AttLis 2022

The Attentive Listener in the Visual World

Storrs, CT, USA **October 27-28**

AttLis addresses multi-modal cognition with an emphasis on the interaction between language and vision. More specifically, there is a central focus on how attentional and visual processes interact with spoken and written language processing. Why are attention and vision crucial to language comprehension? How does each inform and mediate the other in moment-to-moment processing and in language development over the lifespan?

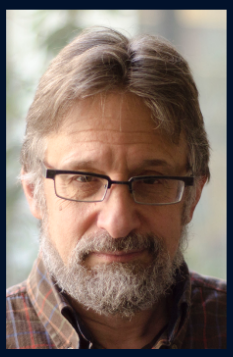
AttLis is supported in part by UConn's NSF-funded training program in *Science of Learning & Art of Communication (SLAC)* and the EDULANG project between NTNU (Trondheim) and UConn.

INVITED SPEAKERS

Elika Bergelson
Duke University

Craig Chambers
U. Toronto

Falk Huettig
Max Planck Institute for Psycholinguistics & Radboud University



Mike Tanenhaus
U. Rochester

Mila Vulchanova
Norwegian U. of Science & Technology (NTNU)

https://slac.uconn.edu/attlis-2022

UCONN
SCIENCE OF LEARNING & ART OF COMMUNICATION

LOCATIONS FOR ATTLIS & EDULANG

Dodd Center

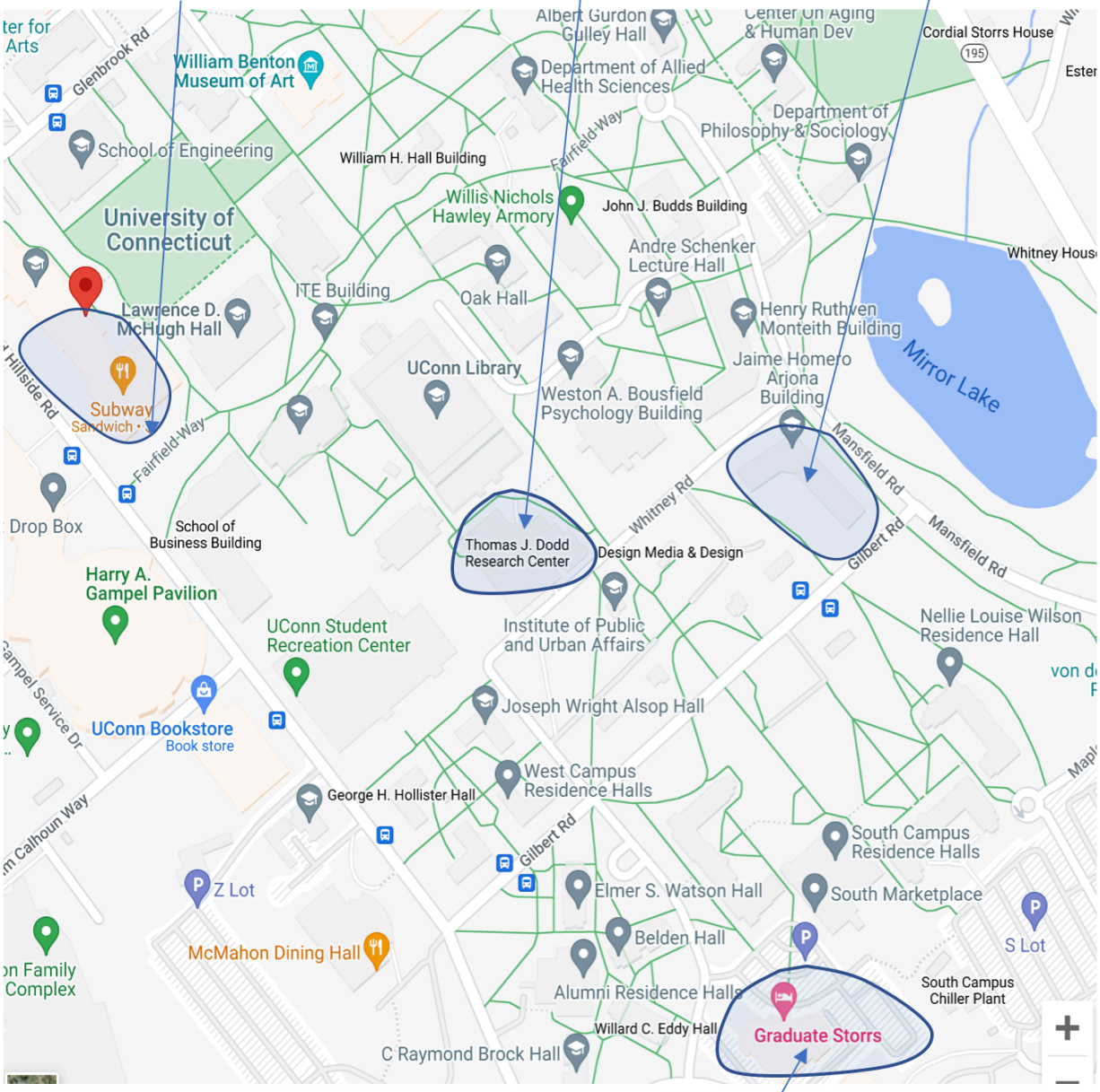
AttLis spoken sessions

Student Union

AttLis lunch (rm. 104) on 10/27
EDULANG workshop 10/25-26

Arjona

EDULANG workshop afternoon of 10/26



The Graduate

Campus hotel;
AttLis dinner 10/27

AttLis 2022 Code of Conduct*

AttLis is dedicated to providing a harassment-free conference experience for all, regardless of gender, gender identity and expression, sexual orientation, disability, physical appearance, body size, race, age, religion, or nationality. We will not tolerate harassment of conference participants in any form. Conference participants violating these rules may be sanctioned or expelled from the conference without a refund, at the discretion of the conference organizers. Harassment includes, but is not limited to:

- Verbal comments that reinforce social structures of domination related to gender, gender identity and expression, sexual orientation, disability, physical appearance, body size, race, age, religion, nationality
- Sexual images in public spaces
- Deliberate intimidation, stalking, or following
- Behaviors intended to make a person or group feel unwelcome, or to encourage ostracism of any individual or group
- Harassing photography or recording
- Sustained disruption of talks or other events
- Inappropriate physical contact
- Unwelcome sexual attention
- Advocating for, or encouraging, any of the above behaviors
- These behaviors are unacceptable both during in-person interactions and on social media.

Enforcement

Participants asked to stop any harassing behavior are expected to comply immediately. If a participant engages in harassing behavior, event organizers retain the right to take any actions to keep AttLis a welcoming environment for all participants. This includes warning the offender, expulsion from the conference with no refund, barring from participation in future conferences or their organization, reporting the incident to the offender's local institution or funding agencies, or reporting the incident to local law enforcement. Organizers may take action to redress anything with the intent or clear impact of disrupting the event or making the environment hostile for any participants. We expect participants to follow these rules at all event venues and event-related social activities. We think people should follow these rules outside event activities too!

Reporting

If someone makes you or anyone else feel unsafe or unwelcome, please report it to conference staff as soon as possible (Jim Magnuson or another UConn faculty member). Harassment and other code of conduct violations reduce the value of our event for everyone. People like you make our scientific community a better place, and we want you to be happy here. You can find University of Connecticut guidelines on discrimination and harassment, as well as a form for reporting incidents that are harmful to members of the community, at this link: https://equity.uconn.edu/discrimination/

Our team will be happy to help you contact hotel/venue security, local law enforcement, local support services, provide escorts, or otherwise assist you to feel safe for the duration of the event. We value your attendance.

Contact information

Should you wish to report an incident, you may contact any UConn faculty in attendance, or otherwise get in touch via e-mail or phone:

- Faculty organizer: Jim Magnuson, 860-617-0853, james.magnuson@uconn.edu
- UConn police:
    - (860) 486-4800 (for non-emergencies)
    - 911 (in case of an emergency)
- UConn Title IX office resources: https://titleix.uconn.edu/get-help/

* *This code has been adapted with permission from:* https://www.umass.edu/linguistics/cuny2020/coc/

COVID POLICIES

COVID is still highly prevalent in Connecticut. Many attendees have expressed concern about COVID policies because they or someone they are in regular contact with is immunocompromised.

Masks. We request that you wear a mask when you are not presenting a talk, except when participating in refreshment breaks or meals. We will have a supply of masks on hand.

Vaccination. The University of Connecticut requires all its personnel to be up to date with COVID vaccinations and boosters. We cannot require that of visitors, but we encourage you to get vaccinated and/or boosted. If you are not vaccinated, we ask that you be vigilant about using a mask, and that you maintain a distance of 2 meters / 6 feet from other participants at all times.

Filtration. We have 3 DIY filtration devices (Corsi-Rosenthal boxes, https://edgecollective.io/airbox) that we will **try** to bring along to all events.

COVID tests. We have a supply of COVID tests. If you have symptoms, we ask that you go to your lodging, and we will arrange to get one or more tests to you. We will do our best to assist you if you need to extend your stay and/or change transportation plans.

LOCAL CONTACT INFORMATION

If you find yourself in a situation where you need assistance or advice, call Jim Magnuson at 860-617-0853.

For non-emergencies requiring police advice, call the UConn police non-emergency line: 860-486-4800.

For emergencies, dial 911.

UCONN LAND ACKNOWLEDGEMENT

We acknowledge that the land on which we gather is the territory of the Mohegan, Mashantucket Pequot, Eastern Pequot, Schaghticoke, Golden Hill Paugussett, Nipmuc, and Lenape Peoples, who have stewarded this land throughout the generations. We thank them for their strength and resilience in protecting this land, and aspire to uphold our responsibilities according to their example.

BRIEF SCHEDULE

Thursday, October 27

8:15	**Breakfast & registration**
8:50	Opening remarks
9:00	***KEYNOTE*** **The task-based visual world paradigm:** **Past (history and context), present, and (perhaps) future** Michael K. Tanenhaus
10:00	**Effect of coordination on perspective-taking: Evidence from eye-tracking** Yipu Wei, Yingjia Wan, & Michael K. Tanenhaus
10:25	**Break -- coffee & tea**
10:40	**Fixations reveal preference: A VWP study of thematic role processing in Spanish** Beatriz Gómez-Vidal, Mirin Arantzeta, & Itziar Laka
11:05	**The use of prosodic cues in syntactic attachment: evidence from eye movements** Aline Fonseca, Andressa Silva, & Gabriela Rodrigues
11:30	***KEYNOTE*** **The nascent lexicon: Early word comprehension in the lab and in the world** Elika Bergelson
12:30	**Lunch** *Served in **Student Union room 331** -- allow 5-10 minutes to walk each way, so please head back to Dodd by 13:50. Afternoon breaks will not have refreshments, so consider bringing a drink or snack with you from lunch.*
14:00	**Tracking Audio-visual Interaction for Semantic Representation Using FrameNet** Frederico Belcavello & Tiago Torrent
14:25	**The role of attention during grammatical gender processing in low and high literacy adults: A field-based study** Jessica Vélez Avilés & Paola E. Dussias
14:50	**Break -- no refreshments**
15:10	**Prosody, context and visual cues in the processing of gapping sentences in Brazilian Portuguese** Andressa Silva & Aline Fonseca
15:35	**Where on the face do we look during phonemic restoration?** Alisa Baron, Vanessa Harwood, Daniel Kleinman, Joseph Molski, Nicole Landi & Julia Irwin
16:00	**Break**
16:10	KEYNOTE **Word learning in ASD: The sensorimotor, the perceptual and the symbolic** Mila Vulchanova
17:10	**Discussion**
17:30	**Break -- on your own**
18:15	**Dinner at The Graduate (campus hotel)**

FRIDAY, OCTOBER 28

8:15	**Breakfast**
9:00	***KEYNOTE*** **Pragmatics and spoken language comprehension in adulthood:** **Using aging as a window into fundamental processes** Craig Chambers
10:00	**The Stance of a Third Person Changes Perspective-Taking** Xiaopei Lin, Hui Chen, Yuxiu Han, & Xiaobei Zheng
10:25	**Break -- coffee and tea**
10:40	**Predictive processing of Mandarin state-change transitive events using morphosyntactic cues** Fang Yang, Martin Pickering, & Holly Branigan
~~11:05~~	*Cancelled due to illness* ~~**Partial Learning of Word Meaning from Referentially Ambiguous Naming Events**~~ ~~Nina Schoener, Sara Johnson & Sumarga Suanda~~
11:05	**Successful communication does not drive language development: Evidence from adult homesign** Marie Coppola
11:30	**Break -- coffee and tea**
11:45	KEYNOTE **What can the visual world paradigm tell us** **about language-vision interactions?** Falk Huettig
12:45	***Discussion (lunch available from 12:45 for those who need to leave)***
13:00	***Lunch served in foyer*** ***Boxed lunches, so you can eat on site or leave if necessary***

ABSTRACTS

KEYNOTE ADDRESS

The task-based visual world paradigm: Past (history and context), present, and (perhaps) future

Michael K. Tanenhaus
University of Rochester

About 30 years ago, in what seems like another lifetime, Michael Spivey and I began conversations about using eye-movements coupled to visually guided reaching to examine language processing. A brief pilot study convinced us that the approach was promising. We were soon joined by Kathy Eberhard and Julie Sedivy. In our weekly meetings, we discussed methodological and theoretical issues about using eye-movements to study spoken language processing, and we designed a series of small experiments to explore some of its potential applications (described in our 1995 paper in Science). I'll try to recreate the context in which that work was initiated, the rationale for the work, some of the early results, including successes, failures, surprises, and missed opportunities. After briefly considering how the visual world paradigm is currently used, I'll argue that one important missed opportunity is using the VWP to explore questions about the interface/division of labor between language and linguistic representations, non-linguistic perception, and action. I'll briefly sketch out how such a research program might proceed by using VR with visually guided reaching and changes to the VR world that are contingent on initiation of saccades and reaching.

Effect of coordination on perspective-taking: Evidence from eye-tracking

Yipu Wei [1], Yingjia Wan [2], & Michael Tanenhaus [3]

[1]*School of Chinese as a Second Language, Peking University*

[2]*Institute of Psychology, Chinese Academy of Sciences*

[3] *University of Rochester*

Coordinated activities foster social bonds by increasing closeness, affiliation, and trust (Hove & Risen, 2009), and improve comprehension between interlocutors (Richardson & Dale, 2005). Moreover, joint tasks that require continuous coordination (e.g., joint music-making) increase prosociality compared to collaborative tasks that focus primarily on a shared goal (Wan et al., 2019). In this study, we investigated the effect of coordination on perspective-taking, a cognitive skill critical to the process of socialization. We manipulated the degree of coordination in a collaborative puzzle task to explore how fine-grained coordination influences perspective-taking in online communication.

75 Chinese participants first finished a screen-based puzzle game with a computer player. Half of the participants were led to believe that they were playing with a real human participant. In the fine-grained coordination condition, the pieces assigned to the two players would always be immediately adjacent to each other once they are in place, thus creating a highly interdependent coordinative experience. In the coarse-grained coordination condition, the pieces were assigned in such an order that the piece a player received would be placed far away from the other player's last-placed piece, leading to a relatively independent collaborative experience.

We then measured participants' eye movements in a perspective-taking task adapted from Heller et al. (2008). Participants heard auditory instructions from the previous computer partner such as *Please give me the big cubic block*. The prenominal scalar adjective *big* elicits pragmatic inference that there is a size contrast of the target. The display included two pairs of size-contrasting blocks (Fig. 1, next page). The factor *ground* included the privileged condition (the competitor-contrast is only visible to the participant) and the shared condition (the competitor-contrast is visible to both sides). If participants were taking into account the speaker's perspective, then they could pre-identify the target at the adjective when the competitor-contrast was privileged.

We performed a growth curve analysis (Barr, 2008) on the proportion of looks to the target and target-set (target and target-contrast) during the scalar adjective region. There is a main effect of ground: the proportion of looks in the privileged condition is higher than those in the shared condition (target: $\beta = 0.33$, $SE = 0.02$, $z = 16.22$, $p < .001$; target-set: $\beta = 0.42$, $SE = 0.02$, $z = 22.48$, $p < .001$), providing further evidence that perspective influences real-time comprehension even in a complex visual setting. There is a significant interaction effect of ground, coordination and time in target-set analysis ($\beta = -0.48$, $SE = 0.11$, $z = -4.51$, $p < .001$), indicating that eye movements over time differ across different ground*coordination groups (Fig. 2, next page). For the fine-grained coordination group, the proportion of looks to the target in the privileged ground condition peaked earlier compared to the coarse-grained coordination group. This suggests that, with fine-grained coordination experience, participants were more attuned to their partner's perspective.

These results show that fine-grained coordination improves perspective-taking in online communication, suggesting that joint actions play an important role in facilitating social cognitive tasks.

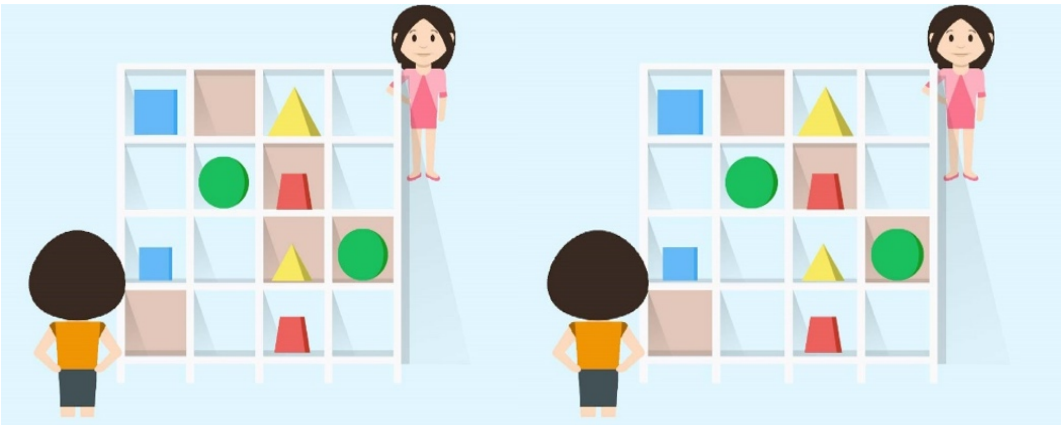


Fig. 1 Example displays of two ground conditions in the online referential communication task (for female participants). Left panel: privileged-ground condition; right panel: shared-ground condition. Four areas of interests were coded for analysis: target (the big cubic block), competitor (the big triangle block), target-contrast (the small cubic block) and competitor-contrast (the small triangle block).

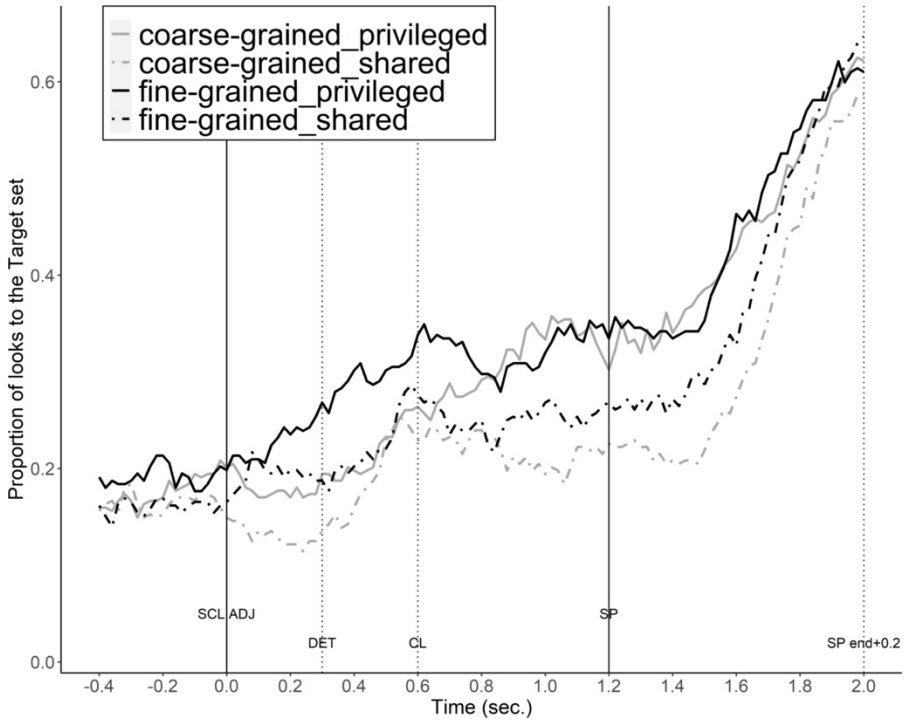


Fig.2 Proportion of fixations on the target-set during the critical time region (from the onset of the scalar adjective to 200ms after the onset of the shape adjective)

Selected references:

Barr 2008. *Journal of Memory and Language*, *59*(4).

Heller, Grodner, & Tanenhaus 2008. *Cognition*, *108*(3)

Wan, Fu, & Tanenhaus 2019. *Psychonomic Bulletin & Review*, 26(2).

Hove & Risen 2009. *Social Cognition*, 27.

Richardson & Dale 2005. *Cognitive Science*, *25*(6).

Fixations reveal preference: A VWP study of thematic role processing in Spanish

Beatriz Gómez-Vidal, Mirin Arantzeta, & Itziar Laka

Dept. of Linguistics and Basque Studies, University of the Basque Country

Previous eye-tracking research [1,2] has found that visual objects related to animate agents received a greater proportion of fixations than those related to patients. The interpretation of this finding is controversial: while some [1] argue that it could reveal greater processing cost, others [2] argue that it indicates a greater preference towards animate agents.

We investigated whether modulating the prototypicality of animacy and thematic role mappings affected proportion of fixations on related visual objects. We contemplated two hypotheses: (i) the Agent Preference, which predicts a preference for agents across the board, and (ii) the Prototypicality Hypothesis, which predicts a preference for prototypical mappings (animate agents, inanimate patients) over non-prototypical ones (inanimate agents, animate patients).

Following a Latin square design, we created 180 Spanish sentences with preverbal subjects, modulating the animacy of the subject (animate, inanimate) and verb type (unaccusative, unergative and transitive) in order to compare agent (unergative and transitive) and patient (unaccusative) subjects. Sentences were paired with visual displays containing 4 gray-scale drawings. Sentential subjects (the sailor) were strongly related to the visual target (ship) semantically. Sixty-two native speakers participated in two VWP tasks which involved listening to stimuli while looking at the visual displays. First, a visual norming task, in which participants heard an NP in isolation (the sailor) while looking at a display containing the visual target (ship). This was done in order to assess the strength of relationships between drawings and subjects, as revealed by eye movements. Next, the experimental task, in which participants listened to the full sentences while looking at the displays. We monitored participants' eye movements on the visual objects using a TobiiX120 eye tracker sampling at 120 Hz.

We used the Growth Curve Analysis technique [3] to analyze the proportion of fixations towards the visual target after the verb. Verb presentation was selected as the onset for the analysis because this was the time in which that participants could assign a thematic role to the preverbal subject. Independent variables were verb type, animacy and different types of polynomials with random slopes (by participant per condition). The dependent variable was the proportion of looks to the visual target per time bin across participants.

Results of the visual norming task showed that animacy of the NP did not modulate the proportion of looks to the semantically-related target. Results of the experimental task showed that visual objects related to animate agents and inanimate patients (prototypical mappings) received a greater proportion of fixations than those related to inanimate agents and animate patients (non-prototypical) at the two analyzed post-verb time frames (from 200 to 1700 ms after verb offset, and from 1700 to 3200 ms after verb offset). At the earliest time frame, objects related to animate transitive subjects received less visual attention than those related to animate unergative subjects. However, no difference in intercept was found between any of the prototypical mappings at the latest time frame. Our results are most consistent with the Prototypicality Hypothesis, and replicate previous findings that show a greater preference for animate agents over patients, revealed by a greater proportion of fixations towards a semantically-related object.

1. Koring L, Mak P, Reuland E. The time course of argument reactivation revealed: Using the visual world paradigm. Cognition. 2012;123(3):361–79.
2. Gómez-Vidal B, Arantzeta M, Laka JP, Laka I. Subjects are not all alike: Eye-tracking the agent preference in Spanish. PLoS ONE. 2022:17(8): e0272211.
3. Mirman D, Dixon JA, Magnuson JS. Statistical and computational models of the visual world paradigm: Growth curves and individual differences. J Mem Lang. 2008. Nov;59(4):475–94.

The use of prosodic cues in syntactic attachment: evidence from eye movements

Aline Fonseca [1], Andressa Silva [2], & Gabriela Rodrigues [1]

[1]Universidade Federal de Juiz de Fora; [2]Universidade do Estado de Minas Gerais

Presenting via video link

The current research investigates how prosodic cues, such as contrastive pitch accents and intonational phrase (IP) boundaries (Ladd, 2008; Nerspor & Vogel, 2007), influence the auditory perception and the attachment of adverbial prepositional phrases in ambiguous Brazilian Portuguese (BP) sentences: *O colega do Paulo revelou que a Camila fumou na varanda do sobrado*. (Paulo's friend revealed that Camila smoked on the balcony). The ambiguous adverbial phrase has two possible interpretations, it could be attached to the verb of the first clause (revelou/revelead), or it could be attached to the verb of the second clause (fumou/smoked). The attachment to the first verb presents a high attachment meaning: *Paul's friend revealed something to him on the house balcony*. The attachment to the second verb presents a low attachment meaning: *Camila smoked on the house balcony*. The Minimal Attachment Principle predicts that the low syntactic attachment is the default interpretation (Frazier, 1979). Previous auditory studies in English and in BP (Carlson & Tyler, 2018; Fonseca, Carlson & Silva, 2019) indicate that prosodic boundaries can influence the final interpretation of that ambiguous syntactic structure, biasing the high attachment of the ambiguous adverbial prepositional phrase. The present study was a spoken language comprehension task (N=28) with the Visual World Paradigm (Tanenhaus & Trueswell, 2006), conducted in an Eyelink 1000. We tested 24 sentences in four prosodic conditions: (1) pitch accent on the first verb (V1), (2) pitch accent on the second verb (V2), (3) pitch accent on V1 + IP before the ambiguous adverbial prepositional phrase (V1IP) (see Picture 1), and (4) pitch accent on V2 + IP before the ambiguous adverbial prepositional phrase (V2IP) (see picture 2). Along with the audio, participants were simultaneously exposed to two pictures that described the two possible interpretations for the ambiguous adverbial phrase (see Pictures 3 and 4). After listening to the sentences and seeing the pictures, they had to answer a comprehension question: *What happened on the balcony? a) Paul's friend revealed something there* or *b) Camila smoked there*. In V1 and V1IP conditions, the picture with high attachment bias was the target (Picture 3) and the picture with low attachment bias was the control (Picture 4). In V2 and V2IP conditions, the picture with low attachment bias was the target and the picture with high attachment bias was the control. The participants' eye movements were measured for total fixation duration (TFD), fixation count (FC), and proportion of looks to both pictures on the screen. The data reveals that participants tend to look more and longer towards the picture that corresponds to the action expressed by focused verb with contrastive pitch accent (see Picture 5). As for the interpretation choices, we found out that the prosodic cues bias the high attachment interpretation of the adverbial prepositional phrase (conditions V1IP x V2IP B=-1.676, SE=0.235, z= -7.138, CI [-2.161 -1.218], $p < 0.001$) (See Graph 1). Overall, the results indicate that participants are sensitive to prosodic cues and that they are able to use the prosodic information in the initial stages of sentence processing (Warren, 1996; Speer & Blodgett, 2006).

KEYNOTE ADDRESS

The nascent lexicon: Early word comprehension in the lab and in the world

Elika Bergelson
Duke University

While a longstanding view in language development holds that infants don't understand words until they begin talking (around age 1), recent research from eyetracking studies in the lab has revealed that infants begin understanding words months earlier (e.g., Bergelson & Swingley, 2012;2015; Tincoff & Jusczyk 1999;2012; Parise & Csibra, 2012; Bergelson & Aslin 2017; Kartushina & Mayer, 2019). In this talk I will explore two branches of my lab's work that begin to unpack the mechanisms of early word learning: studying the learner, and studying the learning environment.

First, I will discuss eyetracking data revealing infants' initially immature expectations about how words sound and what they mean, and how their representations eventually become more adult-like over infancy and toddlerhood as early phonology, semantics, syntax, and morphology come online. Synthesizing across studies, I will discuss recent results showing a robust, non-linear, and arguably *qualitative* improvements in infants' real-time word comprehension just after the first birthday. Drawing from SEEDLingS, my lab's audio and video corpus of home recordings, I will argue that this "comprehension boost" is not well-explained by changes in language input for common words, but rather, by postulating that infants learn to take better advantage of relatively stable input data. I will propose complementary theoretical accounts of what makes older infants "better learners." Time permitting, I will also briefly discuss the dynamics of language learning beyond our typical WEIRD populations, calling on data from cross-cultural collaborations, and early stage work looking at infants with sensory impairment.

Tracking Audio-visual Interaction for Semantic Representation Using FrameNet

Frederico Belcavello & Tiago Torrent

FrameNet Brasil / Universidade Federal de Juiz de For a

Presenting via video link

FrameNet Brasil is building a semantic model to analyze the combination of sentences in spoken audio and visual elements of a TV show for meaning construction. Such a model relies on both a methodology (Belcavello et al., 2020) and a computational tool (Belcavello et al., 2022) for manually annotating multimodal objects with Semantic Frames (Fillmore 1982). In Frame Semantics, Frames are structured representations of interrelated concepts, and words are understood relative to the broader conceptual scenes they evoke (Fillmore 1977). FrameNet (Fillmore et al 2003) implements Frame Semantics as a lexicographic database that describes the words in a language against a computational representation of frames, their frame elements (FEs) and the relations between them. The analysis is attested through the annotation of sentences representing how lexical units (LUs) instantiate the frames they evoke. To apply this framework to multimodal analysis we considered that visual objects in audiovisual data may (i) evoke frames and organize the elements on the screen and (ii) act in combination with the frame evocation patterns of the sentences from spoken audio. In this last case, audio-visual interactive patterns vary along two axes: that of synchronicity and that of identity between the frame-semantic representations generated from annotation for each of the communicative modes. As a proof of concept we conducted the multimodal annotation of one episode of a Brazilian TV Travel Show, critically acclaimed as an example of good practices in audio-visual composition. Through the 23 minutes of the episode, 233 sentences were spoken, for which 1336 lexical annotation sets were created. Annotators also identified 809 visual objects directly or indirectly related to the lexical annotation sets. Results suggest that the framework is capable of building a model that accounts for the combination of visual elements and spoken audio within a fine-grained FrameNet representing the semantic complexity of multimodal corpora.

The role of attention during grammatical gender processing in low and high literacy adults: A field-based study

Jessica Vélez Avilés and Paola E. Dussias
Pennsylvania State University

Past studies have reached to the conclusion that literacy affects cognitive processing. There is evidence showing that low literacy individuals display lower performance in cognitive tasks as compared to their high literacy counterparts. A possible explanation for a consistent advantage in performance of high literacy individuals is that literacy leads to enhanced efficiency in processing and attention. While previous studies have examined differences between low and high literacy individuals focusing on phonological (Loureiro et al., 2004; Schaadt et al., 2013), semantic (e.g., Kosmidis et al., 2004), and lexical processing (e.g., Kosmidis et al., 2006), a linguistic feature that has been less investigated is grammatical gender. Here we explored how general cognitive processes are deployed during morphosyntactic processing using a visual world paradigm task with Spanish-speaking adults with varying levels of literacy. Since grammatical gender is a linguistic feature that is highly entrenched in native Spanish grammars (e.g., it is acquired by 26 months of age; see López Ornat, 1996, 1997), perhaps in this case literacy skill may not affect cognitive efficiency during morphosyntactic processing. We, thus, hypothesized that eye movements of literate and illiterate participants should be similar.

High and low literacy speakers (female = 23, male = 19) from an underrepresented Afro-Hispanic community in San Basilio de Palenque (Colombia) saw two-picture visual displays in which items matched or did not match in grammatical gender. Participants' eye movements were recorded while they listened to 80 Spanish sentences in which target items were preceded by a feminine or masculine article that agreed in gender with both pictures in the visual scene or with one of the pictures (e.g., *Encuentra la/el...* 'Find the$_{\text{FEM/MASC}}$...'). For these individuals, participation in a lab-based study in the field was an entirely new task. Therefore, the demands of the task itself (carefully listening to speech, looking at pictures on a computer screen, clicking on the correct item, and using a mouse to do so—all while the speech signal unfolds) required a heavy attentional load (cf. Olivers et al., 2014; Smith et al., 2014).

Data were analyzed by comparing the proportion of eye fixations on target objects in each condition. The results showed that high and low literacy speakers looked sooner at the target item on different-gender trials than on same-gender trials, replicating results from previous studies with college-educated Spanish speakers. Crucially, both participant groups launched eye movements to the upcoming target object before target word onset. The results suggest that low literacy individuals are able to retrieve morphosyntactic information during spoken language processing and to map this information to the visual environment to the same extent as high literacy individuals. Our study extends the findings to diverse populations and assesses both the validity of theories of language comprehension and the contribution of literacy skill in language processing.

Prosody, context and visual cues in the processing of gapping sentences in Brazilian Portuguese

Andressa Silva [1] & Aline Fonseca [2]

[1]*Universidade do Estado de Minas Gerais;* [2]*Universidade Federal de Juiz de Fora*

Presenting via video link

This research studies Brazilian Portuguese (BP) coordinate sentences with subject versus object ambiguity: *No fim de semana, o Pedro levou a Júlia na festa e o Bruno no churrasco da empresa* (On the weekend, Pedro took Julia to the party and Bruno to the company barbecue. The noun "Bruno" can be the subject of a new clause (gapping structure) or the conjoined object. Speakers usually prefer conjoining objects rather than subjects (Hoeks et al., 2002). The dispreference for subject reading is due to its complex syntactic and information structure, which contradicts the Minimal Attachment Principle (Frazier, 1979). However, the manipulation of grammatical and extra-grammatical features can influence the interpretation of gapping sentences (Carlson, 2002; Hoeks et al., 2009). Therefore, we aim to investigate whether prosodic and contextual cues can bias gapping interpretation. We conducted a sentence-picture matching task (N = 48) manipulating context and prosodic structure: Context (Subject Context/SC; Object Context/OC; No Context) and Prosody (Subject Accent/SA; Object Accent/OA). The Subject Context stated "an action of X and Z towards Y", while the Object Context stated "an action of X towards Y and Z". The nouns "Pedro" and "Bruno" were contrastively pitch accented in Subject Accent Prosody, and the nouns "Júlia" and "Bruno" in Object Accent Prosody. Pitch accented nouns are lengthened and have a similar F0 range. A set of three pictures were designed to explain the two conjuncts of the sentences: the first picture depicted the first conjunct (*Pedro took Julia to the party*) and two other pictures respectively biased subject interpretation (*Bruno took Julia to the barbecue*) and object interpretation (*Pedro took Bruno to the barbecue*) of the second conjunct. We tested thirty experimental sentences plus thirty-two fillers in the PCIbex (Zehr & Schwarz, 2018). The first screen showed the characters' pictures while the context sentence (when available) was played. The second screen showed first conjunct picture at the top, and the biasing pictures side-by-side at the bottom, while the target sentence was playing. Participants had to choose between the two biasing pictures and click on the one that best matched the target sentence. The picture choice data reveal that subject interpretation was mostly chosen only in SCSA and SCOA conditions (63.44% and 55.23% respectively), suggesting that biasing subject contexts prepared participants for expecting two subjects in gapping sentences. Prosody had a considerable effect in promoting subject interpretation in SA condition (31.93%), but the context was stronger in biasing gapping in both SCSA and in the mismatch SCOA condition. A logistic mixed-effects regression model (Baayen et al., 2008) revealed significant statistical differences of picture subject responses between SCSA and SCOA conditions ($B=0.67$, $CI[-0.67 \sim 0.68]$, $p<0.001$), and between SCSA and SA conditions ($B=0.22$, $CI[0.21 \sim 0.22]$, $p<0.001$). The results suggest that context and prosody combined can reduce the dispreference for gapping interpretation in BP. The context is a relevant cue in processing and it is highlighted by the prosodic realization of the utterance.

Where on the face do we look during phonemic restoration?

Alisa Baron [1], Vanessa Harwood [1], Daniel Kleinman [2],
Joseph Molski [1], Nicole Landi [2,3] & Julia Irwin [2,4]
[1]U. Rhode Island, [2]Haskins Laboratories, [3]U. Connecticut, [4]Southern Connecticut State U.

Face-to-face communication typically involves an audio (heard) and visual (seen) speech signal. Previous research on audiovisual speech perception suggests that neurotypical adults use visual articulatory information from the mouth in difficult speaking situations and that visual input enhances auditory speech perception. Little is known about listeners' gaze to the face of a speaker under varying task demands. Given different cognitive resources that may be needed to support effective communication, we wondered whether listener gaze to the speaker's face varied depending on environmental demands, when the need to process the signal is increased due to an attenuated or distorted signal.

To evaluate this, we used a novel visual phonemic restoration paradigm. Forty English-speaking neurotypical adults participated in two eye-tracking experiments which included an audio-visual condition (articulatory information from the mouth was visible) and a pixelated condition (articulatory information was not visible) within passive and active (button press response to a deviant sound) response contexts. Synthesized /ba/ and reduced /ba/ or /*a*/ stimuli were dubbed onto the video of a face speaking /ba/ (AV condition) or a video of a face with a pixelated mouth region with no visible movement (PX condition). These stimuli were presented in an 80/20 oddball design, with /*a*/ serving as the deviant stimulus in both face contexts. The first experiment was passive, and the stimuli were presented in AV and PX conditions. The second experiment was active, and the stimuli again were presented in AV and PX conditions. Prior to analysis, fixations to each interest area were averaged within four consecutive, non-overlapping 300 ms bins. A bin width of 300 ms was used because the visual and auditory information was appropriately segmented, corresponding to the initial rest position (0-300 ms), the mouth opening prior to the consonant closing gesture (300-600 ms), the consonant closing gesture (600-900 ms), and the peak mouth opening for the vowel (900-1200 ms).

Trial-level analyses were conducted with linear mixed-effects models. One model was fit to each of the two interest areas (Mouth/Jaw and Eyes). All models had fixed effects of experiment, condition, time window, and within-task trial number. Results revealed that the greatest fixations to the mouth were present in AV active experiment and visual articulatory information led to a phonemic restoration effect. There were more fixations when the mouth was visible and when the environmental demands necessitated the participants to discriminate between speech tokens and provide a button press response. Thus, when there is information that can be gleaned from the mouth/jaw area, and there is a higher demand/requirement to discern speech, neurotypical adults take advantage of this input. Participants looked significantly more at the eyes when the mouth was pixelated (PX condition) than when the mouth was visible (AV condition). Perhaps when visual articulatory information is not present, participants seek other communicative information and thus gaze at the eyes. These findings have important clinical and real-world consequences, such that general face-to-face communication in which articulatory information is available has the potential to enhance speech perception.

KEYNOTE ADDRESS

Word learning in ASD: The sensorimotor, the perceptual and the symbolic

Mila Vulchanova

Language Acquisition and Language Processing Lab,
Norwegian University of Science & Technology – Trondheim

Word learning requires successful pairing of form and meaning. A common hypothesis about the process of word learning is that initially, infants work on identifying the phonological segments corresponding to words (speech analysis), and subsequently map those segments onto meaning. A range of theories have been proposed to account for the underlying mechanisms and factors in this remarkable achievement. While some are mainly concerned with the sensorimotor affordances and perceptual properties of referents out in the world, other theories emphasize the importance of language as a system, and the relations among language units (other words or syntax). Recent neuro-science inspired approaches suggest that the storage and processing of word meanings is supported by neural systems subserving both the representation of conceptual knowledge and its access and use (Lambon Ralph et al., 2017).

Developmental disorders have been attested to impact on different aspects of word learning. While impaired word knowledge is not a hallmark of ASD, and remains largely understudied in this population, there is evidence that there are, sometimes subtle, problems in that domain, reflected in both how such knowledge is acquired and how words are used (Vulchanova, Saldaña & Baggio, 2020). In addition, experimental evidence suggests that children with autism present with specific problems in categorizing the referents of linguistic labels leading to subsequent problems with using those labels (Hartley & Allen, 2015). Furthermore, deficits have been reported in some of the underlying mechanisms, biases and use of cues in word learning, such as e.g., object shape (Field, Allen & Lewis, 2016; Tek, Jaffery, Fein & Naigles, 2008). Finally, it is likely that symbol use might be impaired in ASD, however, the direction of the causal relationship between social and communication impairment in autism and symbolic skills is still an open question (Allen & Lewis, 2015; Allen & Butler, 2020; Wainwright, Allen & Cain, 2020). Further support for impaired symbol formation in autism comes from the well-attested problems with figurative, non-literal language use (e.g., metaphors, idioms, hyperbole, irony) (Vulchanova, Saldaña, Chahboun & Vulchanov, 2015). Here we propose that embodied theories of cognition which link perceptual experience with conceptual knowledge (see Eigsti, 2013; Klin, Jones, Schultz & Volkmar, 2003) might be useful in explaining the difficulty in symbolic understanding that individuals with autism face during the word learning process.

KEYNOTE ADDRESS

Pragmatics and spoken language comprehension in adulthood: Using aging as a window into fundamental processes

Craig Chambers
University of Toronto

To many cognitive scientists, work on cognitive aging is often thought of as a peripheral area of applied research. On this view, the goal is to simply to connect insights from "core research" (involving young adults) with known patterns of perceptual and cognitive decline to yield a model of age-related change.

In fact, lifespan approaches carry the potential to strongly enrich our understanding of fundamental processes by provoking alternative, deeper, or more nuanced views of cognitive representations and mechanisms. In the domain of language, relevant phenomena include:

- cases showing an intriguing disconnect between cognitive decline and linguistic decline

- cases where declines in language processing abilities are very mild or nonexistent, and whether these cases have a direct connection to the notion of "context usage" or "compensation"

- cases where age-related changes in performance might reflect sensory rather than cognitive decline (evoking the longstanding conundrum of whether processing challenges arise from limited resources or just limited data, cf. Norman and Bobrow, 1975)

In this talk, I will describe the merits of a lifespan approach for understanding the core mechanisms underlying real-time language interpretation. The primary focus will be studies of pragmatic inference, where eye movement measures are used to track listeners' understanding as speech unfolds.

The Stance of a Third Person Changes Perspective-Taking

Xiaopei Lin, Hui Chen, Yuxiu Han, & Xiaobei Zheng
School of Foreign Languages, Shenzhen University, China

Presenting via video link

Conversational participants are always expected to use ground cues. Some propose that shared knowledge is one type of contextual information that people automatically employ. Others suggest that interlocutors infer the partner's mental states to understand language. The present study set up a low-level cue (physical presence) and a high-level cue (mental inference) to explore which of the cues is more effective.

Experiment 1: Thirty college students participated in the study. They sat in front of a frame with one blocked grid. The display either contained four different objects (NON-COMPETITOR condition) or two different and one pair of identical objects in different sizes (COMPETITOR condition) (Fig.1).

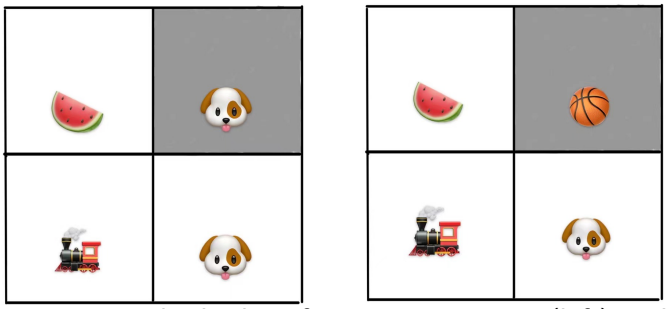


Fig. 1. Sample displays for COMPETITOR (left) and NON-COMPETITOR (right) conditions.

Experimenter 1 (E1) either sat on the same side with the participants, sharing all the objects with them (SHARED) or sat on the opposite side, sharing only three objects (NON-SHARED). Experimenter 2 (E2) always sat on the opposite side (Fig. 2). Either E1 or E2 instructed participants to "*point to the dog*". The target object was always presented in the transparent grids, with the competitor in the blocked one.

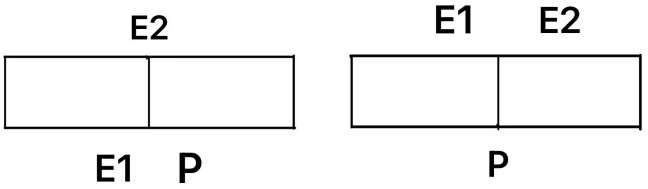


Fig. 2. The object in the grey grids was visible to E1 in the SHARED condition (left), but not visible in the NON-SHARED condition (right).

Participants' fixation on the target object was calculated across the interval of the critical noun (e.g., "dog"). Their eye fixations on the E2 trials displayed a similar pattern to E1 trials: no effect of position was found in the non-competitor trials; a significant position effect was found in the competitor trials, $p < 0.1$, showing that addressees were more likely to look at the shared objects when E1 sat with E2 at the opposite to them (Fig. 3, next page). But it is not clear whether the physical presence or the mental inference of E1 influences participants taking E2's perspective.

Experiment 2: Twenty college students participated in the study. The seating was the same, but the frame was adjusted: When E1 sat on the same side with the participants, they share only three objects (NON-SHARED). When E1 sat on the opposite, they share all the objects (SHARED). E2 always shared only three objects (Fig. 4, next page).

A larger fixation on the shared objects in NON-SHARED for E2 trials would indicate that the mental states of E1 influence how participants evaluate E2's perspective, but the shared object preference in SHARED would indicate that the physical position makes the difference. However, neither of the two possibilities was found: no effect of position was found in both the non-competitor and the competitor trials, $ps < 0.1$, implying that both the physical and mental presentation of the third person may interactively influence participants' perspective-taking.

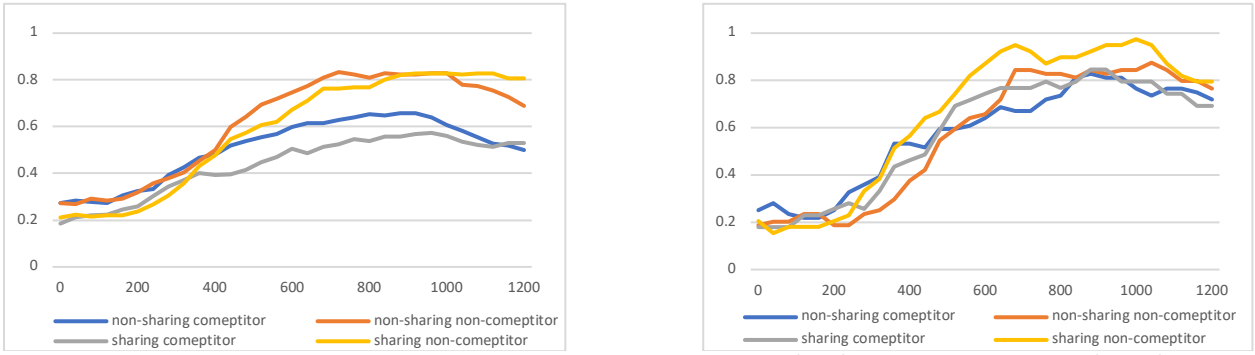


Fig. 3. The proportion of fixation to the target in Experiment 1 (left) and Experiment 2 (right).

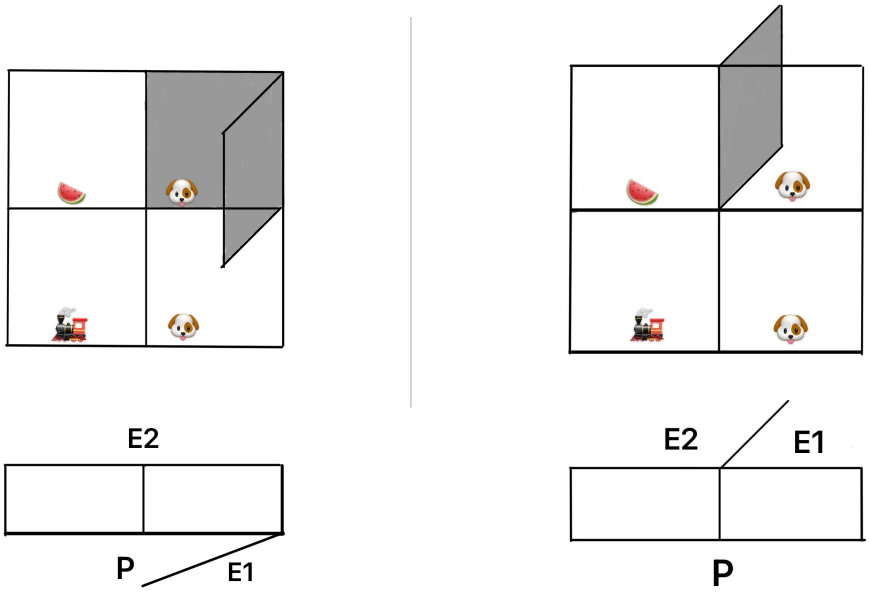


Fig. 4. The objects in the up-right corner are not visible to E1 in the NON-SHARED condition (left), but visible in the SHARED condition (right).

Predictive processing of Mandarin state-change transitive events using morphosyntactic cues

Fang Yang, Martin Pickering, & Holly Branigan
University of Edinburgh
Presenting via video link

Listeners predict prominent information in upcoming utterances [1-2]. In state-change transitive (SCT) events (e.g., Lee broke a plate), one prominent element is the resultant state of the patient (e.g., broken) as it marks the event boundary and facilitates event segmentation [3]. Previous studies show that listeners predict entities' states using verb semantics, tense or aspect as a cue [4-7]. Mandarin SCT events are often described in a BA-construction (e.g., (1)) in which the marker *BA* followed by the patient occur before the verb and the resultant state. Therefore, in theory, Mandarin speakers should be able to use *BA* as a cue to predict the resultant state of the patient in SCT events before encountering the verb. We tested this hypothesis in five experiments.

(1) Lee ***BA*** *yi-ge* *panzi* *nong* *sui* *LE.*
Lee BA one-classifier plate make broken ASPmarker
"Lee broke a plate."

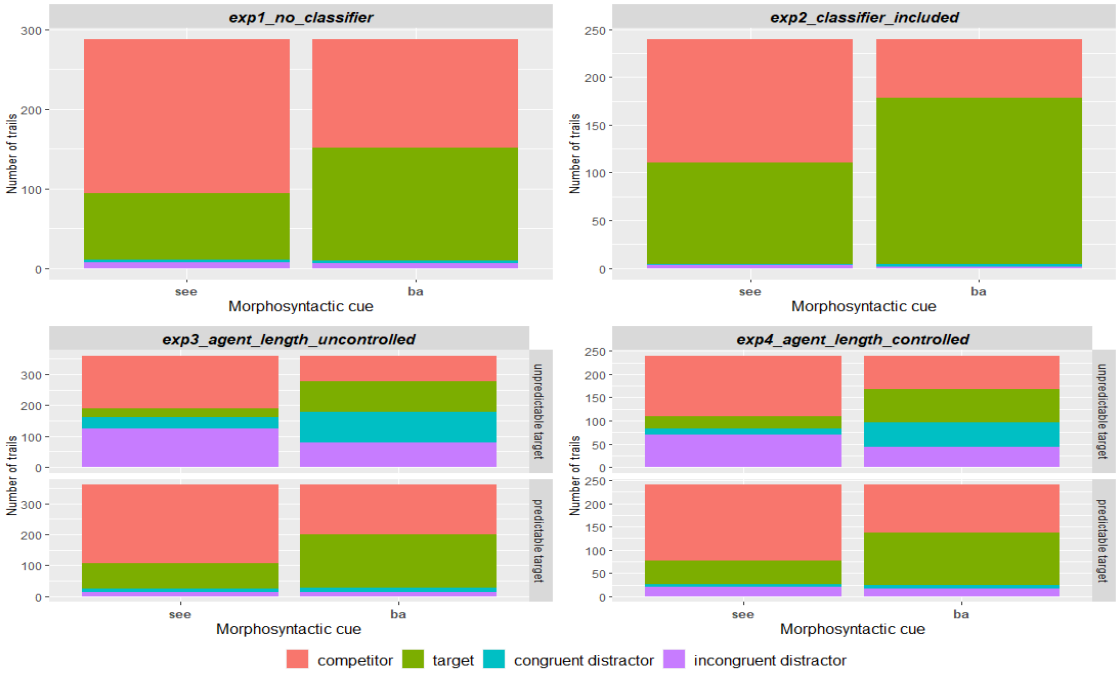
Each experiment used 24 items testing Mandarin speakers' comprehension of SCT events in a "BA" condition and a baseline "see" condition in which *BA* was replaced by verb *kan* "see" thus rendering the resultant state unpredictable (e.g., *Lee* ***kan*** *yi-ge panzi nong sui LE*, "Lee saw that a plate got broken"). Each item was paired with four images containing a target (e.g., a broken plate), a competitor (e.g., a good plate), a congruent distractor (e.g., a broken mirror), and an incongruent distractor (e.g., a good mirror). In Experiment 1, 2 and 5 (n=20, 24, 24), each item had a context sentence (e.g., Lee bought two plates). Experiment 3-4 (n=60, 40) instead manipulated the predictability of the target (e.g., a used wine glass) via the agent (e.g., Mr Drunkard vs Mr Lee) in a 2 (predictable vs unpredictable) X 2 (cue: *BA* vs *see*) design.

Experiment 1-4 used a forced-choice image selection paradigm. In each trial, participants viewed four images complemented with an incomplete sentence (e.g., "Lee BA/see a ...") and selected the image that they thought the sentence was describing. Bayesian analyses show that participants were more likely to select the target image in the BA condition (prob. range: 47%-72%) than the baseline condition (20%-44%) across all experiments. Moreover, they were more likely to select the target image when it was predictable via the agent (48%, 47% in Exps 3&4) than unpredictable (29% and 28% in Exps 3&4).

Experiment 5 used a visual world eye-tracking paradigm in which participants first listened to a context sentence, and then listened to a critical sentence in either a "BA" or "see" condition while viewing four images with 1500ms preview time. GCA and cluster-based permutation analyses revealed no reliable effects of *BA* during the critical time window before patient onset. Potential effects here might be confounded by a "learning effect" as critical sentences always described target images across conditions. Experiment 6 will address this issue by including more competitor-describing fillers using "see" sentences.

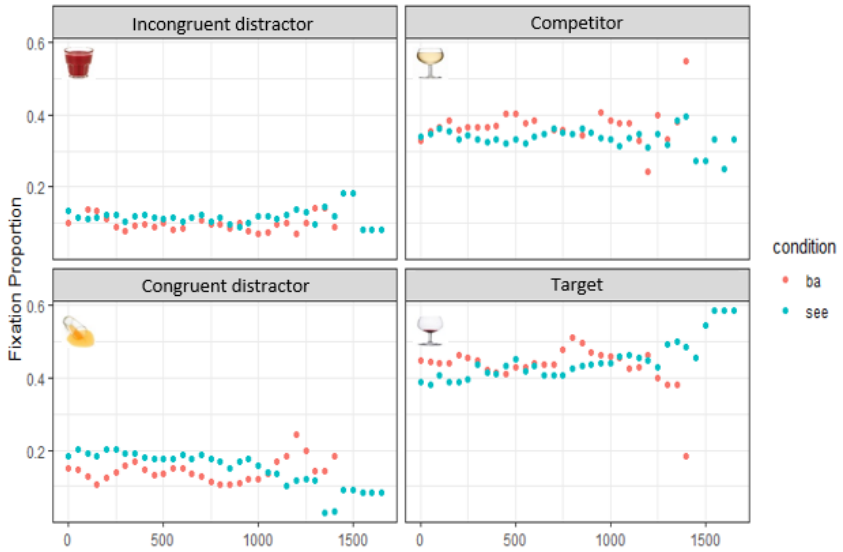
Taken together, these results suggest that when comprehending SCT events, Mandarin speakers may pre-activate the resultant state of the patient using morphosyntactic marker *BA* as a cue.

Fig 1. Frequency of choices following cue "BA" vs "see" in Experiment 1-4



Note: The green segments show participants' choices of the target image (e.g., a used red wine glass). Experiment 1-2 had context sentences. Experiment 3-4 did not have context sentences but instead controlled for the predictability of the target image via the agent (e.g., Mr Drunkard vs Mr Lee).

Fig 2. Proportion of fixations on each image during a critical time window between cue onset and patient onset in Experiment 5



Note: Time window from cue onset to patient onset (e.g., *BA/kan yi-bei,* "*BA*/see one-classifier")

References: **[1]** Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*. **[2]** Huettig, F., Rommers., J., Meyer, A.S. (2011) Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*. **[3]** Ünal, E., Ji, Y., & Papafragou, A. (2019). From event representation to linguistic meaning. *Topics in Cognitive Science*. **[4]** Altmann. G.T.M., and Kamide. Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*. **[5]** Altmann. G.T.M., and Kamide. Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition*. **[6]** van Bergen, G., Flecken, M., (2017). Putting things in new places: Linguistic experience modulates the predictive power of placement verb semantics. *Journal of Memory and Language*. **[7]** Zhou, P., Crain, S., & Zhan, L. (2014). Grammatical aspect and event recognition in children's online sentence comprehension. *Cognition*.

Partial Learning of Word Meaning from Referentially Ambiguous Naming Events

Nina Schoener[1], Sara Johnson[2] & Sumarga Suanda[1]

[1]University of Connecticut; [2]National Institutes of Health

A key debate on early word learning concerns the input that matters for learning. One account argues that early learning stems largely from the small set of input that is referentially transparent (Gleitman & Trueswell, 2020; Medina et al., 2011; Trueswell et al., 2013). A second account argues that early learning emerges from a broad swath of the input, including input that is referentially ambiguous (Smith & Yu, 2008; Yu & Smith, 2012). At the heart of this prominent debate is thus whether learning word-to-referent mappings can take place from referentially-ambiguous naming events. The current work presents new evidence suggesting that the answer may depend on what counts as "learning".

The current study was designed based on Medina and colleagues' (2011) landmark study in which they presented adult word learners multiple muted-vignettes of parent-toddler interactions in which parents uttered a target English word (e.g., "car"). Medina and colleagues found that learners struggled to identify the target word when the vignettes had been independently rated as referentially ambiguous (i.e., vignettes whose target word was difficult to guess). The current study asks whether participants' failure to identify the precise identity of words masks other forms of word meaning knowledge participants may have extracted from referentially ambiguous vignettes.

Participants (N = 52) completed a learning phase in which they were presented with 9 scenes from children's picture books that had contained a common target word (e.g., "dog") in its original text. Scenes were presented with all text removed and all scenes were identified to be referentially ambiguous in a prior norming study. Following the learning phase, and like Medina and colleagues' original study, participants completed a free-response test where they were asked for the identity of the target word. Additionally however, participants completed a series of two-alternative forced choice trials in which they were presented with two new pictures and asked which picture was more likely to contain the target word. Of particular interest is whether participants, despite failing at the free-response test, would nonetheless succeed at the forced-choice test. This learning phase - test phase sequence was repeated for 8 target words.

We replicated Medina and colleagues' finding that many participants failed to identify the precise word meaning from referentially ambiguous stimuli; about 52% (SD = 27%) of words were identified correctly in the free-response test. Importantly however, participants' forced-choice performance following an incorrect free response was reliably above chance, $M = 0.64$, $SD = 0.15$, $t(47) = 6.55$, $p < .001$. These results suggest that when referentially ambiguous naming events were insufficient to yield learning of complete word meaning, they equip learners with reliable and useful partial knowledge about words.

Understanding the kinds of input that shapes early word learning is central to both theories of the mechanisms of learning and to interventions that seek to support it. The current work suggests that a better understanding of that input may depend in large part on the output of learning that one seeks to explain.

KEYNOTE ADDRESS

What can the visual world paradigm tell us about language-vision interactions?

Falk Huettig
Max Planck Institute for Psycholinguistics & Radboud University

The visual world paradigm (VWP) has been very successful for the investigation of key issues in language processing. Much less research using the VWP has been directed at exploring key issues in vision, attention, or language-vision interactions. I will present some VWP experiments that tested whether the nature of the visual environment (in particular the explicit availability of color in the surroundings) are of prime importance for the access and use of 'language-derived' color representations. To explore this question, we tested the effects of color representations during language processing in three visual-world eye-tracking experiments. On critical trials, participants listened to sentence-embedded words associated with a prototypical color (e.g., '...spinach...') while they inspected a visual display with four printed words (Experiment 1), colored or greyscale line drawings (Experiment 2) and a 'blank screen' after a preview of colored or greyscale line drawings (Experiment 3). Visual context always presented a word/object (e.g., frog) associated with the same prototypical color (e.g. green) as the spoken target word and three distractors. When hearing *spinach* participants did not prefer the written word *frog* compared to other distractor words (Experiment 1). In Experiment 2, color competitors attracted more overt attention compared to average distractors, but only for the colored condition and not for greyscale trials. Finally, when the display was removed at the onset of the sentence, and in contrast to the previous blank-screen experiments with semantic competitors, there was no evidence of color competition in the eye-tracking record (Experiment 3). These results fit best with the notion that the main role of visual representations in language processing is to contextualize language in the immediate environment. I will close by discussing the potential of the VWP to illuminate our understanding of how language, attention, memory, and vision interact.