

Simon Dampfhofer*, Edgar Onea* (* University of Graz)

Abstract: New Data on Gender Mismatches and Anaphora

Introduction

If a language allows for distinctions of semantic and grammatical gender, interesting questions arise regarding the possibilities of mismatches between both categories. One particularly interesting and transparent class of examples comprises of what is commonly referred to as *gender shift* cases (see, e.g., Mueller-Reichau 2008), which consist of predicative sentences containing copular constructions with a gender mismatch between two NPs, as illustrated in German example (1) and general structure (2); in a follow-up sentence, one could in principle then use pronouns of either gender when referring to the referent in question.

- (1) Dieser Baum ist eine Ulme. Er/Sie ist groß.
 This tree_M is an elm-tree_F. He_M/She_F is tall.

- (2) NP_{gen1} be NP_{gen2}. PRO_{gen1/gen2} ...

While in some cases, both genders are possible on the pronoun, in others, one gender is dispreferred, sometimes for reasons that are not yet understood completely. In addition, in some situations choosing one gender over the other seems to result in clearly distinct readings, as illustrated by one example in (3):

- (3) a. Dieses Buch ist ein Gesellschaftsroman. Es/?Er hat 800 Seiten.
 This book_N is a society-novel_M. It_N/?He_M has 800 pages.
 b. Dieses Buch ist ein Gesellschaftsroman. ?Es/Er handelt von einer Kaufmannsfamilie im 17. Jahrhundert.
 This book_N is a society-novel_M. ?It_N/He_M deals with a merchant-family in the 17th century.

In this case, a straightforward explanation seems to be that when referring to properties of the first NP, using a pronoun that agrees in gender with that NP is preferred, and analogously for the second NP. However, other examples seem more problematic. For example, in (4), using a pronoun that agrees with *life-saver* is strongly dispreferred and it is hard to come up with continuations where this is more easily possible (analogous to the contrast between (3a) and (3b)):

- (4) Diese Kaffeemaschine ist ein Lebensretter. Sie/??Er leistet uns schon lange Jahre treue Dienste.
 This coffee-machine_F is a life-saver_M. She_F/He_M has been serving us well for several years.

The goal of the present work is to gain a better understanding of factors influencing the availability of either gender in gender shift sentences; this is done by gathering and analyzing empirical data by means of a corpus study. In our contribution, we present the findings of this study and discuss its theoretical implications.

Hypotheses about factors influencing gender shift

A previous exploratory study (in prep) already uncovered some candidates for potential factors influencing the availability of genders in the continuation. In the corpus study, we aim to confirm whether these hypothesized factors actually influence the possibility of gender shift. In addition, we consider some new categories that were not directly mentioned in Onea et al. In particular, the candidate factors we included in our corpus study are: GENERICITY, SEMANTIC TYPE (to allow for a distinction between ABSTRACT and CONCRETE), METAPHORICITY, METONYMY, ANIMACY,

BIOLOGICAL GENDER, WORD LENGTH, DISTANCE between NP and pronoun (with longer distances, pronouns should agree with biological gender more often, as stated e.g., by Köpcke/Panther/Zubin 2010), and NP TYPE. We take some of these factors, such as ANIMACY or WORD LENGTH, to be properties of the NPs contained within sentences, while others, such as METAPHORICITY and METONYMY, concern the relation between the two NPs. Due to the exploratory nature of our study, while we have general ideas about how these factors might influence the availability of gender shift (for instance, based on example (4), one idea is that abstract nouns are less easily available, which would make SEMANTIC TYPE an influencing factor), we didn't formulate precise hypotheses about all of them.

Corpus study

In our corpus study, we look at a total number of 2000 sentences conforming to the general structure in (2) from the deTenTen20 corpus part of the TenTen family of corpora (Jakubíček et al 2013) via SketchEngine, which contains German items from a large variety of web sources; to exclude a large number of false positive hits from our corpus queries, we limit ourselves to sentences containing only masculine and feminine NPs. These sentences are then annotated by a team of three annotators who are tasked with annotating the categories described above for 800 items each; 200 items are shared between all annotators to compute inter-rater agreements (Fleiss' Kappa). Because some categories turned out to be hard to annotate, we iterate several versions of annotation criteria until we achieve an improved version based on which annotators are able to produce results with sufficiently high Fleiss' Kappa.

Theoretical implications and outlook

Depending on our empirical findings, our study could have different theoretical implications. In particular, the goal of the study is to allow us to gain a solid empirical understanding of when and why gender shifts are possible. This will serve as groundwork for later empirical and theoretical work. On one hand, we aim to expand upon this study by gathering data on more targeted and specific phenomena in later empirical studies, including but not limited to an acceptability judgment experiment involving gender shift sentences. On the other hand, our ultimate goal is to develop a syntax-semantics interface for the discourse structural interpretation of gender features. A possible inspiration for such a theory comes from sign languages, where expressions somewhat analogous to pronouns, so-called *referential loci*, are observed to allow for similar shifts as in our cases in a phenomenon called *locative shift* (see Schlenker 2018).

Bibliography

- Jakubíček, Miloš, et al (2013). "The TenTen corpus family". In: 7th International Corpus Linguistics Conference CL, pp. 125–127.
- Köpcke, Klaus-Michael, Klaus-Uwe Panther, and David Zubin. (2010). "Motivating Grammatical and Conceptual Gender Agreement in German". In: Cognitive Foundations of Linguistic Usage Patterns. Ed. by H. Schmid and S. Handl. Berlin: de Gruyter, pp. 171–194. DOI: 10.1515/9783110216035.171.
- Mueller-Reichau, Olav (2008). "Genuswechsel in Kopulasätzen und die Bedeutung indefiniter Nominalphrasen". In: Zeitschrift für Sprachwissenschaft 27.1, pp. 73–98. DOI: 10.1515/ZFSW.2008.004.
- Schlenker, Phillippe (2018). "Locative Shift". In: Glossa: A Journal of General Linguistics 3 (1), p. 115. DOI: 10.5334/gjgl.561.