Self-awareness matters: co-reference in German attitude clauses

Alexander Göbel (UMass Amherst) & Petra Schumacher (University of Cologne)
agoebel@linguist.umass.edu

A sentence like (1) with a third person pronoun that is co-indexed with a higher subject of an attitude verb like *think* (=the attitude holder) can describe one of two scenarios: a) Trump utters "I will become president" (self-aware co-reference, *de se* attitude); b) Trump reads an anonymized tweet that he is impressed by but does not recognize as his own and utters "Whoever wrote this will become president" (accidental co-reference, *de re* attitude).

(1) $Trump_t$ thinks **he**$_t$ will become president.

While personal pronouns in English have been claimed to be compatible with both scenarios, other pronominal expressions have been argued to be sensitive to the distinction between a) and b). For instance, the silent PRO of control infinitives as in (2) requires self-awareness [1].

(2) $Trump_t$ expects **PRO**$_t$ to become president.

However, scenarios like b) are relatively complex and, to our knowledge, have never been tested experimentally. We present three offline studies on German, focusing on the contrast between personal pronouns (PPros) and one type of demonstrative pronoun (DPro). This focus is motivated by claims that DPros in German are sensitive to perspective and avoid co-reference with salient perspective centers [2]. We assume that attitude holders (*Trump* in (1)/(2)) constitute the default perspective center in attitude contexts like (1)/(2) but that this default can be overridden in scenarios like b) for which we assume the speaker to be the perspective center. Thus, we predict DPros to block co-reference in a) but allow it in b).

In Experiment 1, we implemented a 2x3 (CONTEXT vs PRONOUN TYPE) design in a 2AFC. Participants (N=24, Prolific.ac) were given a CONTEXT that either indicated self-aware co-reference or accidental co-reference (Table 1) and had to indicate whether a target sentence was compatible with the context. The PRONOUN TYPES we used in the target sentences were *PPro*, *DPro* and *PRO* (control infinitives) (3). For self-aware co-reference contexts, we expected both *PPros* and *PRO* to be at ceiling with respect to acceptance, while *DPros* were expected to be at floor. For accidental co-reference, we anticipated a general decrease in acceptance rates due to the possibility of the complexity of the contexts affecting comprehension. Abstracting from this potential issue, we expected *PPros* and *DPros* to be overall acceptable and *PRO* unacceptable.

While we found a general decrease for accidental co-reference, there were no differences between PRONOUN TYPES (Figure 1): acceptance rates for target sentences were at ceiling for self-aware co-reference and dropped to chance performance for accidental co-reference independently of pronoun type. The same qualitative pattern was replicated in Experiment 2 with a 7 point Likert-scale instead of binary judgments (Figure 2).

A possible explanation for the lack of differences between pronoun types in Experiments 1 and 2 might be that participants did not process the target sentences deeply enough. To address this worry, Experiments 3a,b presented participants with a target sentence containing a blank in the position of the embedded subject (following the same contexts) and made participants choose between two possible completions (4). The two options were presented below the target sentence on opposite ends of a scale for participants to indicate their relative preference, since PPros were anticipated to be generally preferred over DPros. In addition to the *critical* COMPARISON between PPros and DPros, a *control* COMPARISON of a PPro with an ambiguously false completion was included to measure whether the complexity of contexts would affect ratings independently.

Although Exp. 3a only revealed a significant effect of CONTEXT (Fig 3), the effect of COMPARISON was significant in Exp. 3b (Fig. 4), suggesting that participants are sensitive to the pronoun differences once they are forced to compare them directly. Moreover, both experiments showed a numerical trend towards the predicted interaction, with DPros getting more acceptable in accidental co-reference contexts. Debriefing data suggests that this pattern may depend on whether using DPros is perceived as pejorative, opening an avenue for future research.

Self-aware co-reference	Accidental co-reference
'Mary is rummaging around in a box in the attic where she keeps memorabilia. Therein she finds an old poem that she wrote and likes, and says: "I'm totally a genius!" Later, Mary's housemate Peter, who heard Mary, goes up to the attic himself and is equally impressed by the poem.'	'Mary is rummaging around in a box in the attic where she keeps memorabilia. Therein she finds a poem that she likes, and says: "Whoever wrote this is totally a genius!" Later, Mary's housemate Peter, who heard Mary, goes up to the attic himself and discovers Mary's name on the back of the poem.'

TABLE 1: SAMPLE CONTEXTS

(3) a. **P-Pro**   Peter: "Maria hat behauptet, dass **sie** ein Genie ist."
         *'Peter: "Mary claimed that **she(PPro)** is a genius." '*
    b. **D-Pro**   Peter: "Maria hat behauptet, dass **die** ein Genie ist."
         *'Peter: "Mary claimed that **she(DPro)** is a genius." '*
    c. **PRO**   Peter: "Maria hat behauptet, **PRO** ein Genie zu sein."
         *'Peter: "Mary claimed **PRO** to be a genius." '*

REFERENCES:
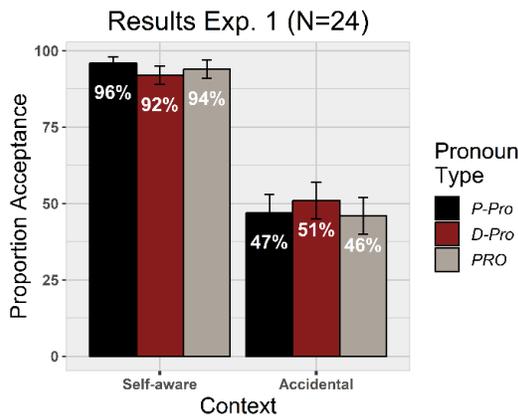[1] CHIERCHIA (1989).
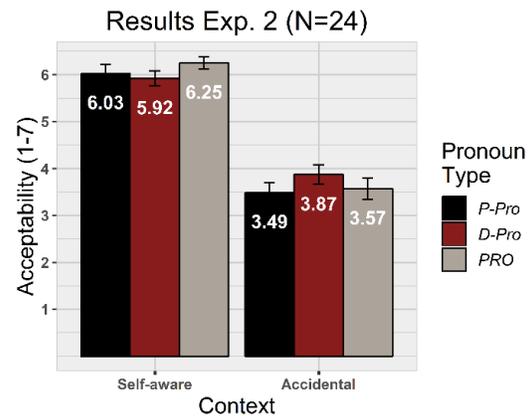[2] HINTERWIMMER & BOSCH (2016)

FIGURE 1: RESULTS EXPERIMENT 1

FIGURE 2: RESULTS EXPERIMENT 2

(4)   *'Peter: "Mary claimed that ＿＿＿ is a genius." '*

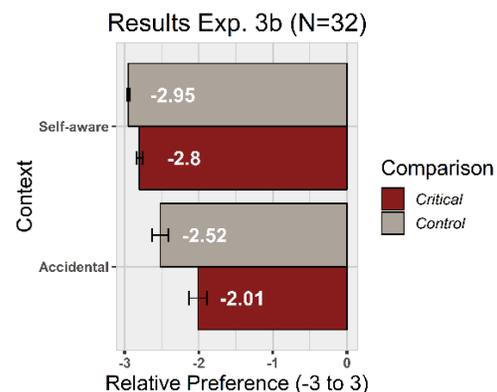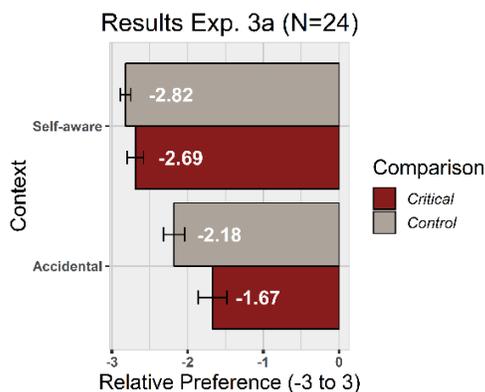| *she(PPro)*	(-3) only left	(-2) strongly left	(-1) slightly left	(0) equal	(1) slightly right	(2) strongly right	(3) only right	*she(DPro) (critical) // Sam (control)*

FIGURE 3 & 4: RESULTS EXP 3A,B (TREND TOWARDS LEFT = PREFERENCE FOR PPRO)