	D1
Centre for Maritime Studies	15 Prince George's Park	A1, A2	B5
Centre for Remote Imaging, Sensing and Processing	S17	A1, A2, A1E, A2E, C, D2, 95	E6
Energy Studies Institute	Blk A, 29 Heng Mui Keng Terrace	A1, A2, A1E, D1, BTC1	A6
Institute for Mathematical Sciences	3 & 4 Prince George's Park	A1, A2, D2, BTC1	C6
Institute of Real Estate Studies	I^3	A1, A2, A1E, D1, BTC1	B5
Institute of South Asian Studies	Blk B, 29 Heng Mui Keng Terrace	A1, A2, A1E, D1, BTC1	A6
Interactive & Digital Media Institute	I^3	A1, A2, A1E, D1, BTC1	B5
Life Sciences Institute	CeLS	A1, A2, A1E, A2E, C, D2, 95	C6
Middle East Institute	Blk B, 29 Heng Mui Keng Terrace	A1, A2, A1E, D1, BTC1	A6
NUS Environmental Research Institute	T-Lab	B,C	D2
NUS Global Asia Institute	S17	A1, A2, A1E, A2E, C, D2, 95	B4/B5
NUS Nanoscience and Nanotechnology Institute	E3	B,C	D2
NUS Risk Management Institute	I^3	A1, A2, A1E, A2E, D1, BTC1	B4/B5
Singapore Institute for Neurotechnology	CeLS	A1, A2, A1E, A2E, C, D2, 95	C6
Singapore Nuclear Research and Safety Initiative	University Hall	A1, A2, C, D2, 95	D5
Singapore Synchrotron Light Source	-	A1, A2, D1, BTC1, D2	C4
Solar Energy Research Institute of Singapore	E3A	B, C	D2
Temasek Laboratories	T-Lab	B, C	D2
The Logistics Institute-Asia Pacific	I^3	A1, A2, A1E, A2E, D1, BTC1	B5
Tropical Marine Science Institute	S2S	A1, A2, A1E, A2E, B, D1, BTC1, 95, 96, 151	D3

ADMINISTRATIVE OFFICES/CENTRES	BLK NO./NAME	BUS	GRID
Alice Lee Centre for Nursing Studies	MD11	A1, A2, A1E, A2E, C, D2, 95	D6
Bioinformatics Centre	CeLS	A1, A2, A1E, A2E, C, D2, 95	C6
Central Procurement Office	University Hall	A1, A2, C, D2, 95	D5
Centre for Development of Teaching & Learning	Central Library Annexe	A1, A2, A1E, A2E, B, D1, BTC1, 95, 96, 151	C3, D3
Centre for English Language Communication	CELC	A1, A2, B, D1, BTC1, 95, 96, 151	C2
Centre for Future-Ready Graduates	Yusof Ishak House	A1, A2, B, D1, BTC1, 95, 151	D3
Centre for Instructional Technology	Computer Centre/ Central Library Annexe	A1, A2, B, D1, BTC1, 95, 96, 151	C3, C6
Centre for Language Studies	AS4	A1, A2, B, D1, BTC1	B3
College of Alice & Peter Tan Office	University Town	D1, D2, 33, 196	H1/H2
Computer Centre	Computer Centre	A1, A2, A1E, A2E, B, D1, BTC1, 95, 96, 151	C2
Design Incubation Centre	SDE2	95, 96, 151	C2
Development Office	Shaw Foundation Alumni House	A1, A2, B, D1, BTC1	B4
Engineering Design and Innovation Centre	E2A	B, C, BTC2, 33, 96, 183, 188	D2
Institute of Systems Science	-	A2, B, D1, D2, BTC1	B5
International Relations Office	Shaw Foundation Alumni House	A1, A2, B, D1, BTC1	B4
Investment Office	University Hall	A1, A2, C, D2, 95	D5
Lee Kong Chian Natural History Museum	-	A1, A2, C, D2, 95, BTC2	E2
National University Medical Institutes	MD11	A1, A2, A1E, A2E, C, D2, 95	D6
NUS Centre For the Arts	University Cultural Centre	A1, A2, B, D1, D2, BTC1, 96, 151	E1/E2
NUS Enterprise	I^3	A1, A2, A1E, D1, BTC1	B4/B5
NUS Entrepreneurship Centre	I^3	A1, A2, A1E, D1, BTC1	B4/B5
NUS Industry Liaison Office	I^3	A1, A2, A1E, D1, BTC1	B4/B5
NUS Museum	University Cultural Centre	A2, B, C, BTC1, D1, D2	E2
NUS Overseas Colleges	I^3	A1, A2, A1E, D1, BTC1	B4/B5
NUS Press	AS3	A1, B, D1, BTC1	B2
NUS Technology Holdings	I^3	A1, A2, A1E, D1, BTC1	B4/B5
NUS (Suzhou) Research Institute Liaison Office	University Hall	A1, A2, C, D2, 95	D5
Office of Admissions	Stephen Riady Centre	D1, D2	G1/G2
Office of Financial Aid	Stephen Riady Centre	D1, D2	G1/G2
Office of Alumni Relations	ShawFoundation Alumni House	A1, A2, B, D1, BTC1	B4
Office of Campus Amenities	Ventus (University Campus Infrastructure)	A1, A2, B, D1, BTC1	B2
Office of Campus Security	17 & 18 Prince George's Park	A1, A2, BTC1, D2	B6, C6
Office of Corporate Compliance	University Hall	A1, A2, C, D2, 95	D5
Office of Corporate Relations	University Hall	A1, A2, C, D2, 95	D5
Office of the Deputy President (Administration)	University Hall	A1, A2, C, D2, 95	D5
Office of the Deputy President (Research&Technology)	University Hall	A1, A2, C, D2, 95	D5
Office of Environmental Sustainability	Ventus (University Campus Infrastructure)	A1, A2, B, D1, BTC1	B2
Office of Estate Development	Ventus (University Campus Infrastructure)	A1, A2, B, D1, BTC1	B2
Office of Facilities Management	Ventus (University Campus Infrastructure)	A1, A2, B, D1, BTC1	B2
Office of Financial Services	University Hall	A1, A2, C, D2, 95	D5
Office of Housing Services	Kent Vale	C, B, BTC1	E1
Office of Human Resources	University Hall	A1, A2, C, D2, 95	D5
Office of Internal Audit	University Hall	A1, A2, C, D2, 95	D5
Office of Legal Affairs	University Hall	A1, A2, C, D2, 95	D5
Office of the President	University Hall	A1, A2, C, D2, 95	D5
Office of the Provost	University Hall	A1, A2, C, D2, 95	D5
Office of Resource Planning	University Hall	A1, A2, C, D2, 95	D5
Office of Risk Management	University Hall	A1, A2, C, D2, 95	D5
Office of Safety, Health & Environment	Ventus (University Campus Infrastructure)	A1, A2, B, D1, BTC1	B2
Office of Student Affairs	Yusof Ishak House	A1, A2, B, D1, BTC1, 95, 151, C	D3
Office of theVice President (Campus Infrastructure)	Ventus (University Campus Infrastructure)	A1, A2, B, D1, BTC1	B2
Office of the Vice President (University & Global Relations)	University Hall	A1, A2, C, D2, 95	D5
Office of the Professional Engineering & Executive Education	E1	B, C	D2
Organisational Excellence Office	University Hall	A1, A2, C, D2, 95	D5
Registrar's Office	University Hall	A1, A2, C, D2, 95	D5
Student Service Centre	Yusof Ishak House	A1, A2, B, D1, BTC1, 95, 151, C	D3
Temasek Defence Systems Institute	E1	B, C	D2
University Health Centre	-	A1, A2, C, D2, 95	D4
University Town Management Office	Stephen Riady Centre	D1, D2	G1
Visitors Centre	Stephen Riady Centre	D1, D2	G1
WHO Collaborating Centre for Occupational Health	NUH	A1, A2, 95, D2, A2E	C7

HALLS OF RESIDENCES	BUS	GRID	RESIDENTIAL COLLEGES	BUS	GRID	STUDENT RESIDENCES	BUS	GRID
Eusoff Hall	A2, B	A2	Cinnamon College	D1, D2	H2	Kuok Foundation House	A2, B, C, BTC1, BTC2, D1, D2, 96, 151	E2
Kent Ridge Hall	A1, A2, D1, BTC1	A4	College of Alice & Peter Tan	D1, D2, 33, 196	H1/H2	Prince George's Park Residences	A1, A2, BTC1, D2	B6/B7
King Edward VII Hall	A1, A2, BTC1, D2	C6/B7	Residential College 4	D1, D2, 33, 196	H1/H2/ I1/I2	UTown Residence	D1, D2	G2
Sheares Hall	A1, A2, D1, BTC1	A5	Ridge View Residential College	A1, A2, C, D2, 95	D4			
Temasek Hall	A2, B	A3	Tembusu College	D1, D2	G2/H2			
Raffles Hall	A1, A2, B, D, 95	D2						

LECTURE THEATRES	GRID	LECTURE THEATRES	GRID	LECTURE THEATRES	GRID
LT 1	D2	LT 16	B4	LT 33	E6
LT 2	D2	LT 17	B4	LT 34	E6
LT 3	C3/C4	LT 18	B4	LT 35 (Peter & Mary Fu Lecture Theatre)	D6
LT 4	C3/C4	LT 19	B4	LT 36 (Alice & Peter Tan Lecture Theatre)	D6
LT 5	D2	LT 20	D5	LT 37	D6
LT 6	D2	LT 21	C6	LT 50	F2
LT 7	D1	LT 22	D6	LT 51	F2
LT 7A	D1	LT 23	D6	LT 52	F2
LT 8	B3	LT 24	D6	LT 53	F2
LT 9	B3	LT 25	D6	Engineering Auditorium	D1
LT 10	B3	LT 26	B6	Hon Sui Sen Auditorium	B4
LT 11	B2	LT 27 (Lim Seng Tjoe Lecture Theatre)	B6	The Ngee Ann Kongsi Auditorium	G2
LT 12	B2	LT 28	D6	UTown Auditorium 1	F2
LT 13 (NUS Theatrette)	B2	LT 29	D6	UTown Auditorium 2	F2
LT 14	C3	LT 31 (Science Auditorium)	D6	UTown Auditorium 3	H2
LT 15	C3	LT 32	D5		

	CAMPUS DINING	LOCATION	GRID
CAFÉS	Butter My Buns	Inside Bookhaven, Stephen Riady Centre	G2
	Dôme	University Cultural Centre	E2
	Fortune Village	Stephen Riady Centre	F2
	Good News Café	EA	D1
	Humble Origins	AS2, Faculty of Arts and Social Sciences	B2
	Humble Origins	Ventus (University Campus Infrastructure)	B2
	Platypus Food Bar	S16, Faculty of Science	D6
	Reedz Café	Mochtar Riady Building	B4
	Route 77 Café	Prince George's Park Residences	B7
	Spinelli	University Hall	D5
	Starbucks	Yusof Ishak House	D3
	Starbucks	MD11	D7
	Starbucks	Education Resource Centre, University Town	G2
	Supersnacks	Prince George's Park Residences	B7
	Waa Cow	Shaw Foundation Alumni House	B4
CANTEENS	Frontier	Faculty of Science	D6
	Techno Edge	Faculty of Engineering	D2
	The Deck	Faculty of Arts & Social Sciences	B3
	The Terrace	School of Computing / NUS Business School	C4
FAST FOOD	McDonald's	Techno Edge Annexe, Faculty of Engineering	D2
	Subway	Yusof Ishak House	D3
	Subway	Town Plaza, University Town	F3
	Subway Mobile Kiosk	E4, Faculty of Engineering	D2
	Subway Mobile Kiosk	Outside LT25, Faculty of Science	D6
FOOD COURTS	E-Canteen	Prince George's Park Residence	B7
	Flavours@UTown	Stephen Riady Centre	F2
	Food Junction	Yusof Ishak House	D3
	Koufu	Town Plaza, University Town	F2
	Pines Food Court	Prince George's Park Residence	B7
KIOSKS	7-Eleven	Yusof Ishak House	D3
	Cheers	Town Plaza, University Town	F3
	Des The Place	Blk E2A, Faculty of Engineering	D2
	Dilys Creation	E4, Faculty of Engineering	D2
	Dilys Creation	S16, Faculty of Science	D6
	Indonesian Express	Deck Level 2	B3
	Liang Ban Kung Fu	Deck Level 2	B3
	Old Chang Kee	Yusof Ishak House	D3
	Platypus FoodBar Capsule	Blk E2A, Faculty of Engineering	D2
	Sapore Italiano	Deck Level 2	B3
	Sarpino's Express	Yusof Ishak House	D3
RESTAURANTS	Alcove. Asian Restaurant. Bar	University Cultural Centre	E2
	A.R.T Food House	E2, Faculty of Engineering	D2
	Hwang's Korean Restaurant	Town Plaza, University Town	F3
	Dilys Creation	MD6, Centre For Translational Medicine	D7
	Pizza Hut	Stephen Riady Centre	G2
	Sapore	Town Plaza, University Town	G2
	Spice Table By Pines	Town Plaza, University Town	F3
	The Royals Bistro	Town Plaza, University Town	F3
	The University Club	Shaw Foundation Alumni House	B4

As at January 2016