

33RD JOINT WORKSHOPS ON  
LINGUISTICS AND LANGUAGE  
PROCESSING

**jwllp33**<sup>seoul</sup>

# **(Mis)matches between Form and Function in Natural Languages**

**April 25th (Fri),  
2025**

---

Virtual event via Zoom

Visit us online for more info:  
<https://khuisli.github.io/jwllp33>



Organized and hosted by  
Institute for the Study of Language and Information (ISLI),  
Kyung Hee University

## The 33rd Joint Workshop on Linguistics and Language Processing

**Theme: (Mis)matches between Form and Function in Natural Languages**  
(Hosted by the Institute for the Study of Language and Information, Kyung Hee University)

April 25, Friday, 2025

Zoom link: <https://khu-ac.zoom.us/j/81106640425>

Session 1-2 - Seoul 10:00-14:00 / Hong Kong, Macau 09:00-13:00 / Paris 03:00-07:00

Session 3-4 - Seoul 14:10-16:20 / Hong Kong, Macau 13:10-15:20 / Paris 07:10-09:20

09:50-10:00	<b>Opening Remarks</b> Jong-Bok Kim (Kyung Hee University) & Yasunari Harada (Waseda University)
	<b>Session 1</b> Moderator: Sanghoun Song (Korea University)
10:00-10:30	<b>Vincent X. Wang</b> (The University of Macau) <a href="#">Do Chatbots Understand What Is Not Said but Implied?</a>
10:30-11:00	<b>Youngju Choi</b> (Chosun University) <a href="#">How Meanings are Constructed in Productive Signs of Sign Languages: Focusing on HOUSE in Korean Sign Language</a>
11:00-11:30	<b>Hyeonah Kang</b> (Chosun University) <a href="#">Building Connections between Linguistic Form and Meaning in L2 Vocabulary Learning and Reading Behavior: Evidence from Eye-tracking</a>
11:30-12:30	<b>Lunch Time</b>
	<b>Session 2</b> Moderator: Jinghang Gu (The Hong Kong Polytechnic University)
12:30-13:00	<b>Yu Wang and Chu-Ren Huang</b> (The Hong Kong Polytechnic University) <a href="#">A Multimodal Dataset of Identifiability for Embodied Tasks</a>
13:00-13:30	<b>Zhaoxin Feng, Jianfei Ma, Xiaojing Zhao, Xiaoyi Bao, and Emmanuele Chersoni</b> (The Hong Kong Polytechnic University) <a href="#">From Unidirectional to Bidirectional: Semantic Probing in Decoder-Only Models</a>
13:30-14:00	<b>Xiaoyi Bao, Jinghang Gu, Zhongqing Wang, Minjie Qiang, and Chu-Ren Huang</b> (The Hong Kong Polytechnic University) <a href="#">Employing Glyphic Information for Chinese Event Extraction with Vision-Language Model</a>
14:00-14:10	<b>Break Time</b>
	<b>Session 3</b> Moderator: Juwon Lee (Jeonju University)
14:10-14:40	<b>Anne Abeillé and Emma Kiouss</b> (LLF, Université Paris Cité) <a href="#">Ellipsis and gender: some new results from French</a>
14:40-15:10	<b>Gabriela Bilbîie</b> (University of Bucharest) <a href="#">Multiple wh-questions in Romanian: An Empirical Approach</a>
15:10-15:20	<b>Break Time</b>
	<b>Session 4</b> Moderator: Peng Bo (The Hong Kong Polytechnic University)
15:20-15:50	<b>Wenxi Fei and Yu-yin Hsu</b> (The Hong Kong Polytechnic University) <a href="#">Dominance of F0 in mapping acoustic cues to pragmatic meaning: Eye-tracking evidence from Mandarin prosodic focus processing</a>
15:50-16:20	<b>Slallichi Lavinia and Yu-yin Hsu</b> (The Hong Kong Polytechnic University) <a href="#">Prediction, semantic evaluation, and sentence constraints: A computational model of reading processing</a>
16:20-16:30	<b>Closing Remarks</b> Chu-Ren Huang (The Hong Kong Polytechnic University)

For more workshop information, visit the website: <https://khuisli.github.io/jwllp33>

# List of Abstracts – Presentations

## Session 1

### **Do chatbots understand what is not said but implied?**

**Vincent X. Wang**

(The University of Macau)

People convey meanings not only by what is said, but, more importantly, by what is implied. Humans handle the latter intuitively, while chatbots rely on pre-trained large language models (LLMs) to compute implied meanings, presumably leading to differences from human linguistic behaviours. Our pilot study reveals that chatbots such as GPT4o, GPT-o3-mini, DeepSeek FW, and Claude 3.5 exhibit differences from human minds in recovering implied meanings from time to time. We attempt to explore reasons for the differences and revisit the notion of intelligence regarding language comprehension in this new light.

### **How Meanings are Constructed in Productive Signs of Sign Languages: Focusing on HOUSE in Korean Sign Language**

**Youngju Choi**

(Chosun University)

This study explores productive signs incorporating the sign HOUSE in Korean Sign Language (KSL). By analyzing how HOUSE interacts with classifiers, affixes, and spatial modifications, it examines the morphological processes that generate new meanings. The study also investigates movement variations and their role in creating semantically rich expressions. Findings reveal that iconicity plays a crucial role in these formations, demonstrating systematicity in KSL's productive morphology. This research contributes to understanding how sign languages expand their lexicon through productive processes, highlighting the interplay between spatial representation and linguistic structure in KSL's morphological system.

### **Building Connections between Linguistic Form and Meaning in L2 Vocabulary Learning and Reading Behavior: Evidence from Eye-tracking**

**Hyeonah Kang**

(Chosun University)

There is strong evidence that engaging in word-focused activities after reading aids vocabulary learning, yet the underlying mechanisms remain unclear. Does such engagement influence the processing of target vocabulary in subsequent readings, thus facilitating vocabulary learning? Using an eye-tracker, we recorded L2 learners' eye movements as they read an English narrative with 16 target non-words before and after completing one of three activities: (a) Comprehension-test (control), (b) Fill-in-the-blank activity, and (c) Summary-writing. Results showed that both Fill-in-the-blank and Summary-writing led to greater vocabulary gains than the Comprehension-test. However, only the Summary-writing group demonstrated a significant correlation between reading time differences and vocabulary test accuracy. Further analysis revealed that words initially not remembered but later recalled were read similarly across readings, particularly in the Fill-in-the-blank activity.

## Session 2

### A Multimodal Dataset of Identifiability for Embodied Tasks

*Yu Wang, Chu-Ren Huang*

(The Hong Kong Polytechnic University)

Existing research on embodied AI often overlooks that objects rarely exist in isolation. For example, with the task "give me that cup," there may be multiple similar cups present. An AI must identify the specific referent to interact correctly. Using linguistic theories of identifiability, we propose a multilingual and multimodal benchmark to evaluate the embodied referent identification capabilities of LVLMs. Results show that current LVLMs have weak identification abilities due to their lack of grounded embodied knowledge and the "self vs others" theory of mind, which is crucial for self-awareness.

### From Unidirectional to Bidirectional: Semantic Probing in Decoder-Only Models

*Zhaoxin Feng, Jianfei Ma, Xiaojing Zhao, Xiaoyi Bao, Emmanuele Chersoni* (The Hong Kong Polytechnic University)

Autoregressive Large Language Models (LLMs) demonstrate exceptional performance in language understanding and generation. However, their application in text embedding tasks has been relatively slow, along with the analysis of their semantic representation in probing tasks, due to the constraints of the unidirectional attention mechanism. Our study aims to explore whether such constraints can be overcome by enabling bidirectional attention in LLMs. We tested different variants of the Llama architecture through additional training steps, progressively enabling bidirectional attention and unsupervised/supervised contrastive learning. Our results show that bidirectional attention improves the LLMs' ability to represent subsequent context but weakens their utilization of preceding context, while contrastive learning training can help to maintain both abilities.

### Employing Glyphic Information for Chinese Event Extraction with Vision-Language Model

*Xiaoyi Bao, Jinghang Gu, Zhongqing Wang, Minjie Qiang and Chu-Ren Huang*

(The Hong Kong Polytechnic University)

As a complex task that requires rich information input, features from various aspects have been utilized in event extraction. However, most of the previous works ignored the value of glyph, which could contain enriched semantic information and cannot be fully expressed by the pre-trained embedding in hieroglyphic languages like Chinese. We argue that, compared with combining the sophisticated textual features, glyphic information from visual modality could provide us with extra and straight semantic information in extracting events. Motivated by this, we propose a glyphic multi-modal Chinese event extraction model with hieroglyphic images to capture the intra- and inter-character morphological structure from the sequence. Extensive experiments build a new state-of-the-art performance in the ACE2005 Chinese and KBP Eval 2017 dataset, which underscores the effectiveness of our proposed glyphic event extraction model, and more importantly, the glyphic feature can be obtained at nearly zero cost.

## Session 3

### Ellipsis and gender: some new results from French

**Anne Abeillé and Emma Kious**

(LLF, Université Paris Cité)

While some gender mismatch are attested in English (Kucerova et al. 2021, Gandon Chapela et al. 2025) (1), such mismatches are debated for languages with grammatical languages (Merchant 2014, Aparicio et al. 2015, Sudo & Spathas 2020).

For example in Greek, Merchant (2014) claims that gender mismatches are possible with adjectives and some predicative nouns, but ruled out with nobility/kinship nouns (2), and all nouns in argument position. Beyond the nobility/kinship noun class, across Romance languages, some predicative nouns (alternating or non-epicene role nouns) are claimed to rule out mismatches with a feminine antecedent (4a).

We will present a more nuanced view for French, on the basis of corpus data and experiments. While gender mismatch are attested and acceptable for adjectives in both ordering (2), we show that for gender mismatch in noun ellipsis (in predicative position), grammatical gender interacts with social gender such that mismatches are sensitive to gender biases (Mizersky et al., 2014). For instance, mismatched non-biased nouns (4a) are more acceptable than mismatched biased-nouns (4b).

The examples and references are provided separately on the following page.

### Multiple *wh*-questions in Romanian: An Empirical Approach

**Gabriela Bîlbîie**

(University of Bucharest)

Like other languages (Bulgarian, Hungarian, Serbo-Croatian, Russian, etc.), Romanian allows the alternation between coordinated *wh*-questions (CWQ) and ‘paratactic’ multiple *wh*-questions (MWQ), where two (or more) *wh*-phrases are fronted with or without a conjunction, regardless of their syntactic function. The existing literature assumes that the two patterns (CWQ and MWQ) are distinct constructions, with different semantic and syntactic properties. However, previous work has been based solely on introspection data, with many examples being artificial and lacking appropriate context. Our main goal is to confront the previous research with corpus data (CoRoLa) in order to account for the behaviour of these two patterns in actual usage.

## Ellipsis and gender: some new results from French

Emma Kiious, Anne Abeillé

While some gender mismatch are attested in English (Kučerová et al. 2021, Gandon Chapela et al. 2025) (1), such mismatches are debated for languages with grammatical languages (Merchant 2014, Aparicio et al. 2015, Sudo & Spathas 2020).

For example in Greek, Merchant (2014) claims that gender mismatches are possible with adjectives and some predicative nouns, but ruled out with nobility/kinship nouns (2), and all nouns in argument position. Beyond the nobility/kinship noun class, across Romance languages, some predicative nouns (alternating or non-epicene role nouns) are claimed to rule out mismatches with a feminine antecedent (4a).

We will present a more nuanced view for French, on the basis of corpus data and experiments. While gender mismatch are attested and acceptable for adjectives in both ordering (2), we show that for gender mismatch in noun ellipsis (in predicative position), grammatical gender interacts with social gender such that mismatches are sensitive to gender biases (Mizersky et al., 2014). For instance, mismatched non-biased nouns (4a) are more acceptable than mismatched biased-nouns (4b).

(1) Mary’s story about her family is funny, but John’s ~~story of his family~~ is not.

(2) a. \*Mary is a princess/sister, John is too.

b. \*John is a prince/brother, Mary is too.

(3) a. Alice est belle et Jean aussi ~~est beau~~.

Alice is beautiful.F.S and Jean too (is beautiful.M.S ).

b. Jean est beau et Alice aussi ~~est belle~~.

Jean is beautiful.M.S and Alice too (is beautiful.F.S ).

(4) a. Alice est une étudiante et Jean aussi ~~est un étudiant~~.

Alice is a.F.S student.F.S and Jean too (is a.M.S student.M.S).

b. Alice est une banquière et Jean aussi ~~est un banquier~~.

Alice is a.F.S banker.F.S and Jean too (is a.M.S banker.M.S).

## References

Aparicio, Helena & Franich, Katie & Xiang, Ming. 2015. Morphological feature mismatches under ellipsis: An eye-tracking study. In Bui, Thui & Özyıldız, Deniz (eds.), *NELS 45: Proceedings of the 45th Meeting of the North East Linguistic Society*: Volume 3, 1–14.

Gandon Chapela, E., (2025). The effect of gender (mis)match on the interpretation of English reflexive pronouns in cases of VP-ellipsis by native and non-native speakers of English, *Glossa*, to appear.

Kučerová, I., Chapman, C., & Moulton, K. (2021). “How to Value Gender: Lexicon, Agree and Feature Transmission under Ellipsis”.

Merchant, J. (2014). “Gender mismatches under nominal ellipsis”. In: *Lingua. Structural Approaches to Ellipsis* 151, pp. 9–32

Mizersky, J. et al. (2014). “Norms on the gender perception of role nouns in Czech, English, French, German, Italian, Norwegian, and Slovak”. In: *Behavior research methods* 46, pp. 841–871.

Sudo, Y., & Spathas, G. (2020). Gender and interpretation in Greek: Comments on Merchant (2014). *Glossa: a journal of general linguistics*, 5(1).

## Session 4

### **Dominance of F0 in mapping acoustic cues to pragmatic meaning: Eye-tracking evidence from Mandarin prosodic focus processing**

**WenXi Fei and Yu-yin Hsu**

(The Hong Kong Polytechnic University)

Understanding how multiple acoustic dimensions are mapped onto linguistic representations is important in speech perception. This study explores how native Mandarin listeners process the communicative intentions of prosodic focus by examining the perceptual weightings of F0, duration, and intensity. Using a Visual World Paradigm, thirty native Mandarin participants listened to re-synthesized audio stimuli and responded to broad-focus or narrow-focus options. Results showed that the acoustic cues significantly influenced focus interpretation, with a greater reliance on F0 than intensity and duration. Eye-tracking data revealed perceptual divergence in the F0 condition, with the divergence of looks occurring at an earlier time window for acoustic processing and later for pragmatic processing.

These findings suggest that native listeners effectively map acoustic variations to communicative demands, emphasizing the critical role of F0. The study highlights the temporal dynamics of interpreting prosodic focus, offering insights into language comprehension.

### **Prediction, semantic evaluation, and sentence constraints: A computational model of reading processing**

**Slallicchi Lavinia and Yu-yin Hsu**

(The Hong Kong Polytechnic University)

Neurolinguistic studies have challenged the traditional separation of semantic and syntactic processing, with some finding P600 but not N400 effects for semantic violations (Bornkessel-Schlesewsky & Schlewsky, 2008). Competing theories exist: the multi-stream account (Kuperberg, 2007) suggests parallel processing, while the

Retrieval-Integration account (Brouwer et al., 2012) posits sequential retrieval (N400) and integration (P600). Computational studies explore cognitive dynamics in these time windows. Surprisal predicts N400 (Michaelov et al., 2024), while semantic similarity influences P600 (Xu et al., 2024; Li & Furtell, 2024). Our study investigates how lexical predictability, semantic similarity, and uncertainty modulate cognitive effort (300-1000 ms).

Using Mandarin Chinese sentences (Jap et al., 2024), we built linear mixed-effect models to predict N400 and P600. A baseline model was compared against models with surprisal, entropy, semantic similarity (cos1: expected vs. actual word; cos5: top five expected words vs. actual word), and global models combining all features.

Results showed that surprisal explained N400, while cos5 predicted P600. Entropy accounted for most variance in both components, suggesting that uncertainty regulates cognitive effort. Findings support sequential processes: semantic prediction (N400) followed by semantic evaluation (P600). Expectations appear to be concept-based, not word-specific. Further research is needed to clarify how expectations, evaluation, and uncertainty interact.