

## How Domain Structure Shapes the Accessibility of Conceptual and Referential Entities

Derya Çokal & Klaus von Heusinger  
University of Cologne

Previous work has highlighted that referential accessibility is tied to prominence – the cognitive status of an antecedent in discourse [e.g., 1; 2]. However, the literature mainly focuses on structural restrictions on referential accessibility. Noun phrases in compounds create a lexical or "anaphoric island" [3] that does not allow a pronoun to access the potential antecedent. Pseudo-incorporated bare nouns show a reduced accessibility of pronouns [4] and indefinite DPs in opaque context are also not accessible for referential pronouns [5]. While these restrictions are well investigated, there is little work on the structural restrictions of conceptual entities (kinds, properties) or the accessibility for conceptual ("sense") anaphora. [6] observes that a plural pronoun can access the concept (or kind) in an opaque context. There is, however, no empirical study on the accessibility of conceptual entities in lexical islands, pseudo-incorporated bare nouns or opaque contexts. Since predictability can influence acceptability, we included GPT-2 surprisal to control this effect. Given that lexical islands categorically block and pseudo-incorporated and opaque contexts reduce the accessibility of referential anaphora, we investigate how these structural constraint the accessibility of conceptual referents. We contrast the German underspecified demonstrative pronouns *das* ('that') for accessing conceptual entities [Experiment 1] with the singular personal pronouns *sie* ('it') for referential entities [Experiment 2] across four domains (see Table 1 for the conditions across the experiments):

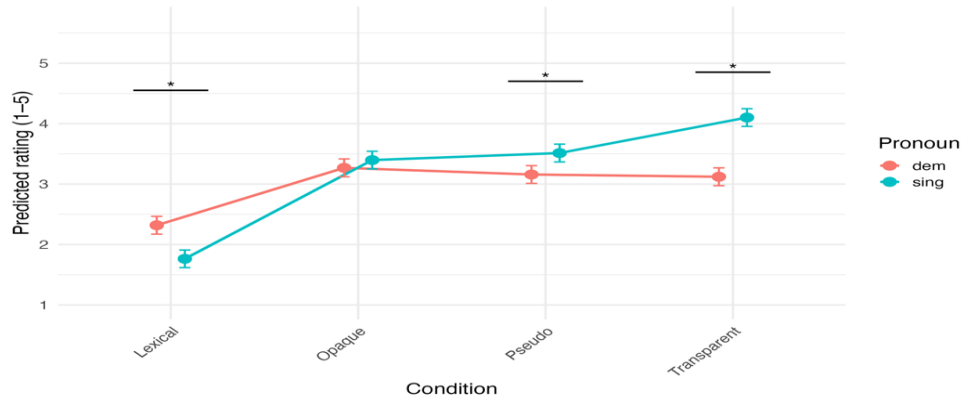
- (1) **Lexical islands** (compounds such as *Zeitungsleser in I am talking to a newspaper reader.*), which block the accessibility of the referential and conceptual entities (i.e., *Zeitung*).
- (2) **Referential opacity (semantic domain:** e.g., *I would like to buy a newspaper.*), where non-specificity lowers prominence; (3) **Pseudo-incorporated direct objects (syntactic domain:** e.g., *I read a newspaper.*), which weaken the accessibility of the embedded noun;
- (4) **Referential transparency** (e.g., *I am buying a newspaper.*), where an antecedent remains maximally prominent and accessible.

We predict that in (1a) and (2a) in Table 1, lexical island would block referential and conceptual accessibility, even though conceptual might be available under restricted conditions. Since semantic opacity and syntactic domains in (2b/c) and (3b/c) have not been tested for conceptual anaphora, these conditions were exploratory but theoretical literature propose that conceptual entities are accessible. Referential transparency in (2d) was the control condition, favouring singular reference over conceptual *das* in (1d). We estimated LLM-based surprisal for each stimulus using GPT-2 [7] to test whether human acceptability aligns with distributional expectations from language experience. We conducted acceptability-judgment Experiments with 308 German speakers. **The results of linear mixed-effects model** reveal a robust hierarchy of accessibility (Fig.1). Transparent contexts yielded the highest ratings ( $M = 3.61$ ,  $SE = 0.05$ ), followed by pseudo-incorporated ( $M = 3.34$ ,  $SE = 0.05$ ) and opaque ( $M = 3.33$ ,  $SE = 0.05$ ) contexts, with lexical islands showing the lowest acceptability ( $M = 2.04$ ,  $SE = 0.05$ ). Our model confirmed a significant effect of domains ( $\beta = 1.535$ ,  $SE = 0.10$ ,  $p < .001$ ). Pronoun type significantly interacted with domains. Demonstrative conceptual anaphora (*das*) was slightly more acceptable than singular anaphora (*sie*) in lexical islands ( $\beta = 0.555$ ,  $p < .001$ ). *Das* was less preferred than *sie* in pseudo-incorporation—domains where antecedents are blocked ( $\beta = -0.355$ ,  $p < .001$ ). No reliable difference emerged in opaque contexts ( $p = .089$ ). The acceptability of *sie* increased in referential transparency contexts, consistent with the prediction that individual reference requires active referential concepts. Surprisal estimates had a small significant negative effect ( $\beta = -0.089$ ,  $t = -2.67$ ,  $p < .001$ ), indicating that LLM-derived predictability aligns with human acceptability judgments. Adding surprisal estimates did not alter the core theoretical pattern, suggesting that conceptual anaphora reflect cognitive mechanisms beyond statistical regularities captured by language models. These results demonstrate conceptual anaphora are significantly more resilient to syntactic and semantic domain restrictions than singular individual anaphora.

**Table 1.** Conditions across the acceptability-judgment experiments

Domain Type	Underspecified Pronoun (Exp. 1) [40 items; 80 fillers]	Singular Pronoun (Exp. 2) [40 items, 80 fillers]
<b>Lexical island:</b> Ich spreche mit <b>einem Zeitungsleser</b> . I am talking to a newspaper reader.	<b>(1a)</b> <i>Das sind</i> <sup>1</sup> informative Medien und überall erhältlich. <i>That are</i> <sup>1</sup> informative media and available everywhere.	<b>(2a)</b> <i>Sie hat</i> ansprechende Bilder und ist frisch gedruckt. <i>It has</i> appealing images and is freshly printed.
<b>Referentially opaque:</b> Ich möchte <b>eine Zeitung</b> kaufen. I would like to buy a newspaper.	<b>(1b)</b> <i>Das sind</i> informative Medien und überall erhältlich. <i>That are</i> informative media and available everywhere..	<b>(2b)</b> <i>Sie hat</i> ansprechende Bilder und ist frisch gedruckt. <i>It has</i> appealing images and is freshly printed.
<b>Pseudo-incorporated:</b> Ich lese <b>Zeitung</b> . I read the newspaper.	<b>(1c)</b> <i>Das sind</i> informative Medien und überall erhältlich. <i>That are</i> informative media and available everywhere.	<b>(2c)</b> <i>Sie hat</i> ansprechende Bilder und ist frisch gedruckt. <i>It has</i> appealing images and is freshly printed.
<b>Referentially transparent:</b> Ich kaufe <b>eine Zeitung</b> . I am buying a newspaper.	<b>(1d)</b> <i>Das sind</i> informative Medien und überall erhältlich. <i>That are</i> informative media and available everywhere.	<b>(2d)</b> <i>Sie hat</i> ansprechende Bilder und ist frisch gedruckt. <i>It has</i> appealing images and is freshly printed.

Note<sup>1</sup>: The pronoun “das” is the singular neuter demonstrative pronoun in German. The copula verb agrees in this kind of (specification) copular sentences in German (and Dutch) with the second DP, while in English with the first DP: *Das größte Problem sind /\*ist deine Eltern*. (The biggest problem is your parents) (cf. Hartmann & Heycock, 2020).



**Fig 1.** Predicted acceptability of singular vs. underspecified pronoun across lexical island, opaque (semantic domain), and pseudo-incorporated direct object (syntactic domain), referentially transparent domain. Dem corresponds to (das). Sing is singular pronoun (sie).

[1] Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). *Cognitive status*. *Language*, 69, 274–307.  
 [2] von Stechow, K., & Schumacher, P. S. (2019). *Discourse prominence*. *J. Pragmatics*.  
 [3] Postal, P. (1969). *Anaphoric islands*. *CLS*, 5, 205–239.  
 [4] Krifka, M., & Modarresi, F. (2023). *Anaphoric potential of bare nominals*. Benjamins.  
 [5] Karttunen, L. (1969). *Discourse referents*. *Proc. Int. Conf. on Computational Linguistics*.  
 [6] Gernsbacher, M. A. (1991). *Comprehending conceptual anaphors*. *Lang. & Cogn. Proc.*  
 [7] Radford, A., Wu, J., & Child, R. (2019). *Language models...* OpenAI report.  
 [8] Fruehwald, J. (2022). *LIN517: NLP — ngrams & perplexity*.  
 [9] Hartmann, J. M., & Heycock, C. (2020). *(Morpho)syntactic variation in agreement*. *Front. Psychol.*, 10:2994.