Argument omissions in multiple German corpora

Josef Ruppenhofer

1 Introduction

In the past several decades, with the publication of results by Lehrer (1970), Mittwoch (1982), Fellbaum & Kegl (1989), our understanding of the contextual factors that license null complements and constrain their interpretations--from constructions to genres to lexical classes--has vastly increased. As Fillmore (1986) made clear, null complementation, while motivated by such pragmatic drives as effort conservation, is linguistically constrained. As examples 1 and 2 show, semantically similar predicates differ in their ability to omit the presumably identical semantic role.[1,2]

1   @jeb140   ah,   ich   {verstehe}   ∅   .   gute   arbeit   !
    'ah, I {understand} ∅. good work! '

2   *   @jeb140   ah,   ich   {realisiere}   ∅.   gute   arbeit   !
    'ah, I {realize} ∅. good work ! '

At the same time, the phenomenon is more systematic than an approach based on lexical idiosyncrasy might suggest. It is well known that certain constructions such as the passive or the imperative allow the omission of a verb's semantic 'deep subject' role (cf. (3)-(4)). But further constructions and contexts exist that can license omissions. Goldberg (2006) argues that the discourse prominence of participants explains why constructions like the English experiential perfect license argument omissions such as (5) that may not occur in episodic contexts. And Ruppenhofer and Michaelis (2010) illustrated the effect of genre on omissibility, as in the case of product label statements such as (6).

3   Eine Wasserader war zwar bereits {gefunden} ∅ … (Passive)
    'Indeed, an underground water course has already been {found} ∅ …'

4   ∅ {Zeig} mir Deine Chromosomen, und ich sag Dir wer Du bist.
    (Imperative)
    '∅ {Show} me your chromosomes, and I'll tell you who you are.'

5   Dieser Mann ist eine Gefahr! Er hat schon eimmal ∅ {getötet} ...

---

1   For the convenience of the reader, I insert the symbol ∅ in order to indicate 'missing' elements. The symbol is not to be taken as an empty element in the grammar.

2   A reviewer questioned whether *realisieren* had a sufficiently similar meaning to *verstehen*. The *duden.de* web-site lists a sense:
  (in einem Prozess der Bewusstmachung) erkennen, einsehen, begreifen
  'recognize, comprehend, understand something (in a process of becoming conscious of it)' (access: May 3, 2013).

http://www.blick.ch/news/ausland/auch-wir-wollen-die-wahrheit-wissen-id144800.html (Experiential perfect)

'This man is dangerous! He has {killed} ∅ once before.'

6   ∅ {Unterstützt} die Hautregeneration nach der Rasur. (Labelese; found on deodorant)

'∅ {Supports} the skin's regeneration after shaving'

At the same time certain conundrums remain. One is, if generalizations are possible for lexically licensed omissions, will they hold only for specific languages such as English or do they, at least to some extent, apply across languages? Another issue is how well any generalizations hold up in the face of data from attested language use given that much research into argument omission was based on introspection or constructed data. Related to the previous point, most work on argument omissions has, implicitly, if not explicitly, focused on some core types of written language, in particular texts found in newspapers. One may wonder if other types of written language, especially ones that represent other communicative settings, offer omission affordances that are not (commonly) found in the written texts that are most often investigated, or if they at least use these affordances in distinct ways.

The remainder of this paper is structured as follows. Section 2 provides a typology of argument omission. Section 3 describes the data sources which are used in this paper. Section 4 examines whether a generalization that Ruppenhofer (2004) proposed for English regarding the interpretation of omissions is plausible for German, too. In Section 5 we compare German data from social media and from spoken and written language to see how they differ in the omission types they admit, or at least in the frequency with which a given affordance is exploited. To this end, we will study the distributions of verb-initial constructions in the data as several of these involve argument omission. In section 6 , we focus on the analysis of various kinds of topic-drop sentences that can be observed in the social media data. In section 7 we provide discussion of our results and offer some conclusions.

2 Overview of null instantiation

Fillmore (1986) distinguished lexically licensed omissions from constructionally licensed ones. With the former class, the argument-omission affordance is licensed by a particular lexical item, and nearly synonymous items may differ in omissibility of a given semantic role. The difference between *verstehen* and *realisieren* in examples (1) and (2) above illustrates lexical licensing.

With constructional omissions, it is the particular construction that determines the omissibility of a given argument, in a given syntactic role. For instance, *von*-phrase agents in passive predications can be omitted regardless of the lexical identity of the passive-form verb (cf. (3)). Likewise, all imperatives can omit their subjects, as in (4).

There also exist constructions that not only allow but actually require the omission of a verbal semantic role in the process of reconciling constructional and verbal requirements (Michaelis 2011). For instance, some types of resultative constructions require omission of a theme argument under an existential interpretation. This is illustrated in examples (7) and (8), where no drink is specified but instead the resultative construction's own patient argument is provided.

7 Die verschwundene Mutter war lustig gewesen und feierte gern ; {trank} sich tot in zweiter Ehe …
   'The mother who disappeared was funloving and liked to party; {drank} herself to death in her second marriage …'

8 Aber Frau Professor konnte mich unter den Tisch {trinken} …(DEWAC)
   'But Mrs Professor could {drink} me under the table …'

Note that the omission-licensing constructions above all are general-language constructions. In addition to these, German also allows for genre-dependent omissions of the types discussed for English by Ruppenhofer and Michaelis (2010). Sentence (9) illustrates the kind of object drop found in recipes; sentence (6) above is an instance of labelese.

9 ∅ 1-2 Tage darin {marinieren} . (Recipe object drop)
   '{Marinate} ∅ in it for 1-2 days'

While we have so far considered cases where an argument can be optionally omitted, it is worth pointing out that there are lexical items that always suppress the expression of core participants of the event they encode. An example is the German verb *zubeissen* 'bite' in (10):

10 Der Hund hat (*mich/in/auf/gegen/mit/zu/für mich) plötzlich {zugebissen} (constructed)
    'The dog suddenly {bit} [me].'

A second major aspect of argument omissions is the interpretation that the omitted argument needs to receive. In some cases, unexpressed arguments are merely existentially bound, in others specific antecedents must be resolvable from the linguistic context (or co-text) or the speech setting. The parameter interpretation type is orthogonal to the licensing parameter. The passive construction allows for the omission of an existentially bound argument, as seen in (3), but so do lexical items like *stricken* in (11). Similarly, the imperative licenses a definite omission of the addressee, shown in (4), but so does the lexical item *verstehen* in (1).

11 Es ist Juli, ich sollte lernen, aber ich {stricke} lieber ∅. #Omatweet (Twitter)

    'It's July, I should be studying but I prefer {knitting} ∅.'

It is worth pointing out the differences between the typology introduced in the foregoing and the typology used, for instance, by Zifonun et al. (1997). Their category of *ellipsis* subsumes much, if not all, of what is discussed here and organizes it differently. Zifonun's (1997: 413-442) subtypes of ellipsis include the following:

- situational ellipsis: omitted referents are recoverable from the speech context (speaker, hearer, objects of joint attention, situations jointly observed)

- empratical ellipsis: predicates (usually) are missing that can be inferred based on the joint activity that speaker and hearer are engaged in

- phatic ellipsis: the speaker abandons their production, leaving it to the hearer to fill in the un-produced material

- structural ellipsis: omissions and ellipses licensed in specific text-types due to considerations of economy and condensation

Thus, unlike our notion of null instantiation, ellipsis encompasses cases of unexpressed predicates such as the example in (12):

12   Hierher.
       '(Over) here.' (instruction to furniture movers) (=Zifonun et al. 1997: 420)

In this study, we will only be concerned with argument omissions. We will not be specifically discussing whether they are, for instance, of the situational or the structural type in the sense of Zifonun et al. (1997), although this is clearly relevant for detailed constructional analysis (see Ruppenhofer and Michaelis 2010).

3 Data

We use several sources of data, as shown in Table 1. First, we use the Huge German Corpus as our default corpus (HGC; Fitschen (2004)). German examples in the text that bear no other identification of a source are taken from the HGC, which contains text from several German newspapers. We als use the DEWAC German web-corpus (Baroni et al. 2009), which represents a wider range of written language. Our third data set is a corpus of messages from the Twitter microblogging service, which exhibits some features of spoken language, despite being medially written.[3] As discussed, for instance, by Richling (2008), social media data combines aspects of spoken and written language and exhibits a good degree of what Koch & Oesterreicher (1985) call *conceptual* orality, even if it is *medially* written. The fourth data set we work with is a corpus of parliamentary speeches from the German Bundestag.[4] While the Bundestag data represents the spoken medium, it is conceptually more on the written end of the continuum between written and spoken language. Our fifth and final data set are 80 transcripts from the CallHome German corpus, which contains telephone conversations between German students on exchange in the US and their family members or friends in Germany.[5]

A quantitative description of the corpora is given in Tab. 1. It should be noted that the units are not fully comparable across corpora. For Twitter, the unit is the individual tweet (tw) of up to 140-characters. For the CallHome corpus (Karins et al. 1997), we give two types of units, sentences (s) and turns (tu). For all other corpora, the unit is the sentence.

Corpus	Tokens	Units
HGC	204.813.118	12.223.281 s
DEWAC	1.627.169.557	92.395.25 s
Twitter	105.074.399	7.311.960 tw
Bundestag	5.756.188	278.160 s
CallHome	202.964	23.791 s / 19319 tu

Tabelle 1: Corpora used

---

3   The Twitter dataset cannot be redistributed at this time due to legal uncertainty surrounding the copyright of social media data.

4   This data set was produced at the University of Stuttgart by Stefan Evert. An online version is available at http://linglit193.linglit.tu-darmstadt.de/CQP/Bundestag/frames-cqp.html. (last access: May 28, 2013)

5   The corpus has 100 transcripts but because of the effort required to pre-process the data to make it queryable, we used only 80 of them.

4 Testing the generalization regarding interpretation type

In this section we test Ruppenhofer (2004)'s prediction that predicates in the same lexical class, defined in terms of FrameNet's frames, will omit a particular semantic role with the same interpretation type, if they can lexically license its omission at all.[6] Ruppenhofer's prediction was tested on English but, given that frame semantics / FrameNet aims to provide a more or less language-independent analysis of lexical meanings, it was also meant to apply to other languages.

FrameNet seeks to implement in lexicographical practice the concepts of Frame semantics (Fillmore 1982, 1985). The basic idea is that many words are best understood as part of a group of terms that are connected to a particular type of situation and the participants and "props" involved in it. The classes of events are the semantic frames. Lexical units (LUs) are said to evoke the frames. The roles associated with an event are referred to as frame elements (FEs). This system of analysis applies not only to events but also to relations and states; the frame-evoking expressions may be single words or multi-word expressions, which may belong to any syntactic category. For each LU, example sentences are extracted from corpus data and annotated with the frame elements, and their phrase types (PTs) and grammatical functions (GFs) vis-à-vis the target word. The goal of annotation is to exemplify every attested combination of FE, GF, and PT.

We consider data from two sources. The first source is the second release of the German Salsa project (Rehbein et al. (2012)), which applied frame-semantic annotation to all instances of selected lemmas in a German newspaper corpus, re-using the English FrameNet's frames to the extent possible. Since the lemma coverage of Salsa is not as high as FrameNet's for English, we also took random samples for additional predicates from the HGC. For these samples, we manually sense-disambiguated 200 lemma instances of our target predicates and recorded for each instance exhibiting the frame of interest how the semantic role of interest was realized. If the role was overtly expressed, we recorded whether the semantic role was morpho-syntactically definite or indefinite. If it was unrealized, we recorded its interpretation type, anaphoric or existential.[7]

Table 2 shows the results of this preliminary study.[8] The predicates marked with an asterisk are ones for which Salsa data was used. In the table, the row total for the predicates for which we took random samples is usually less than 200, the size of our random samples. The main reasons are polysemy (e.g. *ankommen* also has senses 'depend (on)' and 'go down well (with)'); misspellings (e.g. *abreisen* 'depart' instead of *abreissen* 'tear off'); and cases where a potential instance of a particle verb actually consists of the simple verb *reisen* 'travel' occurring with an unconnected instance of

---

6   The prediction is not that all predicates in a particular lexical class will be able to omit a given semantic role. Clearly that would be untenable as, for instance, *ankommen* 'arrive' allows for an unrealized Frame Element Goal but *erreichen* 'reach' does not.

7   For the purposes of the present study, as for Ruppenhofer (2004), the morpho-syntactic form of overt mentions serves as a proxy for their discourse status in accord with observations by Ariel (1988); Fraurud (1996); Gundel et al. (1993) *inter alia* that the two are strongly correlated.

8   The Frame *Create textile* is not an existing FrameNet frame; I posit it here for convenience.

the particle or preposition *ab* 'off'. (So as not to exclude cases where the particle occurred separate from the verb stem, the samples for particle verbs included not only cases where the two occurred together but also all cases where a sentence contained the simple verb and an instance of the particle anywhere in the sentence.)

Verb	Frame	FE	Def	Indef	Zero	NI-type	N
ankommen*	Arriving	Goal	5	2	5	DNI	12
eintreffen*	Arriving	Goal	14	0	9	DNI	23
zurückkehren*	Arriving	Goal	36	5	16	DNI	57
abreisen	Departing	Source	25	2	137	DNI	164
abfahren	Departing	Source	26	0	22	DNI	48
zustimmen*	Agreeing	Content	43	23	5	DNI	74
einwilligen	Agreeing	Content	67	30	53	DNI	150
Bruder*	Kinship	Ego	29	0	8	DNI	44
Mutter*	Kinship	Ego	17	6	17	DNI	44
Schwester*	Kinship	Ego	5	0	2	DNI	19
Vater*	Kinship	Ego	27	4	23	DNI	56
füllen	Filling	Theme	10	60	63	INI	133
schmücken	Filling	Theme	55	62	18	INI	135
sticken	Create textile	Product	9	15	49	INI	170
stricken	Create textile	Product	4	22	36	INI	62
häkeln	Create textile	Product	2	23	23	INI	55
backen	Cooking creation	Produced food	18	54	20	INI	92
kochen	Cooking creation	Produced food	10	24	76	INI	110

Tabelle 2: Morphosyntax of realizations of selected FEs in corpus data

As can be seen, the regularity observed for English also seems to hold for German predicates belonging to the same lexical class. For instance, *abreisen*, *abfliegen*, *abfahren* all require anaphoric interpretation of uninstantiated Sources, while *backen* and *kochen* require existential interpretation of omitted Produced food FEs.

The data in Table 2 also are compatible with Ruppenhofer (2004)'s observation that the interpretation type for omitted instances seems to correlate with the information status of overt instances. That is, if an omissible FE's overt instances tend to be given, as indicated by morpho-syntactic definiteness, its interpretation under omission will be anaphoric, if they tend to be new, as indicated by morpho-syntactic indefiniteness, its interpretation under omission will be existential. This correlation holds for all the predicates in Table 2.

Both of these generalizations, at least in their strongest form, are open to falsification by inspection of additional data. Since there exist, however, a great many predicates with some kind of argument omission affordance, it is not possible to exhaustively test the predictions across all of the German vocabulary in the context of the current study.

5 Constructional omission affordances: verb-initial utterances in German

In this section, we turn to the question whether social media data from Twitter differs from the classical written data found in newspaper corpora in the omission types that are found or the frequency with which different kinds of omissions occur. Since there are quite a few omission-licensing constructions that could be investigated, we focus on constructions that are verb-initial in order to keep the effort needed to identify the constructions under investigation manageable.

Word order in standard German is verb-second in finite main clauses and verb-final in finite subordinate clauses. Additionally, some verb-initial constructions exist. As discussed by Auer (1993), one can subdivide these constructions according to whether the predicates realize all their core arguments or not. We begin by considering the first group, which Auer, following tradition, calls cases of proper V1 (*eigentliche Verbspitzenstellung*). The constructions of this type include some types of exclamatives (13); *yes/no*-questions (14); presentational constructions with a finite (typically, third-person) verb form such as (15), which are often found at the beginning of jokes; *wenn* (*if*)-less protases in conditional sentences (16); and what one may call contrast-inversions such as (17).

13 Ist das schön.
’Is that nice! ’

14 Steht jetzt eine neue Blamage ins Haus ?
’Is a new disgrace imminent? ’

15 Kommt ein Mann zum Psychiater : " Herr Doktor, ich habe ein Problem, alle übersehen mich . " - " Der Nächste bitte . "
’A man comes to see a psychiatrist: "Doctor, I have a problem, everybody overlooks me." - "Next, please".

16 Kommt ein Beschluß durch, der die Verpflichtung zur Offenbarung streicht, stehe ich nicht zur Verfügung .
’If a resolution passes that eliminates the need for disclosure, then I won't serve/stand.’

17 Es ist mit Sicherheit eine der heikelsten Sequenzen in Sally Potters Film . Kann doch Orlandos Hingabe mißdeutet werden als Zugeständnis an die einzig mögliche Utopie, der von Mann und Frau .
’It is certainly one of the most delicate sequences in Sally Potter's movie [given that/since] Orlando's dedication can be misinterpreted as a concession to the only possible utopia, that of man and woman. ’

A verb-initial construction type that is uncommon in standard written German and which Auer does not discuss, but which can be found quite frequently on Twitter, are apodoses/consequent statements of conditional constructions such as (18)-(19), where the conditional protasis (*wenn*-clause) is unexpressed.

18 Gehe ich halt ohne Hose los ! (Twitter)
’So then I'll just go off without pants’

19 Dreh ich halt die #Musik auf . #1live #gotye (Twitter)
’So then I'll just turn up the music.’

A further V1-construction-type is exemplified by cases where an expletive *da* seems to be omitted, as in (20).

20 Und da sage ich so, [blowing through lips] Peter saß mir schräg gegenüber, sage ich so, du da Peter, da kommt raus aus dem Spüle .{Guckt} der ganz irre ja, dem ging es heute den ganzen Nachmittag schon nicht so besonders, ne -- guckt ganz irre, und dann gucke ich wieder hin und dann sehe ich, wie sich ganz langsam und stetig das Waschbecken mit aeh Wasser füllt.

'And then I say, [blowing through lips], Peter was sitting kitty corner from may, and I say, hey you, Peter, it's leaking out of the faucet. He looks real crazy, he wasn't doing particularly well all afternoon today, right—he looks real crazy, and then I look at it again and then I see, how the washbasin fills up with water slowly and continually. (CallHome)

In the above verb-initial clause types, the utterance-initial predicates can have their full set of core semantic roles realized. In other constructions, the ones that Auer (1993) calls cases of improper or pseudo V1 (*uneigentliche Verbspitzenstellung*), one of the arguments is null-instantiated. Of these constructions, the most frequent type may be the imperative (4). Some other subject-drop constructions with a finite verb are found in written language, too. Among them are statements on product labels (6); representations of interior monologue (21) or of speech (22), and diary style (23).

21 Halbfett . Hinter dem Autorinnennamen . Einfach so, der Name und ein Punkt . ∅ {Kommt} gut. Wirkt sicher . Hat was von Definitivität . 'Semi-bold. Behind the author name. Just like that, a name and a period. {Comes across} nice. Looks assured. Looks definitive.'

22 Statt dessen schlägt er den Abschuß eines Balles vom Kopf vor . Ist Tell, sagt er . ∅ {Kommt} gut, sagt er . Hammers ? fragt Dahlmann . Nicht ganz, sagt Kerner .
'Instead he proposes kicking a ball off somebody's head. Is Tell, he says. {Comes across} nice, he says. We ready, asks Dahlmann? Not quite, says Kerner.'

23 ∅ {Bin} immer noch beim Sichten, es werden immer weniger verschiedene Stapel .
'{Am} still sifting, the piles are getting fewer.'

In addition to the preceding subject-drop constructions, German also allows the null-instantiation of objects and clausal complements in verb-initial constructions, typically in the spoken variety. In (24), the bare infinitive complement of *wollen* is omitted, while in (25) an embedded question complement of *wissen* is omitted. (Alternatively, a simple anaphoric pronoun such as *das* may be taken to be omitted in 24 and 25.)

24 So viel Fußball wie in den letzten 5 Wochen hab ich noch nie verpasst . ∅ {Will} ich auch nie wieder . Aber . 3 Wochen Berlin sinds noch . (Twitter) 'Never missed out on as much soccer as in the last 5 weeks. I don't want ∅ ever again. But there's still 3 weeks in Berlin.'

25 Was ziehst du heute an ? — Oehm ∅ {weiß} ich nicht ! :D http://t.co/lTkMgHUP (Twitter)

'What are you putting on today? – Ahem, I don't know ∅! :D'

Actually, if one assumes that an anaphoric *dann* is omitted from examples (18) and (19), then those examples could be said to exhibit topic-drop, too.

Auer (1993:194) also lists other types of pseudo-V1. Among these are, for instance, parenthetical reporting clauses (26), and main clauses of sentences beginning with a subordinate clause.

26 Freilich, {sagte} Eduard, hilft das Hin- und Widerdenken, das Hin- und Widerreden zu nichts. (=Auer 1993:194, example (vi))
'Of course, said Eduad, the thinking to and fro, the talking to and fro doesn't help at all.'

27 Ja, wie er sie auf demPapier sah, {fing} er bitterlich an zu weinen.
'Now,when he saw them on paper, he began to shed bitter tears.'

We will leave these additional types aside for our purposes because they do not have the verb in sentence-initial position and extracting them automatically is not trivial. Generally, for our purposes we use the label V1 in a very surface-oriented way: we will speak of all of the above types of omissions where the finite verb seems to be appear in the V1-position as V1-constructions, even though the topic-drop constructions in particular are treated as involving V2 in syntactic theories that assume empty elements (cf. Schalowski (2009)).

5.1. Frequency of verb-initial utterances

Since the automatic tagger[9] that we use is likely to make a considerable number of mistakes on the Twitter data due to, for instance, unusual punctuation as in (28) and since tweets contain elements that are not directly part of utterances, we hand-checked a set of 200 tweets selected at random from the full corpus. In looking for verb-initial utterances, we also accepted cases where the first token in the post was a form of address marked by "@" (cf. (29)) or a topic marking hashtag (28). In the end, we found 16 tweets with an initial verbal form. 6 of them were *yes/no*-questions, the remaining 10 were cases of subject pro-drop such as (29).

28 158854: {#offline} :) {Schlaft} gut c :
'Sleep well'

29 @nwpxOdi ∅ {Klingt} gut . Wenns Wetter passt bin ich dabei .
'∅ {Sounds} good. If the weather's fine, I'll come along.'

The share of verb-initial sentences in our Twitter data thus is about 8%. To see how this compares with the rate of verb-initial sentences in other corpora, we repeated the exercise by hand-checking samples of 200 randomly sampled sentences.[10] The results

---

9 It is a version of the TreeTagger (Schmid 1994) with a slightly adapted lexicon, kindly provided to us by our colleague Ines Rehbein.

10 An important caveat about the analysis of the tweets is that our sample is somewhat biased: some tweets consist of more than one sentence but we always considered only the first sentence in the tweet. The decision to look only at the first sentence of the tweet is motivated by the fact that automatic sentence splitters perform poorly on Twitter data. So rather than also perform the sentence splits by

are shown in Table 3. As can be seen, Twitter has the highest rate of verb-initial units by far among the five corpora.

Twitter	8%
CALLHOME	3.5%
BUNDESTAG	2.5%
DEWAC	1.5%
HGC	1%

Tabelle 3: Percentage of verb-initial units

5.2. Construction types in a sample of verb-initial sentences

Table 4 shows the distribution of V1-constructions random samples of size N=100 from our five data sets. All samples were extracted by specifying a verbal form as the first token in the sentence. The first column, Type, classifies each construction as either belonging to the proper (*eigentliche*) V1-constructions or to the improper (*uneigentliche*) set.

---

hand, we decided to only look at the first sentence in a tweet. In tweets with more than one sentence, the first sentences may have different semantic-pragmatic characteristics than later sentences and so we are likely not getting the same rate of verb-initial sentences as we would have, had we been able to access all sentences in the Twitter data.

	Type	HGC	Dewac	Twitter	Bundestag	CallHome
Conditional / concessive inversion	P	22	32	2	4	0
Exclamative	P	2	0	2	2	0
Apodosis stranding	P	0	1	0	0	1
Formulas	I/P	0	1	0	0	12
Reporting inversion	P	0	0	0	0	2
da-drop	P	0	0	0	0	4
Presentational inversion	P	0	0	3	0	0
Contrast inversion	P	4	0	0	0	0
Yes-No-questions	P	53	22	31	45	32
Formal imperative	P	0	10	1	32	0
Infinitive imperative	P	0	0	1	0	0
Hortative/Optative	P	0	4	0	16	0
Informal Imperative	I	9	12	13	0	4
Subject topic-drop	I	8	15	34	1	25
Subject expletive drop	I	0	4	6	0	3
Cataphoric subject drop	I	2	0	2	0	0
Object topic-drop	I	0	0	5	0	17
Total		100	100	100	100	100

Tabelle 4: Frequencies of verb-initial constructions

The constructions in Table 3 have all been exemplified above except for what I called formulas. This basically refers to two items, *sag(e) mal* 'say' and *weißt Du* 'you know', used sentence-initially but of somewhat doubtful status as matrix-clause predicates. In the interest of surface-oriented analysis and because they seem to function in a distinct way as interactive units in the sense of Zifonun et al. (1997:62), I counted these items as separate cases of V1 rather than include them under their related form-types, the informal imperatives and the *yes/no*-questions.

Unsurprisingly, the interactive units *sage(e) mal* and *weißt du* are most distinctive for the spoken CallHome corpus, which also has the highest proportion of personal and demonstrative pronouns, response particles, and references to the here (*hier*) and now (*jetzt*), in line with its representing synchronous, though not face-to-face communication. Note that interactive units do occur in the other data sets, too. They just happen not to be represented in our particular random samples of V1-constructions. Other constructions that are distinctive for the spoken CallHome corpus include subject- and object-drop constructions, which according to Zifonun et al.'s typology are instances of situational ellipsis, relying on speaker-hearer synchrony. Object-drop occurs only in the CallHome and, with lesser frequency, in the Twitter samples. On the other hand, subject drop is even more common in the Twitter sample than in the CallHome sample. Altogether this suggests that Twitter data has characteristics of spoken conversation. Of course, the use of drop constructions on Twitter may also be motivated as a kind of structural ellipsis in a medium with constraints on message-length. It is also interesting to note that drop-constructions do not figure much at all in the Bundestag corpus, which is medially oral but conceptually written. With regard to omissions, the parliamentary speeches seem more "written" than the HGC newspaper corpus.

In addition, the CallHome sample is the only one that contains instances of *da*-drop and reporting inversion, shown in (30), which one might consider a subtype of *da*-drop.

30 Ich habe schon gesagt Mensch, ich habe zum Klaus gesagt, ich habe es ihm
    vorgespielt, habe gesagt, sage mal hat die Rosi einen traurigen Ausdruck
    oder aufgeregt . {Sagt}er nee, ganz normal . {Sage} ich, so ungewöhnlich,
    keiner hat Geburtstag, nix fällt an .
    'I said, man, I said to Klaus, I played it for him, I said, say, doesn't Rosi look
    sad or irritated. {Says} he, no, totally normal. I say, so unusual, it's nobody's
    birthday, there's nothing coming up.'

Yes/no-questions are most common in the HGC corpus sample, followed by the
Bundestag sample, and then Twitter and CallHome. It remains to be investigated
what functional roles these questions play in the various corpora. A similar question
arises with respect to the various types of imperative constructions found in the
corpora. They are particularly frequent in the Bundestag data, where the large subset
of optatives/hortatives is notable. The conversational CallHome corpus by contrast
contains very few instances of imperatives, even though these constructions are
hearer-oriented and, thus, very much compatible with synchronous conversation.

Inversions in conditional protases (cf. (16)) are (almost wholly) absent from the
Twitter and CallHome samples, which is not surprising since these constructions
typically go along with considerable sentence-length.

We return now once more to the drop-constructions. In the Twitter and Dewac
data, 'dropped' initial constituents were not always topics: we found quite a few cases
of expletives being dropped (see Section 6.1.). Also, some of the V1-formulations that
we found alternate with overt versions where a cataphoric pronoun would precede the
verb, as in (31).

31 ∅ {Bleibt} nur zu hoffen, daß angesichts der ausgeprägten
    Debattierfreudigkeit der Abgeordneten nicht plötzlich ein Streit über die
    Frage entbrennt, warum der Schwarzwald nicht vertreten ist ? (HGC)
    '[It] {remains} to be hoped that, given the extensive debate-happiness of the
    representatives, there won't arise a dispute about the question why the Black
    Forest is not represented? '

In addition, the distribution of the different drop-constructions differs between the
corpora. Subject drop is more common in the Twitter data than in the other samples,
while object drop is most common in the CallHome sample. Moreover, in the case of
subject drop, the samples differ not only in the relative frequency but also in the the
person feature of the dropped subject referents, as shown by Table 5.

	HGC	Dewac	Twitter	Bundestag	Call Home
1st Ps.	0	10	19	0	0
2nd Ps.	0	0	2	0	4
3rd Ps.	8	9	13	1	41
Total	8	19	34	1	45

Tabelle 5: Person feature of dropped subject referents

The dropped subject topics in the HGC are always 3rd person referents, while in the Twitter data they are somewhat more likely to be 1st person referents than 3rd person referents. This difference may just reflect the fact that 1st person subjects in general are much rarer in the HGC than in the Twitter data: in a sample from the HGC of 200 verbal predicates, there were only 3 cases of overt first-person subjects, while in a parallel Twitter sample there were 31. In the Dewac sample, the distribution is overall similar to that found in the Twitter sample. The most surprising finding is the fact that the CallHome sample patterns most closely with the written HGC corpus: its omissions predominantly concern 3rd person referents and there were no 1st person cases at all.

Finally, we note that the cases of object-drop in the Twitter data all concern 3rd-person referents, which matches Schalowski (2009)'s finding for the instances of object-drop in a corpus of forum posts. Similarly, according to a chi-square test, the distribution of person features among the dropped subject referents in the Twitter data in Table 5 cannot be distinguished from the distribution that Schalowski (2009) found for the dropped subjects in his social media data set.

To sum up: the types of V1-constructions are distributed very differently across the corpora. Some types such as conditional inversion seem to be largely restricted to the newspaper data, while others such as topic-drop occur mostly in the spoken corpus. However, no overall clear picture emerges based on our 5 data sets with respect to which data sets are most similar to each other. What does seem clear is that the Bundestag corpus is the corpus that is most different from all the others with respect to the patterning of V1-constructions. And Twitter is the corpus that uses the largest share of improper V1-constructions that are accompanied by argument-drop. Taken together with the fact that Twitter also has the highest V1-rate, as shown above in section 5.1., it seems that the constraints of the medium with respect to message-length are reflected in a higher use of improper V1-constructions that allow greater information density through omission of recoverable material.

We conclude the section with a quantitative caveat. The counts derived from the five samples may suggest that some constructions figure in only one or two of the two corpora. For instance, cases of contrast inversion were found only in the HGC sample; conditional inversion is mostly a feature of edited written language; instances of presentational inversion were only found in the Twitter sample. While we can be fairly confident about the distribution of conditional inversion where we found relatively large numbers in relation to our small sample size, we cannot be so sure about the other two constructions that have much lower frequencies: Would the differences between corpora showing up in our samples stably reappear in larger

samples? Minimally we have to keep in mind that our corpora do not represent the totality of the respective text or media types involved. To see this we just have to remind ourselves that, for instance, the HGC does contain instances of presentational inversion--after all, example (15) is taken from that corpus--or that Auer (1993) and Günthner (2000:15-19) have shown that presentational V1-constructs play important roles in spoken language (such as creating cohesion with prior context or providing for more lively/dramatic narration).

6 Argument drop in active-form declarative clauses

In this section, we specifically consider sentences in which an argument is omitted by construction in a finite active-form declarative clause. The general question that we ask is whether all such omissions result from a single unitary topic drop construction or whether there are different types of constructions at play. We first consider cases in which the dropped elements are expletives, that is, syntactic arguments only, but without a semantic role. Afterwards, we examine what constraints exist on such omissions and what kind of grammatical analysis should follow from them.

6.1. Expletive-drop in the Twitter data

Expletives are defined to be non-referential nominals that serve to fill certain structural requirements, in particular one for a clausal subject to be present. Given their non-referring status, expletives should not be able to be targeted by topic drop, and this claim is in fact made by Fries (1988:34). Fries is, however, aware that some omissions of expletives can be found but explains them away as cases that may be acceptable to speakers (though still not grammatical) because topic-drop is a phenomenon at the margins of German grammar where judgments are expected to be less crisp.

We will not adopt such a view and instead treat expletive drop as an ordinary construction of spoken German, especially since in our Twitter data, one does encounter omitted expletives quite frequently. Table 7 shows the subject realizations of five verbs denoting precipitation events in the Twitter data set. While the number of instances is small, one can see that expletive drop is attested with all of the verbs, though it may be more common with one of them (*regnen*) than with the others.[11]

Form	regnen	pissen	nieseln	schiffen	schneien
	'rain'	'rain hard'	'drizzle'	'rain hard'	'snow'
zero	11	1	1	10	1
das	3	2	1	2	1
es/'s	23	61	41	33	96
other	0	1	0	0	0
Total	37	64	42	45	98

Tabelle 6: Realizaton of expletive with weather predicates

11 Note the results for *schneien* are from a later version of the Twitter corpus. The version used throughout the paper was collected during the summer of 2012 and therefore contains no mentions of snow and snowing.

We also looked for the verb *regnen* in our other corpora, inspecting 200 random instances (fewer, if fewer were available in the corpus). Basically, we found almost no zero-realizations in the samples from the Bundestag (0/5), CallHome (0/11), Dewac (1/200), and the HGC (0/200). However, since there are very few instances of the predicates involved in the CallHome and the Bundestag data,we cannot come to any confident conclusions about expletive-drop with *regnen* there.

Returning now to the analysis, we may ask whether expletive-drop behaves like topic drop or not. One important aspect in which it does behave like topic drop is the fact that it seems acceptable only if there is nothing occupying the pre-field. While the canonical versions (32a) and (32c) are acceptable, dropping the expletive is only acceptable in (32b) but not in (32d).

32   Today it's {raining} cats and dogs.

     a) Es {regnet} in Strömen heute.

     b) ∅ {Regnet} in Strömen heute.

     c) Heute {regnet} es in Strömen.

     d) # Heute {regnet} ∅ in Strömen.

The unacceptability of (32d) is unexpected if expletive drop is owed to a different, independent licensing mechanism from topic drop: if expletive drop has nothing to do with discourse status, why should sentence position matter?

It is also our impression that expletive drop is associated more with spoken language than with written language, as is the case for topic drop. In the written language, one does not tend to find cases of missing expletives, as noted above. One example contained in the Dewac, though not part of the sample examined above, occurs in a literary quote from Döblin's *Berlin Alexanderplatz,* which*,* suggestively, involves interior monologue:

33   Raus auf die Straße ! Luft ! ∅ {Regnet} noch immer . Was ist nur los ? Ich muß mir ne andere nehmen . Erst mal ausschlafen . Franz, wat is denn mit dir los ? (HGC)
     'Out into the street! Air! Still {raining}. Just what is going on here? I have to find another one. First a good night's rest. Franz, what is the matter with you? '

However, without a larger database we cannot verify how frequent subject-less expletives are with weather predicates in conversational language.

6.2. Constraints on topic drop

In section 5.2. we found that the HGC, Twitter and CallHome data contain instances of topic drop for subject and object arguments. Of the two kinds of arguments, subject drop was much more common for the HGC and Twitter, but less pronouncedly so in the conversational CallHome corpus.

Although topic drop is assumed to be motivated by high accessibility–in Schalowski (2009)'s words, topic drop targets familiarity topics–it seems unable to apply to arguments other than subjects and objects whatever their accessibility. Notably, indirect objects seem not to be droppable; a failed attempt of omitting a third

person indirect object is given in (34).[12]

> 34 # ∅ {Helfe} ich morgen bei den Hausaufgaben.
> 'I {help} [them] tomorrow with their homework.'

Twitter data that we inspected for verbs such as *helfen* 'help', *schenken* 'give as gift', *spenden* 'donate' did not yield any instances that were recognizable as topic drop. The finding that indirect objects cannot be dropped is very surprising given that indirect object referents are generally coded in a way that suggests very high accessibility.

Similarly, it has been held that prepositional phrase arguments cannot usually be dropped. While we didn't find an instance in the Twitter data, we were able to find an example in a web forum. In (35), the Content argument of the verb *erinnern* 'remember' in user Eki's reply is omitted with the verb in utterance initial position and the subject and reflexive object realized, which suggests that the construction used is topic drop.

> 35 http://www.gtrp.de/archive/index.php/t-1152.html
> **Chakra**
> 27.02.2002, 11:55 So, habe jetzt meinen Tommi Mäkkinen IA1 Endurance von 9 Stunden mit "Genau mein Ding" überboten -you remember,Eki-
> Die Kombo macht'nen Riesenspaß, obwohl ich mich irgendwo immer verfranse:rolleyes:
> Ja, Flinx, kann mir gut vorstellen, daß Du 1:31 schaffen kannst, bzw. als virtuelle lap-time schon hast:) Bin jedenfalls gespannt.
> **Eki**
> 27.02.2002, 12:41 ∅ {Erinner} ich mich gut; jaja, der Tommy. …
> 'I {remember} ∅ well; yes, yes, Tommy . ...'

While the above example seems like a very clear case, there are additional instances that could arguably be seen to involve topic drop of a prepositional argument. In (36), the verb *kommen* in the relevant sense ('think of/hit on') requires an *auf*-PP encoding a Content semantic role. However, when the Content semantic role is anaphorically accessible, the anaphoric form *darauf*, which fuses the preposition with a demonstrative pronoun, is standardly used, as in (37). An anaphoric element *da* can also be fronted in a kind of reduplicative construction (cf. (38). The sentence in (36) could thus be seen as involving the dropping of the initial *da* found in (38).

> 36 Die vierte frage konnte ich leider nicht beantworten, da ich mir nicht Sicher war was sie meinte und bevor ich Blödsinn antworte, lieber sage:
> ∅ {Komme} ich gerade nicht drauf, tut mir Leid!
> (http://www.bundeswehrforum.de/forum/index.php?topic=23969.5;wap2)
> 'Unfortunately, I couldn't answer the fourth question because I wasn't sure what she meant and rather than reply with some nonsense, I prefer to say "I can't {think} [of it] right now, sorry! "

---

12 It is possible to drop the indirect-object in some cases: "Hilft beim Abnehmen" 'Promotes weight loss [lit. Helps with losing weight]'. However, this seems to be limited to generic statements about the efficacy of means and instruments but not available when talking about human agents.

37 Ich {komme} gerade nicht darauf/drauf.

38 Da {komme} ich gerade nicht drauf/darauf.

Now, one might object that in (36) above, the anaphoric element is fully instantiated in the fused form and if anything is omitted sentence initially, it might just be an adjunct. This objection would however not apply to cases like (39), where the verb *halten* occurs in a sense 'X holds Y in Z regard' that requires a *von*-PP encoding the evaluated entity. In standard written German, the sentence would have to be formulated as in (40), with an element fusing the preposition and the anaphoric element. In example (39), the anaphoric element is completely missing but has to be understood. Moreover, the initial position of the verb suggests again that a topic drop construction is being used. If that is correct, then it is an instance of topic drop that does not involve a subject or object. (For Fries (1988), the above examples are a special case of the phenomenon that he basically calls Pronoun Zap following Huang (1984). He refers to them as anaphor deletions.)

39 @hertizworld Genau . Ein Spieler der abgestiegen ist, wird uns weit nach vorne bringen . Ne, ∅ {halte} ich nichts von . Weder so noch so . (Twitter) 'Exactly. A player who was relegated will take us forward. Nah, I don't {believe} in [that]. Neither one way nor the other.'

40 Nein, davon {halte} ich nichts.
'No, I don't consider it useful at all.'

Finally, we return to cases of protasis-drop, as seen in example 18 and 19 above. Example 18 is repeated here for convenience:

41 ∅ {Gehe} ich halt ohne Hose los ! (Twitter)
'[So then] I'll just {go} off without pants'

A plausible analysis is that sentence (41) drops an anaphoric *dann*, marking the contextually given protasis. While temporal *dann* is ordinarily an adjunct, the *dann* in such a conditional context has to be seen as one of the two situation-denoting arguments of the conditional construction. On that analysis, sentence 41 also is a case of argument topic-drop, but of an argument of adverbial form. Neither Fries 1988 nor Schalowski 2009 discuss these cases in the context of topic drop.[13]

6.3. Inflectives

Twitter data, like other computer-mediated communication, also exhibits inflectives, a stilistic device that is commonly found in comics (Schlobinski 2001). The prototypical instance of this kind of construction consists of a bare verb stem as in example (42).

42 463085: Dann halt nicht schlafen . *aufräum*

---

13 The description of the kind of usage found in (41) is probably still incomplete. For instance, it seems that the conditional consequent typically is used in an exclamative context, including a discourse marker such as *halt*.

       Then not sleep. *Tidying up*
43 20525: *Bagel mampf* sooo lecker :)
     '*Bagel eat* soooo good :-)

In example (43), the bare stem is preceded by a verbal argument. On Twitter, instances of this construction are usually marked by inclusion in a pair of asterisks.

The verbal forms that occur in such contexts are called *inflectives* following Teuber (1998), who discusses the place of these forms within the morphological paradigm of German. As shown by example (43), inflectives can also occur together with arguments or modifiers.[14] We will refer to the actual occurrences of inflective verb forms together with any arguments and modifiers as inflective constructs.[15]

Based on our corpus work, it seems that the functional purposes of inflectives, or a very similar set of purposes, can also be subserved by forms that are in fact inflected, as shown by (44), or that are not verbally headed at all, as shown by (45).

44 ... nun was das angeht ... öhm ... *schaut sich nervös um* Ähm ... Oh seht mal , eine doofe Mitten im Leben-Folge : http://t.co/S4x1UQNw
    'well, with regard to that … ahem *looks around nervously* ahem … Oh look, a stupid episode of "In the Midst of Life" '
45 92316: @schattenbran empty Du weist warum :D *dreckiges lachen*
    'You know why :D *dirty laughter*

We will reserve the term inflective construct for verbally headed cases, following Teuber and Schlobinski. We will call cases such as (44) amd (45) pseudo-inflective constructs.

What interests us about the verbal cases is their argument structure since normally inflectives lack at least one of their semantic roles, the one that would be realized as subject in finite active-form sentences. To study the argument structures of inflectives, we drew a random sample of 500 tweets. Each candidate tweet to inspect was initially identified based on the the criterion that it contained a token starting with an asterisk. We checked all the candidate tweets by hand and excluded pseudo-inflective constructs such as (44) and (45) above and cases where the asterisks are used for other purposes such as emphasis, as in (46), as well cases where it was not clear what the function of the asterisk was. For instance, in (47), it seems the writer might mark the playful use of a foreign word-phrase (French-English "le me").

46 @BassLove_ ja wenigstens durft ich kaya *ich hab geheult als die karten da warn*
    'at least I was allowed to kaya *i cried when the tickets got here*'
47 *Le me geht jetzt mal zum Briefkasten !
    'Le me [the me] is going to the post box now.'

Regarding the function of inflective constructs, Teuber suggested that they mainly serve to express a subjective response or appraisal. This applies to many examples but as (42) and (43) show, it is not always the case. Let us compare (43) above with the

---

14 There is a question to what extent incorporation of such arguments or modifiers occurs in actual uses of inflectives, correlated to some degree with spelling as a single word. We will not address this issue here (but see Schlobinski 2001).

15 We leave open the question whether the pragmatic/functional characteristics that linguists have observed when studying inflective constructs should be attributed to abstract reflective constructions rather than the verb form.

alternative formulation in (48).

48 (Ich) mampfe eine Bagel. sooo lecker :)
'∅ Eating a bagel.   soooo good :-)

It seems to us that by using the inflective formulation, the speaker asks the hearer to treat the information communicated by the inflective as if they had observed it, rather than been told about it. If that is so, use of the inflective constructs on Twitter serves to re-create a sense of face-to-face or multi-modal communication that distant communication by a broadcast medium like Twitter does not offer. The question is whether one should treat cases such as (42) and (43) as different from ones like (49), where the inflective construct references an emotional state or appraisal.

49   @R3b3ccaCran3 *dich hass* Wir sind extra um 6 aufgestanden, damit wir sie schauen können <3<3<3 morgen auch wieder
'*hating you* We specially got up at 6 so we could watch them <3<3<3 same again tomorrow.

(49) seems like a case where the inflective construct very clearly conveys an internal state. However, one could still think of it as replacing missing multi-modal information: people in general are very good at recognizing emotion in others based, for instance, on facial expression and tone of voice.

We now turn to the analysis of the form and argument structure of the inflective constructs in our sample. The 500 candidate tweets inspected contained 206 inflective constructs. 9 of these cases had reduplicated inflective forms of the type *freu-freu* 'happy-happy', *quietsch-quietsch* 'shriek-shriek', etc. All the inflective constructs with reduplication were based on intransitive verbs. With the exception of example (50), all inflectives were oriented towards the first-person.[16]

50   Geil ! *Wasser im Mund zusammenlauf* RT @ChrissyRamone : La Fischtheke au Wissembourg . http://t.co/89qXIT3Q
'Sweet! *Salivating* '

Table 7 shows the distribution of patterns for the prototypical case, inflective constructs based on a simple inflective verbal form. It lists only the patterns with more than 1 instance. In addition to those patterns, there are 32 other unique patterns represented in the sample. In the table, we distinguish particle verbs from simple verbs by the label PartV in order to draw attention to their high frequency among the inflective forms.

---

16 Of course, example (50) also involves the author of the tweet as an experiencer, though that role is not realized as a subject with the idiom *das Wasser im Mund(e) zusammenlaufen* 'salivate [lit. the water is collecting in one's mouth]'.

V	34
ObjNP-V	32
ObjNP-PartV	28
PP-V	16
Adv-V	7
ObjNP-PP-V	6
Adv-PartV	6
PartV	5
PP-PartV	4
ObjNP-PP-PartV	4
IndObjNP-ObjNP-PartV	4
Adj-V	4
PP-PP-V	3
ObjNP-V-Clause	2
NP-PartV	2
IndObjNP-ObjNP-PP-V	2
Adv-PP-V	2
Adv-ObjNP-V	2
Total	163

Table 7: Distribution of patterns with frequency greater than 1

As shown by Table 7, most of the instances do not match Teuber's characterization as one-word sentences. While one-word instances consisting of just a predicate as in example (42) are frequent, instances where an inflective verb form occurs together with one or more non-subject complements or adjuncts are much more common, Schlobinski's data show many complex inflective constructions, too. Further, may examples in our sample as well as many of the ones shown in Schlobinski's work, contradict Teuber's analysis that inflectives have a syntactically empty valence. An incorporation analysis would not work for cases where specifically referring NPs appear as arguments of inflectives within inflective constructs, as in (51).

51 @HoneyballCookie Ich warte heute mit Papa lecker essen . ^_^ *dir auch etwas rüberschieb*
'Today I had a good meal out with dad. * to-you something slide-over, too *

In the majority of cases, the verbal form is placed at the end of the inflective group, which Schlobinski (2001:206) attributes to the general fact that final position is the norm for non-finite forms in German. In our sample, the exceptions consist of cases where there is a clausal complement that is arguably shifted rightward because of its heaviness (cf.(52)).

52 @fhainalex Amt ? Oha ! *Daumen drück , daß alles einigermaßen gut läuft*
    'Agency? Whoa! *Keeping fingers crossed that everything works out
    tolerably*

We turn now to the verbs that occur as the heads of the inflective constructs.
Within our 206 instances of inflective constructs, a large variety of inflective forms is
represented. Only 6 verbs occur more than 5 times, namely:

53 *winken* 'wave' (6), *drücken* 'cuddle' (7), *reichen* 'pass', (7), *freuen* 'be happy'
    (8), *rüberschieben* 'slide over' (10), *lachen* 'laugh' (11)

To provide some further abstraction over individual words, we labeled the verbal
inflectives heading the inflective constructs in our sample with their appropriate
FrameNet-frame. Table 8 shows the frames that were evoked by the heads of
inflective constructs more than once. In addition, there are 54 other frames that
occurred only once.

Make_noise	20
Cause_motion	18
Body_movement	14
Perception_active	14
Emotion_directed	11
Giving	10
Self_motion	7
Hug*	6
Becoming	5
Cause_impact	5
Cause_to_move	5
Hold_thumbs	5
Placing	5
Cause_to_move_in_place	4
Communication_manner	4
Facial_expression	4
Breathing	3
Experiener_focus	3
Intentionally_act	4
Being_attached	2
Cause_bodily_experience	2
Departing	2
Manipulation	2
Removing	2
Taking	2
Waiting	2

Table 8: Frames with frequency greater than 1 evoked in inflective constructs

Compare Tables 7 and 8 with the following list, which contains the most frequent single-word inflective constructs in all of our Twitter data:

54 *g* 'grin', *hust* 'cough', *lach* 'laugh', *freu* 'be happy', *seufz* 'sigh', *gg* 'grin a lot', *lol* 'laugh(ing) out loud' , *gähn* 'yawn', *grins* 'grin', *sing* 'sing', *kicher* 'giggle', *wink* 'wave'

It seems that when one considers all kinds of inflective constructs, including the ones that consist of more material than just a simple inflective verb form, the semantics of the occurring inflective forms appears more varied. Of course, the usual suspects,

verbs such as *grinsen* 'grin' and *lachen* 'laugh', are well represented among the frequent cases when looking at inflective constructs of any length, but it is also apparent that many other kinds of situations, especially ones involving motion (e.g. Self_motion, Cause_to_move, Cause_motion, Cause_to_move_in_place, Giving), are also common (cf. examples 51 and 52).

We now consider the realization of arguments within inflective constructs. First, it is the case that in all cases in our sample, the subject argument that is filled by the author referent is unexpressed. As noted by Schlobinski (2001: 208), the speaker role is pre-set as a "default" subject. More interesting is the question to what extent objects and indirect objects are realized or omitted. As also pointed out by Schlobinski (2001: 210), within the right scenario, the addressee can also be omitted, as in (55). [17]

> 55   @MiyaSekai @SenjoVal empty woa ich hoff ihr schafft das <;O< *anfeuer ! *
>
> 'Whoa I hope you manage it – *cheering*'

It is notable that the non-realization of an argument role filled by the addressee is not only possible when the referent would appear as a direct object but also with certain indirect objects that normally resist omission in simple assertions in episodic contexts. An example of this is shown in (56).

> 56   @tobi_SE ruhig bluut ... *massage geb*
>   'calm down … *giving a massage*

It is an interesting idiosyncratic fact about argument omission within inflective constructs that the high salience of the addressee seems sufficient to license omission where this same omission is not licensed by the lexical items or by general language constructions.

6.3.1 .   Pseudo-inflectives

Although they are not proper inflectives, we want to briefly consider the cases where inflected verb forms are used by Twitter authors within the asterisk-markup.

One thing to notice is that while the word order is sometimes verb-final as in the case of inflectives proper (cf. 57), the inflected verb is in initial position in other cases such as (58), where there doesn't seem to be a clear need for post-verbal placement of the modifier based on its weight.

> 57   @LederundSpitze Wer darf die neuen Worte denn alles lesen ? *unsicher} schaut*
>   'So who all is allowed to read the new words? * looks uncertainly *
> 58   *pfeift ganz laut*
>   'whistles very loutdly'

In terms of their function, the pseudo-inflective constructs in our data cases still talk about the tweet-author. They are merely presenting information about the author as if

---

17 Of course, reflexive objects co-referring with the authors, as in the frequent
   instances of *freu* 'be happy' is also possible.

there was an external viewer present (hence the third-person person feature). However, while the third person is the dominant option, it is not the only one within pseudo-inflectives. One also finds some instances of first-person pseudo-inflectives, as indicated by the verbal morphology.

59 @Ashqtara Ich habe was für dich , ich weiß nicht ob du das schon gesehen hast . *lieber schnell flüchte* http://t.co/iL6pk1cc
    'I've got something for you, I don't know if you've seen it yet. *better run away quickly*'

Coming to firm conclusions about how pseudo-inflective constructs behave is difficult since we had only 32 instances of pseudo-inflective constructs within our 500-tweet sample. In future work, we would like to test on a larger set of pseudo-inflective constructs whether in terms of variety, complexity, and function, they are different from inflective constructs or not.[18]

6.4. Analysis: a family of argument-drop constructions

What should we conclude about the grammar of argument drop based on the findings in sections 6.1. and 6.2.? First, we would argue that expletive drop is its own phenomenon: though it seems a lot like real subject topic drop on the surface, it cannot have the same functional/pragmatic motivation that needs to be part of the constructional analysis.

The second major question then is what we should do about the treatment of regular arguments. Before we make a proposal let us consider a bit more data. In (60), we have an uninstantiated first person referent in the subject function. If it were to be expressed explicitly, it would be realized as a personal pronoun. In (61) and (62), third person referents in the object role are uninstantiated. If they were to be explicitly realized in the fronted topic position, they would have the form of a demonstative rather than a personal pronoun: the personal pronoun seems completely ungrammatical in (62) and the only acceptable reading for (61) with a personal pronoun would involve contrast, which, however, is not relevant in the context of the actual Twitter thread.

60 (Ich) {Bin} jetzt weg. (Twitter)
    'I {am} off now.'
61 Fady   Malouf.   (Den/?Ihn)   {Mag}   ich   nicht   .   (Twitter)
    'Fady Malouf. [that one/him] I don't {like}.'
62 @KellyKoksNuss öhm (Das/*Es) {weiß} ich jetzt noch nicht :) (Twitter)
    'ahem [that/it] I don't know yet'

Thus, it seems that the activation status requirements for subjects and objects are different. And actually, the activation status requirements for objects seem to be more

18 One might suspect that maybe pseudo-inflectives are produced by authors who have not yet mastered the grammar of proper inflectives, or produced automatically by a software auto-correct feature. While that cannot be ruled out for individual instances, it seems there are too many of them for them all to be just occasional errors.

like those of the omissible prepositional arguments in (36) and (39), whose overt counterparts also involve demonstrative forms with *da*. Not only that but both types of omission also share the constraint on the person feature: fused forms consisting of *da* and a preposition can only be anaphoric to third persons, and third persons are also the only omissible referents in object drop.

Overall, it seems that we should favor an analysis that posits at least three different argument-drop constructions: one for expletive subjects; one for referring subjects; and one for referring objects and prepositional objects. While our discussion still does not explain the non-droppability of indirect objects, the constructional analysis could handle the facts right by paying attention to grammatical functions.

Finally, we note that topic drop may also be facilitated by structure-parallelism contexts, as discussed by Fries (1988). While in (63) the omitted element in the answer would have the same form (and semantic role), there is a change in form and role between B's question in (64) and A's two possible answers.[19]

63 In Köln ist viel Streß, und wie ist es in Tübingen?
'It's a lot of stress in Cologne, what's it like in Tübingen?'
A: ∅ {Ist} alles ziemlich lahm (= Fries' example (110))
'It's all pretty lame [there].'
64 A: Also Berlin mag ich.
'So I like Berlin.'
B: Und Köln?
'And what about Cologne?'
A1: ?? ∅ {Wohn/leb'} ich.
'I live [there]'
A2: ?? ∅ {Spielt} sich doch nichts ab. (= Fries example (109))
'There is nothing going on [there]'

The *in*-PP in (63) is neither a subject nor an object and it's not a regular prepositional complement. Still, the discourse context sets it up as a topic which can be dropped in a sentence where the omitted constituent would play the same semantic and syntactic role. Crucially, structure parallelism is not generally necessary for topic drop, where, for instance, an aboutness-question can precede topic drop as in (65). This contrasts with example (64) where an aboutness-question regarding Cologne does not allow for the subsequent omission of a locative anaphor.

65 Gisbert mag dich, und was ist mit dem Sascha?
'Gisbert likes you, and what about Sascha?'
A: ∅ {Mag} er auch.
'∅ Likes him, too'

Further investigation is needed to ascertain to what extent structural parallelism goes along with various kinds of topic drop for which it is not strictly necessary.

Finally, inflective constructs need to be handled by another independent construction. The word order found with the vast majority of verbal inflective constructs is verb-final, which prevents us from treating them as a subtype of subject

---

19 If we wanted to make the antecedent explicit, it would have the form *da* in both cases. However, that use of *da* would be as a real locative adverbial in contrast to its purely anaphoric uses seen above.

topic-drop. Moreover, they have a strict focus on the here and now and do not allow past-tense or perfect forms reporting on past events or states, which subject-drop does. Finally, inflective constructs can license omissions of arguments, especially indirect objects, that are not omissible through topic-drop.

7 Discussion & conclusion

In section 4 , we used corpus data to check the plausibility of the generalization proposed by Ruppenhofer (2004) that an omitted semantic role receives the same (anaphoric or existential) interpretation across *lexically* licensed omissions by the members of a particular lexical class. For the lexical classes that we considered here, the generalization was indeed found to hold.

In our second study, presented in section 5 , we were interested in *constructional* argument omissions, in particular various forms of argument drop. Comparing data from social media and from spoken and written corpora with respect to verb-initial constructions, we saw that they differed in the frequency with which the various constructional types occurred. Focusing on argument-drop cases in particular, we saw that generally subject topic-drop was more common than object-topic drop. Object drop was most frequent in the conversational CallHome data. In the Twitter data we also found object topic-drop to be less common than subject topic-drop, in line with findings by Schalowski (2009) for another social media set. In addition, we also encountered a less expected result, namely that expletives are dropped quite frequently and that expletive drop seems, in some respects, like topic-drop even though it arguably is not a subtype of the latter.

We also looked at inflective constructions, determining that they have special properties different yet again from regular subject-drop. In the case of inflectives, we did not systematically look for instances in our written corpus, the HGC. Queries for highly frequent verbal inflectives as found on Twitter e.g. *freu* 'be happy' and *guck* 'look' yielded no results, however.

Taking a broader perspective, we note that the findings on the data considered here fit the analysis of Ruppenhofer and Michaelis (2010) for their data, namely that argument omission can be described in terms of "constraints on argument structure, as the relevant conventions target specific semantic and grammatical roles of verbs" (p. 160). Against this background, it is not surprising to observe that, although there are family resemblances among omission-licensing constructions, these constructions also exhibit idiosyncratic differences and limitations that simply call for individual, separate treatment.

Besides the need for specific grammatical treatment, the data also illustrates that not all omission constructions are motivated the same way. While topic drop may primarily be driven by high accessibility, measurable through frequent mention and short distance to a preceding co-refering mention, other constructions may be motivated by other notions of prominence. As discussed by Ruppenhofer and Michaelis (2010), in sentences from English match reports such as (66), the object of play (typically a ball) may go unexpressed simply because it is a globally prominent referent throughout the text via the overall scenario of the game.

66   He hammered ∅ wide of Gary Walsh's exposed net.

If our analysis is on the right track, then even the superficially similar cases of

subject and object drop involve different degrees of accessibility and are not motivated in exactly the same way. In the case of inflectives, it is even clearer that they serve a special communicative function and have corresponding morpho-syntactic constraints.

Finally, let us consider the question what our study has to say about the relation between conceptually spoken language and argument omission affordances. As we saw, the conversational CallHome corpus clearly has a high incidence of improper V1-constructions licensing argument omission (cf. Table 4). The Twitter data actually shows an even higher incidence. As discussed in section 5.2., on Twitter the space constraints of the medium may provide additional motivation towards reduction, beyond what applies to spoken language.[20] All other corpora examined, which are more conceptually literal than the CallHome and Twitter, make much less use of anaphoric omissions enabled by improper V1-constructions. This might suggest that constructional omissions with anaphoric interpretation belong to the language of closeness (*Sprache der Nähe*) that is at the heart of conceptually oral language.

However, we think this point requires further study. First, we would need to accumulate more evidence that Twitter exhibits other features of oral language. And second, we need to relate our observations to the findings of Schwitalla (1988), who reports no significant differences in omission rates between spoken and written language. It is not quite clear, though, to what extent Schwitalla's results are relevant. While we looked only at omissions constructionally licensed by improper V1-constructions, he may have included all types of omissions, including lexically licensed ones, in his analysis. In any event, studying the frequency of lexically licensed omissions is in itself relevant for forming a conclusion about whether argument omissions can serve as an index of the language of closeness. The idea to us makes sense in terms of the communicative conditions and implementation strategies that Koch & Oesterreicher's discuss. In terms of the code, the exploitation of omissions raises the information density of the text, making it more compact. In terms of the communicative requirements for omissions to succeed, some or all of the factors such as face-to-face interaction, involvement, dialog, familiarity, and situational interlocking that Koch & Oesterreicher mention do seem relevant to each of the anaphoric omission constructions. We therefore consider it a worthwhile topic for further research to see to what extent the frequency of anaphoric omission constructions correlates with conceptual orality.

---

20 In future work, we plan on studying a corpus of electronic text messages (sms) in order to verify if constraints on message-length have a similar effect on the frequency of anaphoric omissions there, too.

8 References

Ariel, Mira. 1988. Referring and Accessibility. *Journal of Linguistics* 24**.** 65--87.

Auer, Peter. 1993. Zur Verbspitzenstellung im gesprochenen Deutsch.*Deutsche Sprache* 3.193--222.

Baroni, Marco ; Silvia Bernardini, Adriano Ferraresi & Eros Zanchetta. 2009. The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation* 43**.** 209--226.

Bender, Emily. 1999. Constituting Context: Null Objects in English Recipes Revisited. *Penn Working Papers in Linguistics* 6(1). 53--68.

Culy, Christopher. 1996. Null object in English recipes. *Language Variation and Change* 8. 91--124.

Fellbaum, Christiane & Judy Kegl. 1989. Taxonomic Structures And Cross-Category Linking in the Lexicon. In Kenneth de Jong & Yongkyoon No (eds), *Proceedings of the Sixth Eastern States Conference on Linguistics*, 93--104.

Fillmore, Charles J. 1982. Frame Semantics. In Linguistic Society of Korea (ed.)., *Linguistics in the Morning Calm*, 111--138. Seoul: Hanshin

Fillmore, Charles J. 1985. Frames and the semantics of understanding. *Quaderni di Semantica* 6(2). 222--254.

Fillmore, Charles J. 1986. Pragmatically Controlled Zero Anaphora. In Vassiliki Nikiforidou, Mary Van Clay, Mary Niepukuj, Deborah Feder (eds.), *Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistic Society,* 95--107. Berkeley: Berkeley Linguistic Society.

Fitschen, Arne. 2004. Ein computerlinguistisches Lexikon als komplexes System. Stuttgart: Institut fur Maschinelle Sprachverarbeitung der Universität Stuttgart dissertation.

Fries, Norbert. 1988. Über das Null-Topik im Deutschen. *Forschungsprogramm Sprache und Pragmatik* 3.

Fraurud, Kari. 1996. Cognitive Ontology and NP Form. In Thorstein Fretheim & Jeannette K. Gundel (eds.), *Reference and referent accessibility, 65--67.* Amsterdam/Philadelphia: John Benjamins.

Goldberg, Adele E. 2006. *Constructions at Work.* Oxford, UK: Oxford University Press.

Gundel, Jeanette K., Nancy Hedberg, & Ron Zacharski. 1993. Cognitive Status and the Form of Referring Expressions in Discourse. *Language* 69(2). 274--307.

Günthner, Susanne. 2000. Creating scenic moments: grammatical and rhetoric-stylistic devices for staging past events in everyday narratives. Interaction and linguistic studies 22.

Haegeman, Liliane. 1990. Understood subjects in English diaries: On the relevance of theoretical syntax for the study of register variation. *Multilingua* 9(2). 157--199.

Huang, C. 1984. On the distribution and reference of empty pronouns. *Linguistic inquiry* 15.531--574.

Krisjanis Karins, Robert MacIntyre, Monika Brandmair, Susanne Lauscher and Cynthia McLemore. 1997. CALLHOME German Transcripts. LDC97T15.

Koch, Peter & Wulf Oesterreicher. 1985. Sprache der Nähe--Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte (Langage de la proximité--langage de la distance. L'oralité et la scripturalité entre la théorie linguistique et l'histoire de la langue). *Romanistisches Jahrbuch* 36. 15--43.

Lehrer, Adrienne. 1970. Verbs and Deletable Objects. *Lingua* 25. 227--253.

Michaelis, Laura A. 2011. Knowledge Ascription by Grammatical Construction. In John Bengson & Marc A. Moffett (eds.), *Knowing How: Essays on Knowledge, Mind, and Action,* 261-279. Oxford: Oxford University Press

Mittwoch, Anita. 1982. On the difference between eating and eating something: Activities versus accomplishments. *Linguistic Inquiry* 13(1). 113--22.

Rehbein, Ines, Josef Ruppenhofer, Caroline Sporleder & Manfred Pinkal. 2012. Adding nominal spice to SALSA–frame-semantic annotation of German nouns and verbs, In Jeremy Jancsary (ed.), *11th Conference on Natural Language Processing (KONVENS)*, 89--97. Vienna; ÖGAI.

Richling, Julia. 2008. *Die Sprache in Foren und Newsgroups : eine Untersuchung der konzeptionellen Mündlichkeit und Schriftlichkeit im Wandel der Zeit.* Saarbrücken: VDM Verlag.

Ruppenhofer, Josef. 2004. The interaction of valence and information structure. Berkeley: University of California dissertation.

Ruppenhofer, Josef & Laura A. Michaelis. 2010. A constructional account of genre-based argument omissions. *Constructions and Frames* 2(2). 158--184.

Schalowski, Sören. 2009. Über Topik-Drop im Deutschen: Untersuchung zum Einfluss der grammatischen Funktion und des Merkmals Person. Berlin: Humboldt Universität zu Berlin Master's thesis.

Schlobinski, Peter. 2001. *knuddel – zurueckknuddel – dich ganzdollknuddel* Inflektive und Inflektivkonstruktionen im Deutschen. ZGL. Zeitschrift für germanistische Linguistik 29.2: 192-218.

Schmid, Helmut. 1994. Probabilistic part-of-speech tagging using decision trees. In

Daniel Jones (ed.)*, Proceedings of the international conference on new methods in language processing*, 44—49. Manchester, UK.

Schwitalla, Johannes. 1988. Kommunikative Bedingungen für Ergänzungs-realisierungen. In: Gerhard Helbig (Hrsg): Valenz, semantische Kasus und/oder "Szenen". Berlin, 74-84.

Teuber, Oliver. 1998. fasel beschreib erwähn–Der Inflektiv als Wortform des Deutschen. *Germanistische Linguistik* 141/142. 6—26.

Zifonun, Gisela, Ludger Hoffmann & Bruno Strecker. 1997. Grammatik der deutschen Sprache. Berlin/New York: Walter de Gruyter.