On "Article Omission" in German and the "Uniform Information Density Hypothesis"

Eva Horch
Universität des Saarlandes
e.horch@mx.uni-saarland.de

Ingo Reich
Universität des Saarlandes
i.reich@mx.uni-saarland.de

Abstract

This paper investigates whether Information Theory (IT) in the tradition of Shannon (1948) and in particular the "Uniform Information Density Hypothesis" (UID, see Jäger 2010) might contribute to our understanding of a phenomenon called "article omission" (AO) in the literature. To this effect, we trained language models on a corpus of 17 different text types (from prototypically written text types like legal texts to prototypically spoken text types like dialogue) with about 2.000 sentences each and compared the density profiles of minimal pairs. Our results suggest, firstly, that an overtly realized article significantly reduces the surprisal on the following head noun (as was to be expected). It also shows, however, that omitting the article results in a non-uniform distribution (thus contradicting the UID). Since empirically AO seems not to depend on specific lexical items, we also trained our language models on a more abstract level (part of speech). With respect to this level of analysis we were able to show that, again, an overtly realized article significantly reduces the surprisal on the following head noun, but at the same time AO results in a *more* uniform distribution of information. In the case of AO the UID thus seems to operate on the level of POS rather than on the lexical level.

1 Introduction

It is well-known (see e.g. Sandig 1971; Stowell 1991; Reich, to appear) that headlines (and some related text types) in principle allow for articleless singular noun phrases (1a) which are strictly ungrammatical in other contexts (1b):

(1) a. Größte Dürre seit einem halben
Biggest aridity since a half
Jahrhundert
century
"Biggest aridity since half a century"
(zeit.de: 10.08.2015)

b. *Er dachte an größte Dürre seit
He thought of biggest aridity since
einem halben Jahrhundert
a half century
"He thought of biggest aridity since half a century"

This phenomenon is called article omission (AO) in the literature (even though it is not clear that there is in fact some kind of ellipsis involved). What we do *not* want to claim in this paper is that information theory (IT) can explain why AO is grammatical in some text types, but not in others. However, in text types which do allow for AO, AO is clearly optional. In other words, in production the speaker / writer needs to make a choice. The crucial question that we want to investigate in this paper thus is whether this choice in production is guided by information theoretic principles like the Uniform Information Density Hypothesis (UID).

2 Background and Aim

In a paper on complementizer deletion, Jaeger (2010) showed that the overt realization of a complementizer like "that" can significantly reduce the information carried by the (following) subject, thus contributing to a more uniform distribution of the information at the left periphery in the case of high surprisal subjects. According to Jaeger (2010) this effect guides the speaker when choosing between two grammatical alternatives. The underlying principle he states as follows:

(2) **Uniform Information Density (UID)**
Within the bounds defined by grammar, speakers prefer utterances that distribute information uniformly across the signal (in-

> formation density). Where speakers have a choice between several variants to encode their message, they prefer the variant with more uniform information density (ceteris paribus). (Jaeger 2010: 24)

The parallels to AO are rather straightforward: In both cases there are two grammatical alternatives which convey essentially the same proposition. In both cases a functional expression precedes a noun (phrase). In both cases the speaker / writer has to opt for one of the alternatives during the production process. Now, building on Jaeger's (2010) results one might suppose, firstly, that functional expressions in general lower the surprisal of the lexical items to follow, and, secondly, that the realization of the functional expression depends (at least to some degree) on whether its realization results in a more uniform (local) density profile.[1]

3 Language Modeling

To test this hypothesis with respect to AO in German we trained trigram language models with the SRI Language Modeling Toolkit[2] (Stolcke 2002) on a corpus consisting of 17 different text types with about 2.000 sentences each and compared the density profiles of minimal pairs like *Kampf der Zeiten* ("Battle of times") vs. *Der Kampf der Zeiten* ("The battle of times"), see figure 1.

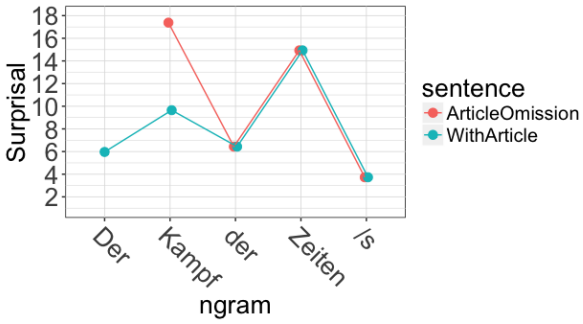

Figure 1: Surprisal profiles (lexical)

Our results show that an article, be it definite or indefinite, does in fact lower the surprisal of the following head noun. This generalizes to sentence-internal positions and to other lexical items of the same syntactic category. In the example chosen, however, the results seem to contradict the UID hypothesis: The original corpus version (*Kampf der Zeiten*) with AO shows a (locally) less uniform profile than the constructed example which overtly realizes the article preceding the head noun.

To get a clearer picture, we abstracted away from the concrete lexical items and trained our language models exclusively on POS structures.[3] The results (trigrams) are shown in figure 2.

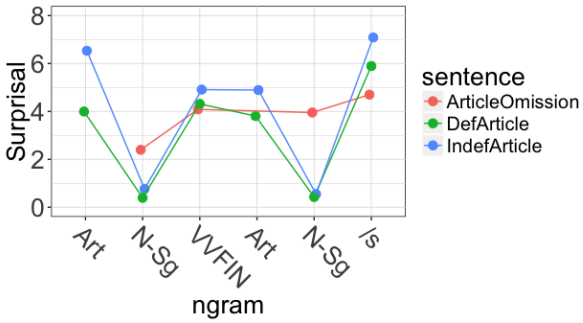

Figure 2: Surprisal profiles (POS)

On this more abstract level, an overtly realized article, whether definite or indefinite, also lowers the surprisal of the following head noun. In contrast to the lexical level, however, an overtly realized article correlates with high surprisal (whether definite or indefinite, whether in sentence-initial or sentence-internal position). As a consequence, the overt realization of an article preceding a head noun results in a peak followed by a trough. Dropping the article, on the other hand, results in a (more) uniform distribution of the information. These results have been confirmed by a 'hybrid' model that combines POS information with information about case, gender and prepositions, see figure 3.

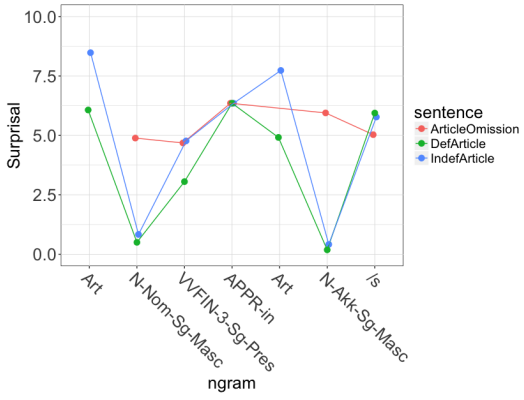

Figure 3: Surprisal profiles (hybrid)

[1] See also De Lange (2008) for a (contrastive) analysis of AO whithin the framework of Information Theory (exclusively based on the number of possible articles in a language).

[2] See `http://www.speech.sri.com/projects/srilm/`. Since smoothing techniques showed no significant effects, we refer to unsmoothed data in this paper.

[3] We used "TreeTagger" by H. Schmid (U Stuttgart) and an expansion of the STTS tagset (see Schmid 1995, 1994).

4 Interpretation

The data suggests two conclusions: Firstly, it seems in fact possible to generalize the observation that functional elements like complementizers and articles systematically lower the surprisal of the lexical item to follow. Secondly, with respect to article omission (and in contrast to complementizer deletion) the UID seems to operate on a more abstract level (POS structure) than on the level of concrete lexical items. This is an important insight into the way article omission (in German) works, and it shows that information theory can in fact contribute to an understanding of that phenomenon.

5 Further Predictions

Given that the two interpretations stated above are essentially on the right track, information theory makes another testable prediction: We expect that if articles are omitted in a sentence, they are in fact omitted across the board. (This is simply because on the level of POS – which has been argued above to be the relevant level for AO – the different surprisal values of different lexical items do not play any role anymore with respect to considerations of uniform information density.) Our corpus suggests that this prediction is in fact correct: The corpus contains a total of 2.127 headlines out of which 308 headlines are in fact subject to AO. Out of those 308 headlines only 137 contain more than one possible target for AO. Out of those 137 headlines, finally, 125 headlines show AO across the board, see (3) (source: SZ.de, 07.06.12) and (4) (source: Bild.de, 04.06.12) for illustration. This is about 91% of the relevant cases, see also figure 4.

(3) Δ *Betrunkene Großmutter schlägt* Δ *Passagier nieder* ('drunken grandma knocks down passenger')

(4) Δ *Fahrer rettet* Δ *Fahrgast aus* Δ *brennendem Bus* ('driver rescues passenger out of burning bus')

As for the remaining 9% it is remarkable that none of them shows the pattern 'overt article followed by null article'. In all of the relevant cases overt articles follow AO. This is in accordance with an observation in Stowell (1991), dubbed "Stowell's Law" in Reich (to appear): In headlines, overt articles must not c-command omitted articles.

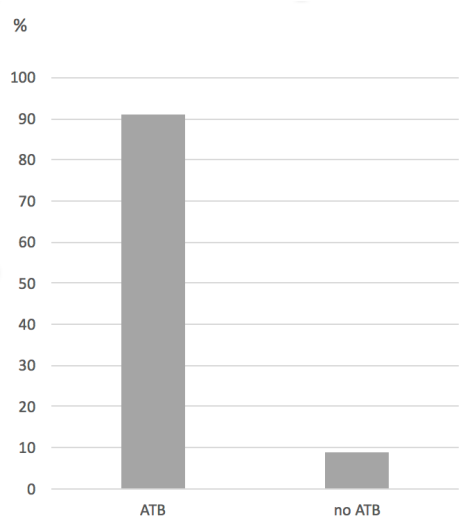

Figure 4: Multiple targets for AO

References

T. Florian Jaeger. 2010. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61:23–62.

J. N. De Lange. 2008. *Article omission in child speech and headlines: a processing account*. Ph.D. thesis, Utrecht University, Utrecht.

Ingo Reich. to appear. On the omission of articles and copulae in German newspaper headlines. In D. Massam and T. Stowell, editors, *Register Variation and Syntactic Theory*. Special issue of *Linguistic Variation*.

Barbara Sandig. 1971. Syntaktische Typologie der Schlagzeile. In *Linguistische Reihe*, volume 6. Hueber Verlag, Ismaning.

Helmut Schmid. 1995. Improvements in part-of-speech tagging with an application to German. In *Proceedings of the ACL SIGDAT-Workshop*, Dublin, Ireland.

Claude Shannon. 1948. A mathematical theory of communications. *Bell Systems Technical Journal*, 27(4):623–656.

Andreas Stolcke. 2002. SRILM - an extensible language modeling toolkit. In *Proc. Intl. Conf. Spoken Language Processing*, Denver, Colorado.

Timothy Stowell. 1991. Empty heads in abbreviated English. In *Proceedings of GLOW 1991*.