

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

When you say that, do you mean it? Crosslinguistic patterns of anaphor resolution in English, German, and Polish

Permalink

<https://escholarship.org/uc/item/64m534qg>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors

Marx, Elena
Iwan, Oliwia Anna
Wittenberg, Eva

Publication Date

2023

Peer reviewed

When you say *that*, do you mean *it*?

Crosslinguistic patterns of anaphor resolution in English, German, and Polish

Elena Marx (marx_elena@phd.ceu.edu)

Central European University, Department of Cognitive Science, Quellenstraße 51
Vienna, 1100 Austria

Oliwia Iwan (iwano@ceu.edu)

University of Vienna, Institute of Linguistics, Sensengasse 3a
Vienna, 1090 Austria

Eva Wittenberg (wittenberge@ceu.edu)

Central European University, Department of Cognitive Science, Quellenstraße 51
Vienna, 1100 Austria

Abstract

When hearing a pronoun, people find its referent effortlessly most of the time. However, across languages, pronominal systems vary: While in one language, a pronoun may point to a referent as a function of its accessibility in discourse, in others, pronoun resolution might rely on a range of different processes, specific to each individual pronominal form. In three studies, using an act-out task in English, German and Polish, we found evidence for an overarching tendency, but also crosslinguistic differences: In general, participants were more likely to relate simple pronouns to single, most salient referents and demonstratives to conceptual composites, but cross-linguistic differences reflect the complexity of each language's pronominal system. Overall, our results extend the empirical basis for anaphora resolution, refining a model of anaphora resolution as a multifaceted interaction of various linguistic and non-linguistic mechanisms at its core.

Keywords: anaphora resolution, demonstratives, pronouns, cross-linguistic comparison

Introduction

Pronouns function as shortcuts in discourse: After mentioning that I broke the big, polka-dot coffee mug my best friend gave me for my birthday, using a pronoun *it* enables me to efficiently share novel information about that mug without repeating the whole complex noun phrase. In that sense, pronouns index and sometimes bundle mental representations which have been previously evoked by linguistic or extralinguistic context (Gundel, Hedberg, & Zacharski, 1993; Tily & Piantadosi, 2009; Trott, Bergen, & Wittenberg, 2022).

Without doubt, pronominal indices are highly efficient in language production. From a comprehender's point of view, however, they often introduce ambiguity: In (1), *it* can either refer to the Theme of the previous sentence (i.e., *the mug*), the Goal (i.e., *the saucer*) or to a combination of both (i.e., *the mug on top of the saucer*).

- (1) *I placed the mug on the saucer. Then I carried it to the kitchen counter.*

A central question is which factors determine how pronominal reference to objects is resolved during language comprehension. Specifically, what are the psycholinguistic mechanisms, guiding people to link pronominal elements to their mental representations?

One prominent approach relates the interpretation of pronouns to the accessibility of discourse referents (e.g. Ariel, 1990; Givón, 1983; Gundel, Hedberg, & Zacharski, 1988; Gundel et al., 1993). Under the Accessibility Hierarchy approach, simple pronouns restrict the comprehender's search for a referent to topical or maximally salient linguistic entities in working memory, privileging topical subject noun phrases from the preceding linguistic context. Demonstrative pronouns, on the other hand, signal to select non-topical, but activated referents from previous discourse (e.g., less salient object NPs). In doing so, demonstratives shift the focus of attention to the new indexed entities (Diessel, 1999; Grosz, Joshi, & Weinstein, 1995; Grosz & Sidner, 1986; Gundel et al., 1988; Krasavina & Chiarcos, 2007). Accessibility-based accounts thus presume that people keep track of saliency differences between linguistic constituents and establish reference according to this dimension. Importantly, accessibility still remains one of the most prominent factors of analyzing pronoun resolution (see in robotics e.g., Pal, Clark, & Williams, 2021; Williams & Scheutz, 2017; or Xu & Xiang, 2021, for language processing).

However, increasing evidence in psycholinguistic research, mostly on English, has suggested that reference resolution differs across individual pronominal lexemes (e.g., Brown-Schmidt, Byron, & Tanenhaus, 2005; Kaiser, 2011; Kaiser, Runner, Sussman, & Tanenhaus, 2009; Kaiser & Trueswell, 2011; Wittenberg, Momma, & Kaiser, 2021). Rather than generally relying on access to previous linguistic discourse as a determining principle of reference, speakers and listeners seem to monitor a variety of linguistic, but crucially, also non-linguistic factors and weigh them differently for each referential expression.

For instance, a variety of studies has shown that people are sensitive to syntactic constraints (e.g., Chow, Lewis, & Phillips, 2014; Kaiser et al., 2009), information structure

(e.g., Kaiser & Trueswell, 2011), coherence relations (e.g., Kaiser, 2011), verbal features (e.g., Bevacqua, Loáiciga, Rohde, & Hardmeier, 2021; Francey & Cain, 2022) as well as to the type of referred entity, privileging concrete entities for simple pronouns and events or propositions for demonstratives (e.g., Bevacqua et al., 2021; Çokal, Sturt, & Ferreira, 2018; Loáiciga, Bevacqua, Rohde, & Hardmeier, 2018; Wittenberg et al., 2021).

Unlike the Accessibility Hierarchy approach, reference-specific accounts reject a single unifying mechanism that tracks saliency in discourse and argue that processes of reference resolution differ between pronouns: While simple pronouns may indeed trigger a quick search for the most accessible NP, demonstratives tend to function as “conceptual bundlers” that merge bits of conceptual structure into single linguistic units to make them available for subsequent reference (Wittenberg et al., 2021).

While pronominal reference has been extensively studied cross-linguistically, much of the research on object and event reference has focused on English, investigating different interpretative preferences for *it* and *that* (for exceptions see Bevacqua et al., 2021). For instance, Brown-Schmidt et al. (2005) asked participants to manipulate objects (e.g., a cup and a saucer), with imperatives that either contained a simple or a demonstrative pronoun and could refer to the Theme (cup), the Goal (saucer) or to their conceptual composite (cup on saucer). Results showed that people tended to choose the Theme when hearing *it*, but the composite when hearing *that*.

Here we ask: Is such a distributional pattern cross-linguistically stable? Given that pronominal paradigms vary across languages, comprehenders might rely on different criteria when they search referents. To answer this question, we therefore used English, German, and Polish – three languages with informative similarities and differences in their pronominal systems (Table 1).

	English	German	Polish
simple pronouns	<i>it</i>	<i>er/sie/es</i>	<i>go/ją/je</i>
d-pronouns	-----	<i>der/die/das</i>	-----
demonstrative pronouns	<i>this</i> <i>that</i>	<i>diese/r/s</i> <i>jene/r/s</i>	<i>ten/ta/to</i> <i>tamten/a/o</i>

Table 1: Pronominal paradigm in English, Polish and German (in Singular Nominative). Forms used here are marked in bold.

In English, non-human entities are not gender-marked, so there is only a distinction between *it* and the two demonstratives *this* and *that*.

In German, both simple and demonstrative pronouns are marked for gender. A particularity of the German pronoun system, however, is the existence of so-called *d*-pronouns, which are largely form-congruent with definite determiners. Crucially, *d*-pronouns have been argued to behave complementary to simple pronouns: they should never refer to the maximally salient referent. Unlike *dieser* und *jener*, which are associated with a more formal register however, they can be used without marking the contrast of distal and proximal referents (Bader & Portele, 2019; Bosch &

Hinterwimmer, 2016; Bosch & Umbach, 2007; Helbig & Buscha, 2001; Patil, Bosch, & Hinterwimmer, 2020; Portele & Bader, 2017). Thus, *d*-pronouns are most similar to English and Polish demonstratives.

Finally, in Polish, third-person simple pronouns are marked for grammatical gender. Furthermore, in addition to gender-marked demonstratives (as well as a contrastive distal demonstrative *tamten*), Polish has the gender- and number-neutral demonstrative *to* for generic object reference (Brooks, 1975; Czardybon, 2017; Fokker & Smolikowska, 1971). Polish *to* is therefore comparable to the English *that*.

Based on these cross-linguistic differences as well as previous findings, several predictions arise: The first concerns the differential distribution of single object and composite choices for different pronoun types. For English, we expect to replicate Brown-Schmidt et al.’s (2005) findings: People should understand the pronoun *it* as primarily referring to single objects (e.g., the most accessible NP) and the demonstrative *that* as referring to conceptual composites. For German and Polish, pronoun resolution might either mirror the expected English pattern – as predicted by reference-specific accounts – or follow from the accessibility of discourse referents. In such a case, people should choose the most accessible NP for simple pronoun, and the activated NP for demonstrative pronoun reference.

The second prediction is specific to within-single-object choices: People should choose Goal objects significantly more often in cases of demonstrative pronoun reference – in line with predictions of the accessibility-based accounts.

Our third prediction addresses crosslinguistic differences: Here, we expect German *d*-pronouns to be driven by information structure as described in the literature, resulting in higher proportion of Goal choices for demonstratives compared to English. Furthermore, given its similarities with *that* in English, we predict the Polish demonstrative *to* to align with English interpretative preferences.

Conceptually replicating Brown-Schmidt et al. (2005), we used an act-out task adopted from Wiese et al. (2020) to tap into people’s preferred interpretations of pronominal reference as a function of pronoun type. The rationale of the method was that people’s enactments, more than other measures, give somewhat unfiltered access to the referential interpretation of a given anaphora. In Experiment 1, we contrasted the English pronouns *it* and *that* in a context that rendered them potentially ambiguous between a previously mentioned Theme or Goal object or their conceptual combination. Experiment 2 was conducted in German, using gender-matched simple pronouns *er/sie/es* and *d*-pronouns *der/die/das*. In Experiment 3, we compared people’s referent choices for Polish gender-marked simple pronouns and the unmarked generic demonstrative *to*.

Experiment 1: English

Participants

We recruited 32 English speakers from CEU's internal recruitment platform. Experimental sessions lasted approximately 35 minutes. Only complete data sets were included in the statistical analysis, and when participants performed 5 out of 8 sanity check trials correctly. Based on these criteria, no participant was excluded.¹

Materials

We created eight object sets for the act-out task (see Fig. 1). Each set contained four toys: one human agent (e.g., pilot), at least one animal (e.g., horse) and at least one artifact (e.g., a broom and a basket).

As linguistic prompts, we constructed two instruction paragraphs per object set, consisting of an introductory sentence and a target sentence. Introductory sentences (1a) established a containment or support relationship between two toys, one of which was introduced as a Theme (i.e., the object being moved or moving on its own) and the other as a Goal (i.e., target location of the Theme, see 1a). Target sentences described a change of location, either using the pronoun *it* or the demonstrative *that* (1b).

- (1) a. *The pilot put the broom into the basket.*
- b. *Then he placed it/that next to the horse.*

Each instruction paragraph was concluded by two filler sentences (e.g., *The horse ran into the basket*).

Furthermore, for each object set, we included three independent filler trials. To control for participant's attentiveness, an additional sanity check trial was included, describing two sequential events in a complex sentence (e.g., *The pilot climbed on the back of the horse that had previously drunk from the basket*). Sentences were pre-recorded via an online text-to-speech software, and presented using PsychoPy, which was also used for live-coding; each experimental session was recorded on video.

Procedure

Object sets were presented as blocked trials: Each set started with a short familiarization in which participants were asked to sort the toys to their matching label (Fig. 1). After familiarization, each trial followed the same three steps: First, a pre-recorded sentence was played. In order to prevent participants from using two hands (and manipulating more than one object at a time), participants held a rubber duck in their non-dominant hand and squeezed it to emit a squeak after each sentence. After the squeak, participants then enacted the sentence they heard with their dominant hand, using the provided toys.

¹ For logistic reasons, we were unable to reach the preregistered sample size of 48 participants in English and Polish. Full results will be published at <https://osf.io/9nt6p> after complete data collection.

Three practice trials were executed with a separate object set. After practice, object sets were presented in a pseudo-randomized order. Participants always performed 2 pronoun instruction paragraphs (see 1), one sanity check and 3 filler trials before a new object set was introduced. Within sets, trial order was pseudo-randomized. There were 99 trials in total: 3 practice trials, 8 sanity checks, 24 unrelated fillers, and 64 trials per critical instruction paragraphs, with 8 critical trials per pronoun condition (*it* vs. *that*). Participants were assigned randomly to one of four lists, counterbalanced by a Latin-square design.

Statistical Analysis

A series of two pre-registered nested regression models was constructed which all included pronoun type as independent contrast-coded variable, and people's choice of referent as the dependent variable (see Fig. 2). The dependent variable was coded differently across the two models: In the first model, we examined the effects of pronoun type on people's conceptual composite and single object choices. Single object choices therefore collapsed over different choice possibilities, provided by the linguistic antecedents in the introductory sentence (i.e., the Theme or Goal). In the second model, we asked whether different pronouns privilege different single choice referents. We therefore only compared Theme and Goal choices as a function of pronoun type, excluding trials with conceptual composite choices from the analysis.

For both models, we built a mixed effects logistic regression in the R statistics environment (R Core Team, 2014) with pronoun type as fixed effect and participants and object sets as random intercepts, controlling for variability caused by these two factors. We included a minimal intercept-only structure (i.e., no random slopes), to ensure model convergence. For the same reason, we excluded object sets as random intercepts from model 2. Statistical significance was assessed by model comparisons. A preregistration of all experiments, all materials, data, and analyses are public at <https://osf.io/9nt6p>.

Preliminary Results

Participants passed 93% of the sanity checks, demonstrating that performing the task was easy and that interest did not drop over the course of the trials.



Figure 1: Example of an object set.

English speakers' referent choices are illustrated in Figure 3 (left two bars). As predicted, participants were more likely to choose one of the single objects after hearing *it* (mean=0.55, SD=0.50), compared to hearing *that* (mean=0.22, SD=0.42, $z=7.86$, $p<0.0001$), when they preferred moving the two objects combined (mean of single object choices in each condition). Zooming in on single object choices in Model 2, we found a main effect of pronoun type: Participants were less likely to choose the goal object when critical sentences contained a simple pronoun (mean=0.02, SD=0.17) than when they heard a demonstrative pronoun reference (mean=0.21, SD=0.41, $z=3.00$, $p=0.003$).

Experiment 2: German

Participants

48 German participants were recruited from CEU's and Vienna CogSciHub recruitment systems. Only complete data sets were analyzed, and participants had to meet the sanity check criterion. As in Experiment 1, no participant was excluded.

Materials and Procedure

We constructed German equivalents to the English sentence stimuli in Experiment 1. Pronoun instruction sets started with an introductory sentence (2a), followed by a critical sentence (2b) that either contained a simple pronoun (*er/sie/es*) or a *d*-pronoun (*der/die/das*) and two additional fillers.

- (2) a. *Der Pilot stellte den Besen in den Korb.*
The pilot put the broom_{M.SG} in the basket_{M.SG}
- b. *Nun stellte er ihn/den neben den Löwen.*
Then put he pro_{M.SG}/that_{M.SG} next to the lion

Since gender is grammatically marked in the German pronoun system, Theme and Goal pairs were matched in grammatical gender, such that pronoun and *d*-pronoun references were referentially ambiguous.

Object sets containing an agent, one or two animal or artifact toys were matched to the linguistic materials. The experimental procedure was the same as in Experiment 1.

Statistical Analysis and Preliminary Results

The same statistical analysis was conducted as in Experiment 1. Participants performed correctly in 93% of all sanity checks. This suggests that they understood the task and remained attentive throughout the experimental session.

Experiment 2 yielded a similar pattern of results as Experiment 1: There was a main effect of pronoun type in both regression models so that participants chose composites more often when hearing *d*-pronouns (mean=0.54, SD=0.50) than when hearing simple pronouns (mean=0.80, SD=0.40, $z=7.99$, $p<0.0001$). Furthermore, they moved the Goal more often for *d*-pronouns (mean=0.10, SD=0.30) than for simple

pronouns (mean=0.02, SD=0.15, $z=3.63$, $p=0.0003$, see Fig. 3, middle bars).

Experiment 3: Polish

Participants

We recruited Polish native speakers via CEU's internal recruitment system, as well as via language exchange projects (e.g., language tandem cafes, online forums). 20 full data sets were collected, with no exclusions.¹

Materials and Procedure

For the Polish experiment, we constructed instruction sets that closely followed the English and German materials: An initial sentence (3a) was followed by a critical sentence, either containing a simple pronoun (*go/ją/je*) or a demonstrative (*to*, see 3b). Theme and Goal were matched in grammatical gender, so that Polish simple pronouns were referentially ambiguous.

- (3) a. *Pilot postawił beczkę na ciężarówkę.*
Pilot put barrel_{F.SG} on truck_{F.SG}
- b. *Nagle przeniósł ją/to przed owcę.*
Suddenly carried pro_{F.SG}/that in front of sheep

As Polish is a pro-drop language that tends to omit subjects (McShane, 2009), pronoun manipulation in (3b) always related to the sentential objects of the introductory sentence (3a), to keep pronominal elements largely unmarked. Two filler sentences concluded each trial; object sets were tailored to the linguistic materials. The experimental procedure was adopted from Experiments 1 and 2.

Statistical Analysis and Preliminary Results

The statistical analysis was the same as in Experiments 1 and 2. As in previous experiments, performance accuracy was high at 96%.

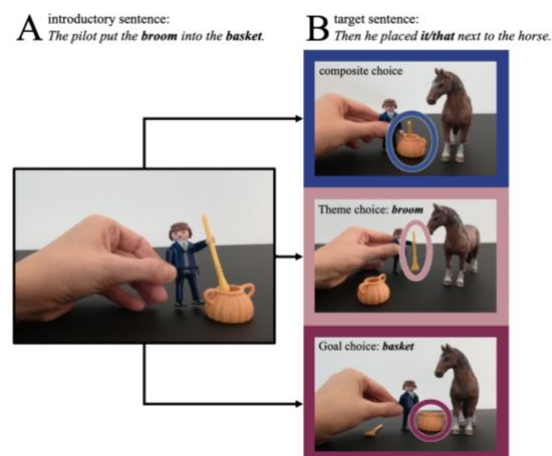


Figure 2: (Staged) example of an enactment of the introductory sentence (A) and possible referent choices in the target sentence (B).

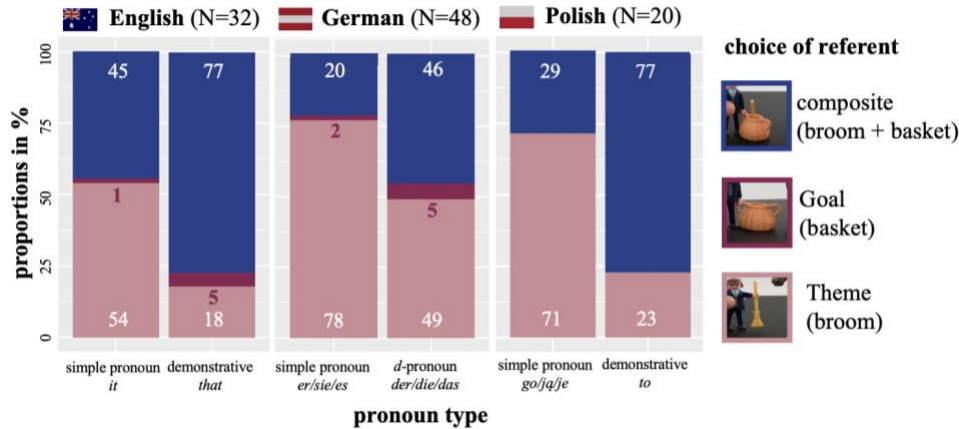


Figure 3: Proportional reference choices as a function of pronoun type in English (left), German (middle) and Polish (right), digits represent percentage of Theme, Goal and composite choices.

For Model 1, the Polish data replicated the pattern of results found in both previous experiments (see Fig. 3, rightmost bars): Participants were more likely to choose single objects for simple pronoun reference (mean=0.71, SD=0.45), while moving both objects more often when instructions included a demonstrative (mean=0.23, SD=0.42, $z=8.07$, $p<0.0001$). Unlike in the previous experiments, however, participants never chose Goal objects in reference to either pronoun type, rendering the second logistic regression superfluous.

Cross-linguistic Comparison

Statistical Analysis

We pooled Polish, German and English data together and ran two nested logistic regression analyses of the main effects of pronoun type, language, as well as the interaction between the two predictors. In line with the language-specific analyses, we first compared conceptual composite and single object choices (i.e., the Theme or Goal) in Model 1, and excluded conceptual composite trials to investigate the effect of pronoun type on Theme and Goal choices in Model 2. Languages were coded using a Helmert coding scheme: English and German referent choices were grouped as one Helmert contrast pair and compared with the mean of Polish referent choices as a second Helmert contrast pair.

Furthermore, we ran planned pairwise comparisons within pronoun types, to investigate the interaction between language and pronoun type on people's referent choices more thoroughly. Since pairwise comparisons were pre-registered, we did not adjust p -values (see <https://osf.io/9nt6p>).

Preliminary Results

We excluded Model 2 because of sparse data points for people's goal object choices with pooled languages. Comparing people's composite and single-object choices in Model 1, we found a main effect of language: Polish participants interpreted pronouns more often in reference to conceptual composites (mean=0.47) compared to English and German participants (mean=0.51), while, at the same

time, English participants performed significantly different (mean=0.39) from German native speakers (mean=0.63, $Df=2$, $\chi^2=18.78$, $p<0.0001$).

Furthermore, there was a main effect of pronoun type such that conceptual composite choices generally occurred more often after demonstratives (mean=0.38) than after simple pronoun reference (mean=0.70, $Df=1$, $\chi^2=220.39$, $p<0.0001$).

Importantly, the interaction between the two fixed predictors was significant ($Df=2$, $\chi^2=13.59$, $p=0.001$). Pairwise comparisons between individual languages revealed that, for demonstratives, the interaction was driven by the significantly higher proportion of single object choices of German as compared to English ($\beta=-1.14$, $z=-5.75$, $p<0.0001$) and Polish participants ($\beta=-1.09$, $z=-4.23$, $p<0.0001$). For simple pronouns, on the other hand, English speakers differed significantly from the other language groups, yielding more composite choices than Polish ($\beta=0.76$, $z=2.81$, $p=0.005$) or German participants ($\beta=0.92$, $z=5.03$, $p<0.0001$).

General Discussion

Our study investigated if speakers of different languages would employ different mechanisms when they map the pronominal elements of their language onto the space of possible referents. More precisely, we asked whether accessibility of discourse entities would be the sole determinant of pronoun resolution or whether different criteria would come into play depending on specific pronominal forms. Furthermore, we asked whether reference resolution is mediated by the specific properties of a language's pronominal systems.

In line with our first prediction, we largely replicated Brown-Schmidt et al.'s (2005) data pattern for English, and found a similar main effect in German and Polish: Compared to demonstratives, simple pronouns were more driven by accessibility considerations, frequently relating to the maximally salient referent of the previous discourse – in our case, the Theme of a discourse. At the same time, demonstratives privileged composites that comprised two

previously mentioned entities. In line with our second prediction, demonstratives furthermore evoked more Goal referent choices than simple pronouns, although Goal interpretations were overall rather low, and absent in Polish.

This pattern of results was stable across languages, but its profiles varied. For one, English participants had a composite bias, compared to German and Polish speakers, while German speakers had a Theme bias, against our first cross-linguistic prediction according to which pronoun resolution in German should primarily draw on information structural differences. The second cross-linguistic prediction, however, was confirmed, with the Polish demonstrative *to* mirroring its English counterpart *that*.

In principle, our results lend support to reference-specific accounts that assume different mechanisms of reference resolution for different pronoun types (e.g., Kaiser & Trueswell, 2011; Wittenberg et al., 2021): Not only did simple and demonstrative pronouns differ systematically within each of the three languages, but moreover, specific items took effect on how reference was resolved between languages. The Accessibility Hierarchy approach, on the other hand, can only account for the data found here on simple pronoun interpretations, apart from the English composite bias for *it*.

Notably, such a composite preference for *it* deviated from Brown-Schmidt et al.'s (2005) findings where Theme choices generally prevailed in the *it* condition (but see their Exp. 2b). One reason for this discrepancy may relate to the stimulus modifications we included, using transitive and ditransitive declaratives instead of imperatives, and inanimate as well as animate referents as potential conceptual composites (e.g., bird on dragon). Increased composite choices for *it* might therefore reflect the greater variability in our materials. Alternatively, instruction paragraphs in our study only established support and containment relations, whereas the original study also included non-contact relations (i.e., *next to*) that generally rendered conceptual composites less accessible and boosted single object choices. Our focus on support and containment relations may have reduced participant's sensitivity towards single object referents.

Another open question is: Why did German *d*-pronouns not show sensitivity towards information structural differences as standardly predicted by the literature (Bader & Portele, 2019; Bosch & Umbach, 2007; Patil et al., 2020; Portele & Bader, 2017; Schumacher, Dangl, & Uzun, 2016)? In our study, *d*-pronouns rarely selected the less salient discourse referent (i.e., the Goal), but tended to point to Theme referents, similar to simple pronouns. While shifting discourse focus does not appear to be the primary function of German *d*-pronouns, their behavior was nevertheless different from English and Polish demonstratives as people generally privileged referents with a linguistic antecedent. One possibility of explaining these results relates to our auditory stimuli: In order to differentiate *d*-pronouns from definite articles, vowels were artificially lengthened, resulting in subtle clicks in some items. This slight acoustic marker may have led

participants to expect a simple noun following the *d*-pronoun, biasing their interpretations towards single object referents.

Here, however, we propose an alternative explanation, aligning the single object bias in German with the composite bias for *it* in English. Although the cross-linguistic differences we found for pronoun resolution should be taken with a grain of salt, given that the English and Polish data collection is not completed,¹ according to our view, they may bear on the complexity within the languages' pronominal system: Speakers of languages with a more differentiated pronominal paradigm were more likely to attach pronominal forms to single object referents. For simple pronouns, this was the case in both German and Polish where third-person pronouns are invariably gender-marked, and for demonstratives, this was the case in German, which unlike Polish and English lacks a gender-neutral demonstrative form. In that sense, our results could be taken as suggesting that obligatory gender-marking generally narrows down people's search space to single object referents.

But why would this be the case? We suggest it may be a result of statistical learning in both the domain of morphosyntactic structures and how they map onto semantic space: We know that speakers track the probabilities of pronouns for a given referential context and adjust their interpretation strategies accordingly (e.g., Arnold, Strangmann, Hwang, Zerkle, & Nappa, 2018; Johnson & Arnold, 2022).

In line with such findings, we propose that resolving reference to specific pronominal forms may be driven by cross-linguistic differences in referential frequencies: As gender-neutral pronouns (i.e., English *it/that*, Polish *to*) may apply to a broad range of referential contexts, relating to single object referents regardless of grammatical gender as well as object composites, events and propositions, they may be more readily treated as linguistic operators that resolve to conceptual structure bundles. Certain gender-marked pronouns, on the other hand (i.e., Polish *go/ją* and German *die/der*), establish reference conditional on grammatical gender specifications, frequently picking out single object referents, but never events, and object composites only in cases where the components' grammatical gender match. The frequency of single object reference of a pronoun within a language's pronominal paradigm, compared to the frequency of conceptual bundle reference, may therefore create different biases of reference resolution across languages. To target this question, follow-up studies are planned, systematically investigating whether gender-marked Polish demonstratives *ten/ta* shift interpretative preferences towards single object choices, similar to *d*-pronouns, or whether neuter forms in German (i.e., *das/es*) systematically exhibit a composite bias.

In sum, our data suggest that pronoun resolution processes rely on different factor weights across languages, depending on each language's pronominal system; and the factors weighted are both linguistic and non-linguistic, supporting a model of anaphora resolution that connects to linguistic processes as tightly as to broader cognition.

Acknowledgements

We thank Attila Balla for his support in data collection, Jesse Snedeker, Alper Demircan, and Natalia Jardon Perez for their critical comments and review and, of course, the SWs Julius and Nils for providing their toys to scientific purposes.

References

- Ariel, M. (1990). *Assessing Noun-Phrase Antecedents*. New York, London: Routledge.
- Arnold, J. E., Strangmann, I. M., Hwang, H., Zerkle, S., & Nappa, R. (2018). Linguistic experience affects pronoun interpretation. *Journal of Memory and Language*, 102(December 2017), 41–54. <https://doi.org/10.1016/j.jml.2018.05.002>
- Bader, M., & Portele, Y. (2019). The interpretation of German personal pronouns and d-pronouns. *Zeitschrift Fur Sprachwissenschaft*, 38(2), 155–190. <https://doi.org/10.1515/zfs-2019-2002>
- Bevacqua, L., Loáiciga, S., Rohde, H., & Hardmeier, C. (2021). Event and entity coreference across five languages: Effects of context and referring expression. *Dialogue and Discourse*, 12(2), 192–226. <https://doi.org/10.5210/DAD.2021.207>
- Bosch, P., & Hinterwimmer, S. (2016). Demonstrative pronouns and perspective. In P. G. Grosz & P. Patel-Grosz (Eds.), *The Impact of Pronominal Form on Interpretation* (pp. 189–220). Berlin, Boston, MA: De Gruyter Mouton.
- Bosch, P., & Umbach, C. (2007). Reference determination for demonstrative pronouns. *ZAS Papers in Linguistics*, 48(1), 39–51. <https://doi.org/10.21248/zaspil.48.2007.353>
- Brooks, M. Z. (1975). *Polish Reference Grammar*. The Hague: Mouton.
- Brown-Schmidt, S., Byron, D. K., & Tanenhaus, M. K. (2005). Beyond salience: Interpretation of personal and demonstrative pronouns. *Journal of Memory and Language*, 53(2), 292–313. <https://doi.org/10.1016/j.jml.2005.03.003>
- Chow, W. Y., Lewis, S., & Phillips, C. (2014). Immediate sensitivity to structural constraints in pronoun resolution. *Frontiers in Psychology*, 5, 1–16. <https://doi.org/10.3389/fpsyg.2014.00630>
- Çokal, D., Sturt, P., & Ferreira, F. (2018). Processing of It and This in Written Narrative Discourse. *Discourse Processes*, 55(3), 272–289. <https://doi.org/10.1080/0163853X.2016.1236231>
- Czardybon, A. (2017). *Definiteness in a language without articles: A study on Polish*. <https://doi.org/10.1515/9783110720426>
- Diessel, H. (1999). *Demonstratives: Form, function and grammaticalization*. Amsterdam, Philadelphia: John Benjamins.
- Fokker, A. A., & Smolikowska, E. (1971). *Anatomy of a Word-Class: A Chapter of Polish Grammar*. The Hague: Mouton.
- Francey, G., & Cain, K. (2022). The influence of the temporal characteristics of events on adults' and children's pronoun resolution. *Journal of Child Language*, 1–26. <https://doi.org/10.1017/S0305000921000878>
- Givón, T. (1983). *Topic continuity in discourse: The functional domain of switch-reference*. <https://doi.org/10.1075/tsl.2.06giv>
- Grosz, B. J., Joshi, A. K., & Weinstein, S. (1995). Centering: A Framework for Modelling the Local Coherence of Discourse. *Computational Linguistics*, 21(2), 203–225.
- Grosz, B. J., & Sidner, C. (1986). Attention, Intentions, and the Structure of Discourse. *Computational Linguistics*, 12(3), 175–204.
- Gundel, J. K., Hedberg, N., & Zacharski, R. (1988). On the generation and interpretation of demonstrative expressions. *Proceedings of the Twelfth International Conference on Computational Linguistics, John von Neumann Society for Computing Sciences*, 216–221. <https://doi.org/10.3115/991635.991679>
- Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive Status and the Form of Referring Expressions in Discourse. *Language*, 69(2), 274–307.
- Helbig, G., & Buscha, J. (2001). *Deutsche Grammatik*. Berlin, München: Langenscheidt KG.
- Johnson, E. D., & Arnold, J. E. (2022). The Frequency of Referential Patterns Guides Pronoun Comprehension. *Journal of Experimental Psychology: Learning Memory and Cognition*. <https://doi.org/10.1037/xlm0001137>
- Kaiser, E. (2011). On the Relation between Coherence Relations and Anaphoric Demonstratives in German. *Proceedings of the 2010 Annual Conference of the Gesellschaft Für Semantik, Sinn Und Bedeutung*, 337–351.
- Kaiser, E., Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2009). Structural and semantic constraints on the resolution of pronouns and reflexives. *Cognition*, 112(1), 55–80. <https://doi.org/10.1016/j.cognition.2009.03.010>
- Kaiser, E., & Trueswell, J. C. (2011). Investigating the Interpretation of Pronouns and Demonstratives in Finnish. In E. A. Gibson & N. J. Pearlmuter (Eds.), *The Processing and Acquisition of Reference* (pp. 323–354). <https://doi.org/10.7551/mitpress/9780262015127.003.0013>
- Krasavina, O., & Chiarcos, C. (2007). A corpus-based study of demonstratives in German, Russian and English. *Proceedings of the Corpus Linguistics Conference*, 1–16.
- Loáiciga, S., Bevacqua, L., Rohde, H., & Hardmeier, C. (2018). Event versus entity co-reference: Effects of context and form of referring expression. *Proceedings Ofthe Workshop on Computational Models OfReference, Anaphora and Coreference*, 97–103. <https://doi.org/10.18653/v1/w18-0711>
- McShane, M. (2009). Subject ellipsis in Russian and Polish. *Studia Linguistica*, 63(1), 98–132.

- <https://doi.org/10.1111/j.1467-9582.2008.01155.x>
- Pal, P., Clark, G., & Williams, T. (2021). Givenness Hierarchy Theoretic Referential Choice in Situated Contexts. *Proceedings of the 43rd Annual Meeting of the Cognitive Science Society: Comparative Cognition: Animal Minds, CogSci 2021*, 1816–1822.
- Patil, U., Bosch, P., & Hinterwimmer, S. (2020). Constraints on German diese demonstratives: Language formality and subject-avoidance. *Glossa*, 5(1). <https://doi.org/10.5334/GJGL.962>
- Portele, Y., & Bader, M. (2017). Personal Pronouns and D-Pronouns in German: Connecting Comprehension to Production. *Proceedings of Linguistic Evidence 2016*, 1–16.
- R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Schumacher, P. B., Dangl, M., & Uzun, E. (2016). Thematic role as prominence cue during pronoun resolution in German. *Empirical Perspectives on Anaphora Resolution*, 213–240. <https://doi.org/10.1515/9783110464108-011>
- Tily, H., & Piantadosi, S. (2009). Refer efficiently: Use less informative expressions for more predictable meanings. *Proceedings of the Workshop on the Production of Referring Expressions: Bridging the Gap between Computational and Empirical Approaches to Reference*.
- Trott, S., Bergen, B., & Wittenberg, E. (2022). Spontaneous, controlled acts of reference between friends and strangers. *Language Resources and Evaluation*, 1–25.
- Williams, T., & Scheutz, M. (2017). Reference in Robotics: A Givenness Hierarchy Theoretic Approach. In B. Abbott & J. Gundel (Eds.), *The Oxford Handbook of Reference* (pp. 457–474). Oxford, UK: Oxford University Press.
- Wittenberg, E., Momma, S., & Kaiser, E. (2021). Demonstratives as bundlers of conceptual structure. *Glossa*, 6(1). <https://doi.org/10.5334/GJGL.917>
- Xu, W., & Xiang, M. (2021). Is there a predictability hierarchy in reference resolution? *Proceedings of the Annual Meeting of the Cognitive Science Society*, 276–281.