The interplay of consituent order and givenness in German and Korean: An Expectation-based account

Torsten Kai Jachmann[1] (jachmann@lst.uni-saarland.de), Heiner Drenhaus[1], Francesca Delogu[1], Myeongju Lee[1], Matthew W. Crocker[1]
[1]Saarland University

Languages with flexible word order typically exhibit increased processing effort for non-canonical constituent orders (Bahlmann et al., 2007; Schlesewsky et al., 2003). Information status (IS) independently shapes real-time interpretation, with comprehenders systematically preferring given referents to appear before newly introduced ones (Kaiser & Trueswell, 2004; Krifka & Musan, 2012; Arnold et al., 2013). Expectation-based approaches jointly explain these patterns: canonical orders and given-before-new linearizations are more strongly expected and therefore yield lower surprisal. Under this view, implied referents—entities inferable but not explicitly mentioned—should occupy an intermediate position, being partially predictable yet not fully discourse-accessible.

Building on a series of studies by Jachmann et al. (2024), which localized Word Order (WO) and IS interactions primarily at the sentence-initial position in German, we test the broader proposal that sentence processing is guided by an expected-first principle in two self-paced reading (SPR) studies. This principle predicts that comprehenders rely heavily on expectations about the properties of the initial NP—based on discourse accessibility, grammatical role probabilities, or both—and that processing costs emerge when the initial NP diverges from these expectations. We further assess whether implied referents pattern with given or new referents or rather occupy an intermediate position.

Study 1 (German, 60 participants, age: 20–47, 120 items) manipulated three IS levels—Given, Implied, New—across Subject-first and Object-first sentences. For masculine singular nouns as used in our stimuli, German's case marking enables clear subject/object identification even in non-canonical orders, making it possible to identify WO effects at the earliest possible moment. Study 2 (Korean, 60 participants, age: 20–34, 120 items) implemented the same 3×2 design with materials translated from German. Korean is a flexible word-order language with productive topic/subject drop, yet processing reliably favors canonical SOV orders (e.g., Kim, 1992; Lee and Cho, 2003; Son, 2001); initial objects may reflect either topic-drop or OSV analyses, creating a natural test of how comprehenders resolve uncertainty about NP1's grammatical role using both structural and discourse cues.

Linear mixed-effects models contrasted SOV vs. OSV and both Given and New vs. Implied referents. German showed a graded IS hierarchy (Given < Implied < New) and early WO effects at NP1. Korean displayed a binary IS contrast (Given < Implied = New), with Implied and New patterning similarly, and WO effects emerging later at NP2 rather than on NP1. An additional NP1 interaction (SOV-Given < OSV-Given) suggests reanalysis, such that NP1 may have initially be interpreted as the object of a topic-dropped structure rather than that of a dispreferred OSV structure.

Overall, the cross-linguistic convergence provides strong support for syntactic and information structural expectancy to jointly drive comprehension, and that these expectations are rapidly updated during incremental sentence processing. Comprehenders in both German and Korean process sentences more easily when NP1 matches expectations derived from discourse context or canonical grammatical role probabilities. When these expectations are violated, processing slows, though the timing and magnitude of WO effects differ across languages in ways that interact with typological features such as case marking, argument drop, and the availability of topic interpretations. These findings support the view that incremental sentence processing relies on a unified expectation system that integrates grammatical and discourse-level cues, but whose surface manifestations depend on language-specific morphosyntactic resources.

Table 1: Example Items in the German and Korean experiments.

		Ein Bäcker ging auf ein Konzert. Ich habe gesehen, dass ... *A baker went to a concert. I saw that ...*	한 제빵사가 콘서트에 갔다. 나는 ... *A baker went to a concert. I ...*
Givenness	Word Order	German Sentence Continuation	Korean Sentence Continuation
G-first	SOV	... der Bäcker dort gestern {I:den Musiker ; N:einen Piloten} ... *... the baker [SUBJ] there yesterday {I:the musician ; N:a pilot}* [OBJ] ...	... 제빵사가 어제 그곳에서 {I:음악가를 ; N:조종사를} ...
I-first	SOV	... ein Musiker dort gestern {G:den Bäcker ; N:einen Piloten} ... *... a musician [SUBJ] there yesterday {G:the baker ; N:a pilot} [OBJ] ...*	... 음악가가 어제 그곳에서 {G:제빵사를 ; N:조종사를} ...
N-first	SOV	... ein Pilot dort gestern {G:den Bäcker ; I:den Musiker} ... *... a pilot [SUBJ] there yesterday {G:the baker ; I:the musician} [OBJ] ...*	... 조종사가 어제 그곳에서 {G:제빵사를 ; I:음악가를} ...
G-first	OSV	... den Bäcker dort gestern {I:der Musiker ; N:ein Pilot} ... *... the baker [OBJ] there yesterday {I:the musician ; N:a pilot} [SUBJ] ...*	... 제빵사를 어제 그곳에서 {I:음악가가 ; N:조종사가} ...
I-first	OSV	... den Musiker dort gestern {G:der Bäcker ; N:ein Pilot} ... *... the musician [OBJ] there yesterday {G:the baker ; N:a pilot} [SUBJ] ...*	... 음악가를 어제 그곳에서 {G:제빵사가 ; N:조종사가} ...
N-first	OSV	... einen Piloten dort gestern {G:der Bäcker ; I:der Musiker} ... *... a pilot [OBJ] there yesterday {G:the baker ; I:the musician} [SUBJ] ...*	... 조종사를 어제 그곳에서 {G:제빵사가 ; I:음악가가} ...
		... angesprochen hat. ... talked to.	... 부르는 것을 목격했다. ... call out to witnessed.

(G - Given, I - Implied, N - New, S - Subject, O - Object, V - Verb)

Table 2: Results of the German and Korean experiments.

	NP1	NP2
	GER	
I:G	$\beta = -26.64$ / $z = -3.75$	$\beta = -15.31$ / $z = -2.70$
I:N	$\beta = 26.90$ / $z = 3.41$	$\beta = 8.03$ / $z = 1.19$
WO	$\beta = 44.50$ / $z = 5.28$	$\beta = 5.66$ / $z = 0.79$
I:G & WO	$\beta = -14.16$ / $z = -1.25$	$\beta = -21.47$ / $z = -2.26$
I:N & WO	$\beta = 14.19$ / $z = 1.27$	$\beta = 0.49$ / $z = 0.04$
	KOR	
I:G	$\beta = -56.84$ / $z = -5.52$	$\beta = -65.44$ / $z = -3.34$
I:N	$\beta = -0.79$ / $z = -0.09$	$\beta = 40.70$ / $z = 1.73$
WO	$\beta = 3.18$ / $z = 0.36$	$\beta = 187.80$ / $z = 5.19$
I:G & WO	$\beta = 41.26$ / $z = 2.69$	$\beta = -23.23$ / $z = -0.55$
I:N & WO	$\beta = 12.08$ / $z = 0.77$	$\beta = 44.38$ / $z = 1.22$

(G - Given, I - Implied, N - New, WO - Word Order)

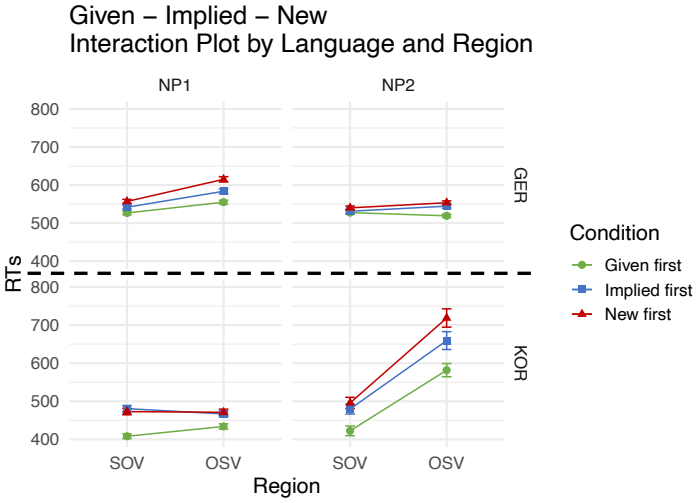

Figure 1: Residualized (by Word Length) Reading Times for German (top) and Korean (bottom) on NP1 (left) and NP2 (right) averaged by NP1 Information Status

References:

[1] Arnold et al., (2013), WIREs Cognitive Science; [2] Bahlmann et al., (2007), Human Brain Mapping; [3] Jachmann et al., (2024), HSP 2024; [4] Kaiser & Trueswell, (2004), Cognition; [5] Kim, (1992), The University of Wisconsin-Madison; [6] Krifka & Musan, (2012), In The Expression of Information Structure; [7] Lee & Cho, (2003), Studies in Generative Grammar; [8] Schlesewsky et al., (2003), Brain and Language; [9] Son, (2001), The University of Wisconsin-Madison.