

Den Klaus kennt jeder.
On the use of the definite article with proper names in German

Edgar Onea, Melanie Loitzl, Lea-Sophie Kravanja and Sophie Pfundner
University of Graz

edgar.onea-gaspar@uni-graz.at, melanie.loitzl@edu-uni-graz.at,
lea.kravanja@edu-uni-graz.at and sophie.pfundner@edu.uni-graz.at

1 Introduction

While proper names (PNs) are usually treated as complete definite descriptions in the literature and are assigned the semantic type *e*, they can be combined with the indefinite article leading to sortal interpretation, (1b), quantification over stages, (1c), or manifestations, (1d), of the individual denoted (von Heusinger 2010). PNs can also be combined with the definite article (DA). While an unmotivated use of the DA is unacceptable in English, (1e), the use of the DA is acceptable if the PN is modified by a relative clause, (1f), or if it can be understood as an individual brand, (1g). Finally, in some languages such as Greek the usage of the DA is arguably obligatory with PNs.

- (1) a. Klaus worked on these data.
b. I met a (certain) Mary. (Longobardi 1994: 636)
a person called Mary
c. Durch die Tür kam ein wütender Paul. (German)
'a furious Paul entered' (von Heusinger & Wespel 2007: 337)
d. We need another Roosevelt. (Payne & Huddleston 2002: 521)
another manifestation of the real Roosevelt / Someone like Roosevelt
e. (?? The) Klaus is semanticist.
f. The/An Anselm I met yesterday was born in Aosta. (Longobardi 1994: 639)
g. The Donald did it my way. (January 22 2017, *The Sunday Times*)

In (1) the usage of the definite or indefinite article is associated with clear semantic effects. However, in German, the usage of the DA with PNs denoting persons does not have any immediate semantic correlates, (2).

- (2) Ich habe (den) Klaus besucht.
'I visited Klaus.'

The literature mentions several potential reasons to use the DA with PNs (Werth 2020: 2–9). Firstly, there seems to be regional variation to the extent that northern German speakers tend to accept the usage of the DA with PNs less and associate it with lack of politeness, whereas speakers of southern German varieties accept the usage more readily, cf. (Bellmann 1990). Social and politeness related effects may further depend on which kind of proper noun (e.g., first names, nick names, full names) occur with the DA. Moreover, some discourse structuring aspects associated with the usage of the DA with PNs in spoken German have been suggested in Golato (2013) and Betz (2015), which remain to be further investigated. Finally, the usage of articles may have a syntactic function in clarifying the syntactic relations in a sentence, since case morphology is virtually non-existent in PNs in current German, which contrast with historical practice of case marking with PNs (see Paul 1917: 153–163 and Nübling 2012: 225–229).

This paper is part of an attempt to elucidate the parameters of variation for the usage of the

DA with person-denoting PNs in written standard German. Here, we report the first results of a large-scale corpus study on this variation.

2 The study

The alleged case-marking function of the DA on PNs appears to be general enough to be detected in written language. Thus, we focus on case marking as a predictor of DA with PN.

Hypothesis If case-marking is a motivation to use the DA with PNs, we expect that the DA will more often be used when the DP is not in nominative. This is because nominative is the unmarked default case in German. Further, we expect a correlation between morphological case-marking and the use of the DA, e.g. genitive is often marked morphologically on the noun (*Peter* vs. *Peters*).

Method We have extracted four data-sets from the German Reference Corpus (DeReKo) based on different queries: 1) MA: contains the most frequent 15 male first names, 2) ARMA: contains a DA directly followed by a male, 3) FE: contains the most frequent 15 female PNs, 4) ARFE: contains a DA followed by a female PN. The first names were chosen from a statistical evaluation of the most commonly used first names in Austria in the past years compiled by Statistik Austria (2020). For each of the data-sets, 200 examples were screened for errors and annotated for gender and case. To name a few criteria, samples were regarded as erroneous when they were used for organisations (*Thomas Hardy-Gesellschaft*), when relevant (linguistic) information was missing, e.g. the case of the DP could not be determined, or if the article before a PN was used as a relative pronoun. Importantly, in MA and FE, the use of a DA with the PN was considered an error, because this way, all data sets became disjunct.

For the evaluation of the results, we were interested in the probability of using the DA given a certain case: $p(\text{article}|\text{case})$. The direct way to calculate this would have been to search for all PNs in a certain case and check the proportion of article usage with them. However, since DA is used very rarely with PNs in written language, this would have been an unrealistic task. Fortunately, using Bayesian reasoning, our data allowed estimating the desired probability. In particular:

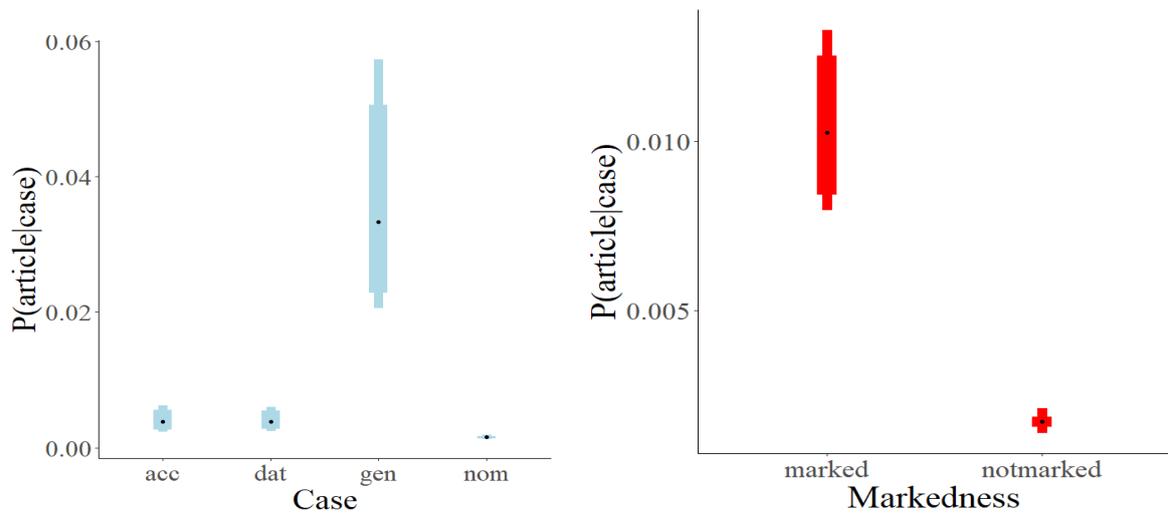
$$p(\text{art}|\text{case}) = \frac{p(\text{art}) \cdot p(\text{case}|\text{art})}{p(\text{case})} = \frac{p(\text{art}) \cdot p(\text{case}|\text{art})}{p(\text{case}|\text{art}) * p(\text{art}) + p(\text{case}|\text{noart}) * p(\text{noart})} \quad (1)$$

From these, $p(\text{case}|\text{art})$ and $p(\text{case}|\text{noart})$ are the proportion of DA/no-DA usage given a certain case, which can be estimated from the annotated data and $p(\text{art})$ is derived from estimating the rate of DA in the sample given the rate of errors in the sample and the total hits.

We estimated the $p(\text{case}|\text{art})$ for each case using a hierarchic Stan simulation using the rstan package (Stan Development Team 2020), with 10 chains of 10.000 iterations and a burn-in of 5.000 and flat priors for all variables; the models achieved good convergence as witnessed by the potential scale reduction statistic (Rhat) of approximately 1 for all cases and the usual visual diagnostics. Thereby, we analytically derived $p(\text{art}|\text{case})$ in each simulation iteration.

Results We show the results in Figure 1 for each case and for a variant of the model in which all marked cases are merged (i.e. feminine dative and genitive and masculine accusative, dative and genitive). The results clearly show that the usage of the DAs in written standard German is well predicted by the case of the DP. At the same time, the probability of using an article with a PN is generally extremely low.

Figure 1: Probability of article usage depending on case



3 Concluding discussion

The data suggest that case-marking is an important factor in determining the usage of the DA with PNs in German. While this has been suggested in the literature before, our method provides a quantitatively more solid estimation of the distributional facts as compared to prior studies such as Schmuck & Szczepaniak (2014) and Werth (2014).

However, our data presented here are insufficient to provide a more detailed explanation of this observation. Are articles actively used to disambiguate case? Is this a reflex of spoken German data or of regional variation? What other factors influence the distribution? For example, we clearly see that for the genitive case, alternation between pre- or postponed possessor (relative to the head noun) is a relevant factor. However, we still lack a clear enough understanding of the semantic and pragmatic effects of such alternations. We hope that by exploring further details of our annotated data we can contribute to explaining the factors governing the distribution of the DA with PNs in German.

References

- Bellmann, Günter. 1990. *Pronomen und Korrektur. Zur Pragmalinguistik der persönlichen Referenzformen*. Berlin: De Gruyter.
- Betz, Emma. 2015. Recipient design in reference choice. Negotiating knowledge, access, and sequential trajectories. *Gesprächsforschung* 16. 137–173.
- DeReKo – *Deutsches Referenzkorpus*, <https://cosmas2.ids-mannheim.de/cosmas2-web/> (2nd February 2022).
- Golato, Andrea. 2013. Reparaturen von Personenreferenzen. *Deutsche Sprache* 41. 31–51.
- von Heusinger, Klaus. 2010. Zur Grammatik indefiniter Eigennamen. *Zeitschrift für germanistische Linguistik* 38(1). 88–120.
- von Heusinger, Klaus & Johannes Wespel. 2007. Indefinite proper names and quantification over manifestations. *Proceedings of Sinn und Bedeutung* 11. 332–345.
- Longobardi, Giuseppe. 1994. Reference and proper names: A theory of N-movement in syntax and Logical Form. *Linguistic Inquiry* 25(4). 609–665.
- Nübling, Damaris. 2012. Auf dem Weg zu Nicht-Flektierbaren: Die Deflexion der deutschen Eigennamen diachron und synchron. In Björn Rothstein (Hrsg.), *Nicht-flektierende Wortarten*, 224–246. Berlin: De Gruyter.

- Paul, Hermann. 1917. *Deutsche Grammatik*, vol. 2. Halle a. S.: Niemeyer.
- Payne, John & Rodney Huddleston. 2002. Nouns and noun phrases. In Rodney Huddleston & Geoffrey K. Pullum (Hrsg.), *The Cambridge Grammar of the English Language*, 323–524. Cambridge: Cambridge University Press.
- Schmuck, Mirjam & Renata Szczepaniak. 2014. Der Gebrauch des Definitartikels vor Familien- und Rufnamen im Frühneuhochdeutschen aus grammatikalisierungstheoretischer Perspektive. In Friedhelm Debus, Rita Heuser & Damaris Nübling (Hrsg.), *Linguistik der Familiennamen*, 97–137. Hildesheim: Olms.
- Statistik, Austria. 2020. *Statistik der natürlichen Bevölkerungsbewegung. Statistik der 60 häufigsten Vornamen 1984–2019 in Österreich – etymologisch zusammengefasst*. <https://www.statistik.at/> (2nd February 2022).
- Werth, Alexander. 2014. Die Funktion des Artikels bei Personennamen im norddeutschen Sprachraum. In Friedhelm Debus, Rita Heuser & Damaris Nübling (Hrsg.), *Linguistik der Familiennamen*, 139–174. Hildesheim: Olms.
- Werth, Alexander. 2020. *Morphosyntax und Pragmatik in Konkurrenz. Der Definitartikel bei Personennamen in den regionalen und historischen Varietäten des Deutschen*. Berlin: De Gruyter.