Self-embedding and complexity in oral registers

Abstract

This article reports the results of a study on the self-embedding depth of nominal, verbal and clausal projections in spoken corpora of German. In order to gain insights about the intentional processes that influence the choice of recursive structures, we compared two spoken registers featuring public and non-public conversation by measuring the depth of self-embedding in C, V, and N projections. The findings confirm the hypothesis that speakers use more or less self-embedding in speech production depending on the familiarity of the speech situation (public vs. non-public). In addition, we examined previous assumptions about the differences between right, left, and center embedding in C projections. The results confirm a preference against center embedding in non-public texts, which reflects the complexity of these structures and the decline of complexity in non-public spoken data. Finally, we find evidence that the depth of self-embedding between different projections is correlated across speakers. This suggests that self-embedding is an entity of grammar which is relevant for the speaker's reality, i.e., may be selected for particular (in this case, stylistic) purposes.

Keywords: complexity, embedding, register variation, spoken language, corpus study

1   Introduction

The understanding of self-embedding and its repercussions for phenomena of language use figure prominently in several fields of linguistic research. Thus, the role of recursive embedding as a core property of the language faculty has been a central issue in theoretical syntax since the early days of generative linguistics (see Hauser et al. 2002; van der Hulst ed. 2010; Arsenijević & Hinzen 2012; Trotzke & Bayer eds. 2015). The claims about the role of recursion open up questions about the cross-linguistic variation concerning the potential of embedding (cf. Everett 2005; Nevins et al. 2009), the role of cognitive limitations in processing these structures (e.g. Frazier 1985; Gibson 1998; Demberg & Keller 2008; Christiansen & MacDonald 2009; Roeper & Speas ed. 2015; etc.) as well as the acquisition of recursive structures (e.g. Roeper 2011; Pérez-Leroux et al. 2012). Beyond the debates that evolved in several areas of research, these investigations scrutinize the relation between the potential structures of natural languages and the structures that are realistically produced or processed in language use.

Focussing on language use, the influence of register on structural complexity contributes to our understanding of how speakers select certain structural properties depending on their communicative intentions. It has been observed that recursive syntax predominantly appears in written varieties but less so in spoken varieties; cf. Sakel & Stapert (2010); Kornai (2014). Karlsson (2007) notes, for instance, that repeated center embedding, which is associated with high processing load, is almost exclusively found in written language. The difference between spoken and written varieties in complexity is a recurrent issue in register research; cf. Halliday (1979); Biber (1995); Miller & Weinert (1998); Maas (2006); (2008); Biber & Gray (2010); Biber (2012) among many others. Further dimensions of register variation, most notably the contrast between formal and informal interaction, are also reflected in complexity and are sometimes orthogonal to the written-spoken distinction, i.e., formal oral varieties (e.g.

parliamentary talks) display higher levels of complexity than informal written varieties (e.g. chat communication); cf. Ágel & Hennig (2006); Koch & Oesterreicher (2007).

In the present study, we investigate structures of recursive embedding in different registers of spoken language. 'Register' is thereby conceived as the sum of linguistic behaviour in a given functional setting (Labov 1972; Biber 1995; 2008). Speakers select from an inventory of linguistic structures, notably among structures of different depth of embedding, depending on functional properties that are appropriate for the discourse situation. Why should different registers differ in terms of depth of embedding? Particular registers, e.g. public speech, are designed to have an impact on the audience, i.e., the speaker may intend to draw attention not only to what is being said but also by how it is said. Hence, complexity might be an instance of expensive signaling in the sense of Zahavi & Zahavi (1997) or an instance of the Effort Code (see Gussenhoven, 2004: 85-89 for similar phenomena in intonational phonology). Speakers increase their effort in particular registers, which leads to a greater exploitation of the potential made available by the grammar (Givón 1979; Ochs 1979).

The present article examines the influence of registers properties (notably, the distinction between private and public oral speech) on the depth of self-embedding; the conceptual prerequisites, the motivation and the hypotheses as well as the goals of this study are outlined in Section 2. The empirical study compares the complexity of public and non-public registers in spoken German (see Section 3 on the text sample and the data mining procedure). In particular, we investigated the frequencies of several levels of embedding in three types of syntactic projections, i.e. nominal, verbal and clausal (the results are presented in Sections 4 to 6, respectively). A crucial question is whether the depth of embedding correlates between projections, i.e., whether the observed findings can be reduced to an abstract choice between more or less complex structures of whatever projection; see Section 7. The reported empirical findings are discussed in Section 8; Section 9 concludes.

2 Prerequisites

2.1 Self-embedding

Self-embedding is defined with reference to two structural properties: (a) the bracketing of two constituents, i.e. the notion of embedding, and (b) the labelling of the involved constituents. As seen in the definition given in (1), 'self-embedding' refers to the output of syntactic rules. The discussion about the type of rules that derive these structures is an independent issue that is beyond the aims of this study (see Luuk & Luuk, 2011 on the derivation of self-embedding by recursion or iteration).

(1)   Self-embedding

$[_\alpha\,[_\alpha\,]]$

A constituent labelled α is self-embedded if it is dominated by another constituent with the same label α (see Miller and Chomsky 1963; Miller and Isard 1964; Gibson 1998: 5).

This definition applies to several layers of syntactic structure: subordinate clauses are embedded in higher clauses, e.g., [CP *she said* [CP *that he is sleeping*]]; nominal projections can be embedded within higher nominal projections, e.g. [DP *the book* [*on* [DP *the shelf*]]], verb projections can be embedded in higher verb projections, e.g. [VP *wants to* [VP *dance*]], etc. The definition in (1) does not require the embedded structure to be immediately dominated by a projection of the same category. In theories of syntax which assume that lexical projections have functional structure, immediate dominance of this type does never occur (see Arsenijević & Hinzen 2012).

There are three logical possibilities regarding the placement of the embedded constituent within the dominating constituent: it may be embedded either within or on the left or on the right of the material of the dominating constituent; see (2).

(2)   a.   Center embedding: $[_\alpha\,X\,[_\alpha\,Y\,]\,Z\,]$

A constituent labeled *a* is center-embedded in another constituent with the same label if some non-null element that belongs to the dominating constituent intervenes between the edges of the embedded constituent and the edges of the dominating constituent.

b.   Left embedding: [*α* [*α* X] Y]

A constituent labeled *a* is left-embedded in another constituent with the same label if the left edge of the embedded constituent immediately follows the left edge of the dominating constituent.

c.   Right embedding: [*α* X [*α* Y]]

A constituent labeled *a* is right-embedded in another constituent with the same label if the right edge of the embedded constituent immediately precedes the right edge of the dominating constituent.

Although grammars may not have discrete limits on the depth of embedding, language must be produced and processed in a finite amount of time, which entails that an infinite application of a rule will never be observed. Moreover, self-embedding in language use must cope with a realistic exploitation of processing resources, e.g. limitations in memory. Hence, it comes as no surprise that language users only exploit a small subset of the computational potential.

The intuition that center embedding is constrained in language performance can already be found in Chomsky (1965: 13): "repeated nesting contributes to unacceptability". Multiple center embedding is associated with processing load; see illustration in (3) from Gibson (1998: 4). The sentences in (3) convey the same propositional content while exhibiting different embedding structures. The center embedding of the temporal clause within the conditional clause in (3b) requires more processing effort than the structure without center embedding in (3a). An additional level of center embedding as in (3c) leads to unacceptability. Similar phenomena are reported for VP projections; see, e.g. Christiansen & MacDonald (2009) and de

Vries, Christiansen & Petersson (2011) on the processing of center embedding in Dutch and German VPs.

(3) (a) [CP [CP *If the mother gets upset* [CP *when the baby is crying*]],

the father will help], [CP *so the grandmother can rest easily*].

(b) [CP [CP *If* [CP *when the baby is crying*], *the mother gets upset*]],

the father will help], [CP *so the grandmother can rest easily*].

(c) #[CP [CP *Because* [CP *if* [CP *when the baby is crying*], *the mother gets*

upset], *the father will help*], *the grandmother can rest easily*].

(Gibson, 1998:4)

For initial clausal embedding, predominantly found in written language, corpus results in several languages (including English, German, Finnish, Latin, Swedish) display a maximal depth of two as an upper limit (Karlsson, 2007a). Furthermore, the maximal depth of center embedding reaches three levels in written and two levels in spoken English according to Karlsson (2007b). Final embedding seems to be less limited: speakers tend to restrict the depth of embedding to three in simple varieties (e.g. everyday conversation, textbooks) and five in complex varieties (e.g. written); however, examples with depth up to ten do occasionally occur (Karlsson, 2010a: 93). It is thus no surprise that most embedded clauses are finally embedded (around 80 % in corpora of spoken and newspaper texts from different languages). These tendencies are sensitive to register: some varieties such as legal language show a higher preference for initial and center embedding (final embedding down to 60 %). Similar asymmetries depending on depth and placement of embedding are reported in speech processing studies (see e.g. Miller & Chomsky, 1963; Miller & Isard, 1964; Gibson & Thomas 1996; Gibson 1998; Demberg & Keller 2008; Nakatani & Gibson 2010). Left embedding also appears to be more difficult than right embedding, presumably since the head must be anticipated in the latter case (Ueno & Polinsky 2009).

The two generalizations emerging out of the diverse studies on speech processing, corpus frequencies and acceptability are outlined in (4): lower degrees of self-embedding are more frequent and easier to process than higher ones, which directly reflects the structural complexity of these constructions; see (4a). Secondly, there is a general preference for types of embedding that do not affect the continuity of the constituents, i.e. a preference against center embedding. Whenever the syntax allows for both left and right branching, as in the case of subordinate clauses, there is a preference for the right branching option; see (4b).

(4)   Asymmetries in language use (reflected in frequencies and ease of processing)

    a.   self-embedding depth n > self-embedding depth $n+1$

    b.   right-embedding > left-embedding > center-embedding

The present study is devoted to the reflexes of the asymmetries in (4) concerning register variation. We already saw that complexity increases in written communication, which provides a flexible time window for planning and processing (Ochs 1979; Beaman 1984; Karlsson 2009; Sakel & Stapert 2010). Written registers, as e.g. literary and academic prose or newspaper texts, have been described as showing higher levels of syntactic complexity including self-embedding while spoken language is often characterized as less complex than written language (among many others Chafe & Tannen 1987; Paolillo 2000; see also the above statements on different types of clausal embeddings from Karlsson). Similarly, Miller & Weinert (1998) observe for several languages (English, Russian, German; also building on earlier work by Hawkins 1969; Sirotinina 1974; Biber 1988; a.o.) that noun phrases in written texts are generally more complex than in spontaneous spoken language.

However, differences in complexity cannot be exhaustively accounted for by the contrast between written and spoken varieties, so that reducing the choice of complexity to the advantage writing/reading has over speaking/listening in exploiting a larger time window for planning or processing is not feasible. The structural patterns that emerge within registers are

part of the register competence of the speaker: i.e., speakers consciously select more or less complex structures in order to convey social meaning depending on the discourse situation at issue. Crucially, the differences between written and spoken communication mentioned above cannot be generalized across languages. For instance, Besnier (1988) reports that written speech in Nukulaelae Tuvaluan (Polynesian) does not differ from spoken speech in complexity; Ong (1982: 37f.) claims that subordination is more frequent in languages with established literary traditions. These findings suggest that the rise of complexity is not a necessary concomitant of the time flexibility of written communication, but instead evolves with the emergence of particular registers, which may be oral or written.

Studies comparing written and spoken communication under identical discourse conditions offer a more differentiated picture: Beaman (1984) – comparing written and spoken narrations of the *pear story* – finds that the modalities differ in the types of complexity; where written narrations were lexically more dense, integrated and compact, spoken narrations showed relatively more subordinate clauses.[1] Moreover, comparisons between registers within the same modality show that complexity varies depending on further factors. Biber's register studies identified a distinction between face-to-face conversations and public conversations/spontaneous speeches on the (English) dimension of *involved vs. informational production,* relying among other features on complexity differences such as different types and forms of subordinate and embedded structures (Biber 1995; 2008). Paolillo (2000) distinguishes

---

[1] Beaman's study of the *pear film* is instructive for the distinction between (narrative) written and spoken language because it does not confound modality with other parameters (including formality, topic, purpose, etc.). She concludes that higher degrees of complexity in written (vs. spoken) language are a result of the influence of factors such as formality and purpose of registers often associated with written language (cf. Beaman 1984:79).

different spoken registers in Sinhala that vary in formality, the formal variety showing a higher complexity in terms of the coding of grammatical features than the less formal variety.

Finally, conceptualizing complexity as a conventionalized property of particular registers explains why complexity asymmetries do not apply across the board, i.e. equally for any type of syntactic projection. Biber & Gray (2010) show for English that academic writing is characterized by nominal complexity and conversation by a greater amount of clausal embedding. The same contrast is reported for German; see Neumann (2014: 77).

2.2 *Aims of the study*

The research questions of the present study are outlined in (5).

(5)   Research questions

    a.   Do speakers modulate syntactic complexity (in terms of depth of embedding) in speech production depending on register?

    b.   Does the preference for complex or less complex structures equally apply to different syntactic projections, i.e., is there evidence that a preference for self-embedding structures exists in linguistic reality independently of particular projections?

In order to approach the research question (5a), we examine register variation within the same modality, i.e., between varieties of spoken language. This comparison circumvents the risk of confounding further factors that may influence complexity, e.g., the flexibility of the time window for planning/processing in written communication; see discussion in Section 2.1. In particular, we will compare public and non-public registers of oral speech. Assuming that speakers employ a greater amount of planning in public speech than in private conversations, public speech is expected to involve reflexes of complexity, leading us to predict higher depths of self-embedding and more costly types of self-embedding (see (4)) in this variety.

The present study examines three different syntactic projections in German, namely NPs, VPs, and CPs. German NPs are predominantly right-branching. Left embedding is mainly

restricted to proper names (in use). German VPs are left-branching, though the highest verbal head (finite verb) may occur on the left of the non-finite verb as a result of head movement. This means that for NPs and VPs left vs. right embedding is not a type of variation that can be attributed directly to register. This contrasts with the direction of embedding in the CP, which is largely free. Center embedding is possible in all three types of syntactic projections, but with NPs, center embedding has further requirements: it appears with modifying adjectives or participles having further nominal dependents, i.e. a center-embedded NP is not a possible alternative linearization for any type of NP. In order to examine the research question (5a), we will first examine the frequencies of self-embedding and the location of embedding in the mentioned syntactic projections (Sections 4 to 6) followed by an examination of whether the depth of self-embedding is correlated between projections (Section 7).

3   Method

3.1   Text sample

The analyzed data stem from two corpora of spoken German provided by the *Datenbank für gesprochenes Deutsch* (DGD, both available at http://agd.ids-mannheim.de, IDS Mannheim). The first corpus, *Grundstrukturen: Freiburger Korpus* (FR), contains conversations recorded from 1960 to 1974 in public settings (e.g. panel discussions, open councils, press conferences etc.) and private situations (e.g. at home, during lunch etc.) in the Freiburg area. The data of the *Forschungs- und Lehrkorpus Gesprochenes Deutsch* (FOLK) contains recordings made between 2003 and 2016 in various German speaking areas, among them private everyday talk (at home, informal gatherings of friends, theatre breaks, etc.) and public meetings (e.g. panel discussions, open councils etc.). The investigated corpus comprises 24 conversations: 12 public, 12 non-public (1000 tokens each).

The choice of texts was based on the contextual indicators for register classifications that are available in the database (see Eggins & Slade 2005; Kunz 2010; Halliday & Matthiessen 2013; Neumann 2014). The relevant indicator for the creation of two subcorpora (public vs. non-public) is the dimension public vs. private. We chose texts with close-to-standard speakers in order to keep dialectal variation to a minimum and only considered oral and phonic communication (indicator *mode of discourse*). Medially transmitted recordings such as talk shows and telephone conversations were excluded. Both subcorpora (public, non-public) in our study share the same array of contents (indicator *field of discourse*): political and social contents with only some instances of private subjects in the non-public subcorpus. The style of the conversations includes mainly argumentative reasoning and expositive intercourse, though argumentation dominates the public texts and exposition the non-public ones. The number of speakers is a concomitant of the distinction between public and non-public (Schikorsky 1990: 34): the average number of speakers is 2.75 in the non-public subcorpus and 7.08 speakers in the public subcorpus. Finally, public and non-public texts differ with respect to the social distance between speech participants (indicator *tenor of discourse*): speakers in the non-public subcorpus tend to know each other informally, having casual or even intimate relationships, while the public subcorpus involves speakers that are not or less acquainted to each other. Thus, the situational aspects of both subcorpora differ maximally in the factors that relate to the public vs. non-public dimension, but are generally similar in all other aspects.

A crucial limitation of this type of data is that the corpus does not allow for observations of the same individual under different registers. The statistic treatment of this sample requires a model in which the random factor SPEAKER is nested within the fixed factor REGISTER.

3.2 Data mining

All texts were converted into the TCF format for compatibility across platforms. Due to a concentration on oral communication features, the segmentation in the existing annotation was

based on speaker turns and time intervals, thereby separating parts of clauses and sentences in the linearization. As the aspired syntactic analysis concentrates in part on clausal structures and because the available tools only allow annotating linearized structures, we had to revise the segmentation, thereby following the guidelines developed in the NoSta-D project for the syntactic annotation of non-standard language varieties (here in particular "Guidelines Vorverarbeitung", Reznicek 2013). In the new segmentation, each segment contains only non-overlapping utterances including a matrix clause as well as all its dependent clauses. Conjunctions of asyndetic coordinations start a new segment. In addition, we inserted a new token at the beginning and end of a segment, the former functioning as a root node necessary for the annotation and as an indicator of the respective speaker, receiving the original speaker ID as a token tag and the "_" as a lemma tag, while the latter marks the end of a segment via a full stop. Even though the blending of metadata (speaker ID) and object data was undesired, we found that the current tools did not provide a better solution without losing valuable information.

WebLicht (Hinrichs et al. 2010) provides a service environment for automatic annotation of text corpora, granting access to the MaltParser (Hall et al. 2009), a data-driven dependency parsing system, that was used to add automated dependency annotations to the texts. The parser's results were then revised to comply with the NoSta-D dependency annotation scheme for non-standard language annotations utilizing the web-based multi-layer annotation tool WebAnno (Eckart de Castilho et al. 2014). Even though the NoSta-D dependency annotation scheme meets the demands of annotating spoken texts syntactically, its being based on the syntactic TIGER annotation scheme (Albert et al. 2003) means that some distinctions cannot be found in the data (missing label distinctions), in particular concerning the various characteristics subsumed under the MOD label. Therefore, we instead chose to conform to the

TüBa-D/Z annotation scheme (Telljohann et al. 2012) on rare occasions, e.g. annotating prepositional phrases in predicative constructions as PRED instead of MOD.

For the distinction between adverbial complements and adjuncts, we strictly followed E-VALBU (Kubczak 2016), an online valency dictionary based on corpora analyses. To allow later reviews of these somewhat controversial cases, we assigned them the new label KADV. A second addition to the label set was necessary to exclude unfinished ('terminated') sentences as in *das möcht ich noch* 'I'd also like to' (FR30). Analoguous to the COR label, the new T label intersects with the appropriate label of the putative clause, resulting in TS for the example above. Therefore, whenever a finite verb is missing an obligatory complement such as a missing subject, which renders the sentence incomprehensible, the T label is used.

A script-based analysis with R (R Development Core Team 2008)[2] uses the POS-tags and the dependency annotation to retrieve the layers of embedding with R by listing the heads of the desired structures, e.g. finite verbs when looking at CPs, and relating embedded structures of the same type, provided that non-relevant dependency labels have been excluded. As there are a few labels that cannot be excluded, thus leading to some false hits, we manually checked the results. The resulting data frame provides all occurrences for each depth of embedding per projection. These were then manually annotated for type of embedding to allow for a more fine-grained analysis; see individual projections for more details.

3.3 Annotations

Extracting the count and depth of self-embedding required determining every instance of each projection type (CP, VP, NP) and checking for whether it is contained in another instance of the same projection type. To count as an instance, the projection had to be formally complete,

---

[2] The groundwork was kindly provided by Burkhard Dietterle.

which means that all cases with a T-label were excluded, i.e., an utterance with two finite verbs where the structurally lower CP is incomplete was not included in the counts as a case of embedding.

Coordinated instances of the same projection type count for the respective level as they do not augment the level of embedding. Thus, a CP embedding two paratactically joined clauses is counted as two instances of embedding, see the sentence *wir haben diese preise [die uns vorgelegt wurden] und [die wir zurückgerechnet haben] verglichen* 'we compared the costs which we were given and which we recounted' (FOLK70). The two embedded clauses are counted individually since any of the embedded instances may contain further embeddings, potentially resulting in different depths (see also Karlsson 2010a).

The annotation of depths of embedding follows the definitions in (2) (see also Karlsson 2007b, 2010a). A structure contains initial embedding when the embedded XP precedes all the elements of the projection except for coordinators, which may precede initial embedded elements. Center embedding has the embedded XP after and before elements of the embedding phrase, e.g. between determiner and noun for NPs or between finite verb and other elements for CPs. Instances count as final embedding when they are not followed by any parts of the superordinate phrase.

Matrix phrases without any embedded elements count as depth 1. The depth is increased by one if an instance is embedded in another instance, so a CP with an embedded complement clause has the depth 2, constituting one level of embedding, while a CP that embeds a complement clause which itself embeds another clause has the depth 3 with two levels of embedding. Coordination and parallel modifications do not increase the depth but count individually, introducing new non-linear strands of depth.

4 Nominal projections

4.1 Overview of the data

The present corpus study only considers lexical nominal projections, i.e. nominals containing a lexical head (excluding pronouns). We assume a DP structure as the extended projection of German nouns.[3] DP-within-DP embedding is mainly attested with two types: either the embedded DP is a genitive phrase as in (6a) or it is dominated by a prepositional projection, as in (6b). In the latter case, a few tokens (*n=5*) do not have the PP embedded immediately within the nominal projection, but it is part of an Adjective Phrase as illustrated in (6c). Finally, the corpus contains some cases in which the embedded DP is not case-governed by the head N, instead resembling a (cited) fragment; see (6d). We adopt the CP-layer as upper bound for embedding structures, i.e., we ignore DPs embedded in a CP that is itself embedded within a higher DP.

(6) a. [$_{DP1}$ *das Problem* [$_{DP2}$ *der Freizeit*]] (FR180)

the   problem   the(GEN)   freedom

'the problem of the freedom'

b. [$_{DP1}$ *die Zeit* [*nach* [$_{DP2}$ *m Krieg*]]] (FOLK220)

---

[3] The difference between theories of DP and NP structure is not crucial, since our assumptions about embedding relate to the bracketing and not to the labeling of these structures. In terms of the DP analysis, the genitive phrase is located in the specifier position of the DP, i.e. within a functional layer that expands the projection of a lexical N head: [$_{DP}$ DP$_{gen}$ [$_{D'}$ D$\varnothing$ [$_{NP}$ …N°…]]] (e.g. Haider 1988). In terms of an NP account, the genitive NP is embedded within the specifier position of the NP, which is again contained by the extended projection of the head N: [$_{NP}$ NP$_{gen}$ [$_{N'}$ …N°…]] (cf. Vater 1986; for a discussion of further accounts see Machicao y Priemer 2017, sect. 4.6). Our notion of embedding is compatible with any theory that assumes that the genitive phrase is contained by the extended projection of the head noun.

        the   time   after   the(DAT)war

        'the time after the war'

c.   [$_{DP1}$ *das* [[*von*   [$_{DP2}$ *Herrn Böhme*]] *genannte*] *Problem*] (FR199)

    the   by   Mr.   Böhme   mentioned problem

    'the problem mentioned by Mr. Böhme'

d.   [$_{DP1}$ *die Frage*   [$_{DP2}$ *Schuttablade*]] (FR205)

    the   question   dumping

    'the question of dumping'

A part of the genitive vs. PP alternation in German depends on register: apart from the morphological conditions that enhance the selection of a PP (relating to the loss of overt marking of the genitive; Smith 2003), embedded PPs are more frequent in colloquial styles than in formal and written language (see Scott, 2014). Indeed, embedded genitive phrases are the most frequent option in the public part of our corpus (54.4%), while their relative frequency is lower in the non-public texts (39.5%); see Table 1.

In order to estimate the choice between functionally equivalent expressions, we should consider those PPs that can be replaced by a genitive. The relevant subset are *von*-phrases with a possessor role: we found 17 such phrases out of 52 PPs in the non-public subcorpus and 21 out of 125 in the public subcorpus. The ratio of genitives and possessor *von*-phrases is 34/17=2 in non-public texts, and 156/21=7.4 in public texts.

Table 1. Types of nominal embedding.

	non-public		public		total	
	n	%	*n*	%	*n*	%
genitive DP	*34*	39.5	*156*	54.4	*190*	50.9
PP	*52*	60.5	*125*	43.6	*177*	47.5
not governed	*0*	0.0	*6*	2.1	*6*	1.6
Total	86	100	287	100	373	100

Nominal projections are right-branching in German: complement and adjunct DPs or PPs follow the N°; see (6). Right embedding applies to 97.6% of the analyzed DPs; see Table 2. Center embedding only appears with adjective phrases in the corpus (*n*=5); see (6c). Finally, left embedding occurs in the case of genitive DPs occupying the specifier position of the DP (*n*=4); see DP_2 in (7a) and DP_3 in (7b). Left embedded genitive phrases within left embedded genitive phrases are not attested in our corpus (although this possibility is grammatical in German, see e.g. Haider 1988).

(7) a. $[_{DP1} [_{DP2} Frischs]$ *Meinung*] (FR212)

Frisch(GEN)   opinion

'Frisch's opinion'

b. $[_{DP1} eltern$ *von* $[_{DP2} [_{DP3} gustafs]$ *klasse*]] (FOLK201)

parents   of   Gustaf(GEN)   class

'parents of Gustaf's class'

Table 2. Direction of branching in nominal projections.

	non-public		public		total	
	n	%	*n*	%	*n*	%
[DP1 [DP2 …] …]	*2*	2.3	*2*	0.7	*4*	1.0
[DP1 …[DP2 …]]	*84*	97.7	*280*	97.7	*364*	97.6
[DP1 …[DP2 …] …]	*0*	0.0	*5*	1.7	*5*	1.0
Total	86	100	*287*	100	*373*	100

4.2 Depth of embedding

The examined (public, non-public) sample contains 2914 simple or complex DPs; see exact counts in Appendix, Table 10. The majority (2580; 88.5%) are simple DPs (depth = 1) not containing an embedded DP. The remaining DPs are complex, involving up to three degrees of embedding: 297 cases with a single embedded DP as in (8a) (depth = 2), 35 instances of embedded DPs within embedded DPs as in (8b) (depth = 3), and finally 2 instances of threefold embedding as in (8c) (depth = 4). In total, this dataset contains ($297\times1+35\times2+2\times3=$) 373 embedded DPs (independently of depth).

(8)   a.   depth = 2

[DP1 *dem*   *Eindruck*   [DP2 *des*   *Zusammenbruchs*]]

the(DAT)   impression   the(GEN)   collapse(GEN)

'the impression of collapse' (FR196)

   b.   depth = 3

[DP1 *im*   *Interesse*   [DP2 *der*   *Zukunft*   [DP3 *Europas*]]]

in.the(DAT)   interest   the(GEN)   future   Europe(GEN)

'in the interest of Europe's future' (FOLK126)

c.   depth = 4

[DP1 *diese*   *gute*   *Art*   [DP2 *der* *Mitbeteiligung* ... [DP3 *an* *der*   *Ordnung*

this   good way the(GEN)   participation   at   the(DAT)   order

[DP4 *des*   *Gottesdienstes*]]]]

the(GEN)   church_service

'this good way of participation at the order of the church service' (FR199)

The influence of register on the depth of embedding is presented in Figure 1; see counts in Appendix, Table 10. Embedding in DP projections is more frequently attested in public registers; see Figure 1a. The density plot (Figure 1b) is based on the mean-depth values of each speaker separately in each particular text of the corpus. Most speakers in the non-public sample have a mean depth that is very close to 1, i.e., these texts have almost no embedded DPs at all. The density of the public data reveals a larger spread, which indicates greater variability.
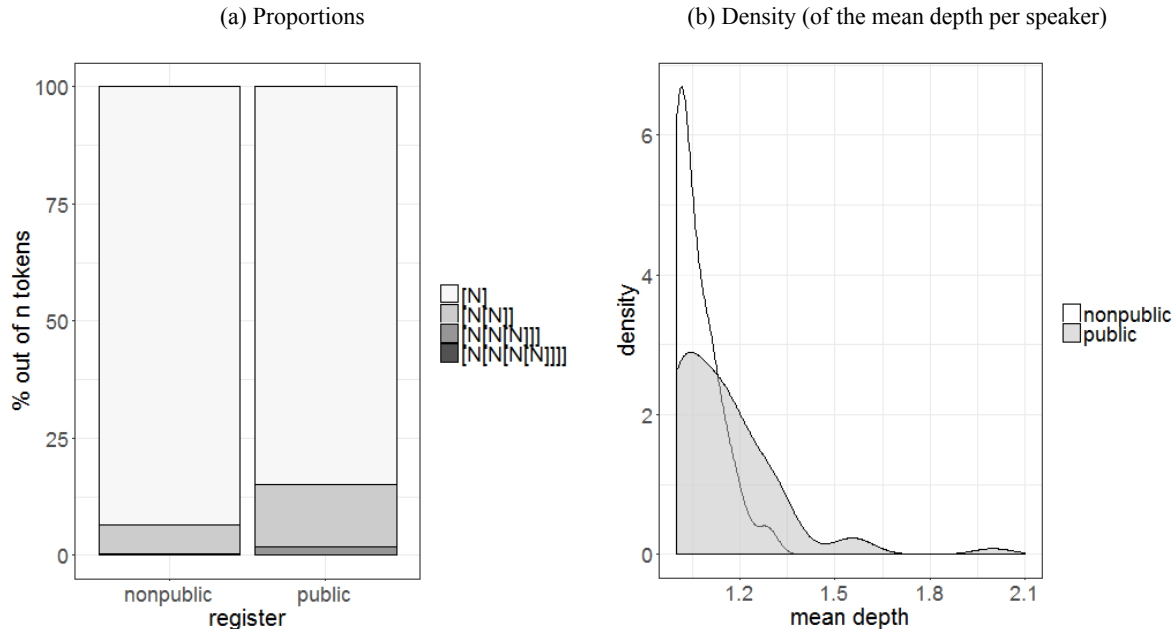

(a) Proportions   (b) Density (of the mean depth per speaker)

Figure 1. Depth of embedding nominal projections.

In order to test the impact of register on the DEPTH of DP-embedding, we fitted a generalized mixed-effects model on the data. The fixed factor of interest is the binary factor REGISTER (public vs. non-public). The dependent variable DEPTH ranges between 1 and 4. The variation

that is due to the different SPEAKERS is captured as a random factor in this model. The parameters of the model of maximal fit are given in Table 3. The model with the factor REGISTER has a better fit (AIC= 6331) than the corresponding model without this factor (AIC = 6336). A Log-Likelihood Test reveals a significant difference; $\chi^2(1) = 6.9, p < .01$.

Table 3. Generalized linear mixed-effects model on the depth of N-projections (Poisson distribution; random factor: speaker).

fixed factor	β	SE	z	p
INTERCEPT	.07	.03	2.4	< .05
REGISTER	.09	.04	2.6	< .01

5   Verbal projections

5.1   *Overview of the data*

From a morphological perspective, all elements bearing verbal inflection are verbs, hence verbs comprise lexical as well as functional verbs, i.e., auxiliaries (e.g. the perfect auxiliaries *haben* 'have' and *sein* 'be' or the future auxiliary *werden* 'will') and modal verbs (e.g. *wollen* 'want'*, dürfen* 'may'). Assuming functional projections, these elements are heads of different projections: a lexical verb is generated as the head of a VP whereas functional verbs are heads of functional projections, such as TP (=Tense Phrase), ModP (=Mood Phrase) and AspP (=Aspect Phrase). For German, the syntactic evidence for this distinction is particularly controversial and some accounts consider all these types of verbs as projecting VP structures (see Sternefeld 2006: 507ff.). In order to understand the behavior of verbal clusters, it is useful to conflate the different categories of verbs assuming that they create projections of the same type embedded in each other; see (9) (see previous analyses of verb clusters in this vein; Schmidt & Vogel 2004; Salzmann 2013).

(9)   *daß*   [VP1 *die Schauspieler*   [VP2 *das nur gemimt*]   *haben*] (FR106)

that   the   actors   this   only   mimic(PTCP)   have

'that the actors only mimed that'

The frequencies of functional and lexical verbs in our dataset are reported in Table 4. Simple lexical verbs without any functional verb occur more frequently in non-public (68%) than in public texts (57.1%); see Appendix, Table 11. Combinations of more than one functional verb – as illustrated in (10) – occur 19 times in non-public texts and (36+1=) 37 times in public texts; see Table 4. Furthermore, our dataset contains tokens with more than one lexical verb: 30 (out of 1449+30=1479) in non-public texts and 63 (out of 1320+63=1383) in public texts; see example (11) below.

(10)   *wo*   [VP1 [VP2 *wir*   [VP3 *uns irgendwie bewerben*] *wollen*]   *würden*]. (FOLK44)

where   we   us   somehow   apply(INF)   can(INF)   would

'where we could somehow apply'

Most tokens in our corpus contain verb clusters of up to three (either lexical or functional) verbs. Clusters of more than three verbs rarely occur in the corpus of spoken data (1 token with 2 functional and 2 lexical verbs and 1 token with 3 functional and 1 lexical verb in the public texts); see (14) below and counts in Table 4.

Table 4. Frequencies of lexical and functional verb combinations (Grand total: n=2862).

		non-public				public				total			
n of lexical V		1		2		1		2		1		2	
		n	*%*	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*
n of functional V	0	1005	69.4	25	83.3	790	59.8	41	65.1	1795	64.8	66	71.0
	1	425	29.3	5	16.7	493	37.3	21	33.3	918	33.2	26	28.0
	2	19	1.3	0	0	36	2.7	1	1,6	55	2.0	1	1.1
	3	0	0	0	0	1	0.1	0	0	1	0.1	0	0
	total	1449	100	30	100	1320	100	63	100	2769	100	93	100

German verb clusters generally follow the linearization patterns of V-final languages. Embedded VPs are projected on the left side of the corresponding verbal heads; see (9)-(10). In main clauses, the finite verb is fronted to the head position of the CP projection rendering a verb-second linearization (V°-to-C° movement; Thiersch 1978; Den Besten 1989); see (11).[4] Further instances of V°-to-C° movement appear in questions and in subordinate clauses without a subordinating conjunction; see (16a) below.

(11) *man*   *hat*ᵢ [VP1 [VP2 *schon* *pferde* [VP3 *vor*   *der*

     somebody has   already horses   in_front_of the

     apotheke *kotzen*]   *sehen*]   tᵢ]. (FOLK69)

     farmacy   vomit(INF) see(INF)

     'somebody has already seen horses vomiting in front of the pharmacy'

A particular linearization appears in constructions involving a perfect auxiliary and a modal verb. The perfect auxiliary is fronted to a position immediately preceding the verb cluster, while the modal verb appears in the infinitival form (and not as a participle, as otherwise in perfect

---

[4] Verb-fronting is indicated by the index *i* at the finite verb and its trace at the base position.

tense); see (12), see discussion in Sternefeld (2006: 644-664). This type of cluster creates cross-dependencies: V₁ (*hätten*) intervenes between V₃ (*wissen*) and its argument (*das*). This construction is attested six times in our dataset.

(12)   *so daß*   [VP1 *die* [VP2 [VP3 *das hätten*ᵢ   *wissen*]   *können*]   tᵢ].

      so that   they   it would_have   know(INF)   can(INF)

      '…so that they could have known that.'

Next to cases with bare infinitives illustrated in the preceding examples, also cases with *zu* 'to' infinitives are included. These generally involve extraposition to the right; see VP3 in (13).

(13)   *ich   möchte*ᵢ *nur* [VP1 [VP2   *hinterher   die Freiheit* tVP3   *haben*] tᵢ]

      I   like   only   afterwards   the freedom   have

      [VP3 [VP4 *unter   Punkt eins   etwas   sagen*]   *zu dürfen*]. (FR213)

            under   point one   something   say(INF)   to may(INF)

      'I just would like to have the freedom afterwards to be able to say something under point one.'

VPs are embedded on the left side of the verbal head in German. Finite verbs in main clauses are fronted to an earlier position, which results in a linearization in which the embedded VP follows the finite head. Under the assumption of V-fronting (precisely, V°-to-C° movement in terms of Thiersch 1978; Den Besten 1989), the constituent structure of these sentences also involves leftwards embedded VPs; see (11). However, in order to assess potential left-right asymmetries in the linearization, as stated in (4b), we should inspect the corresponding frequencies of different levels of embedding with final vs. fronted finite verbs; Table 5. These frequencies reveal that the likelihood of embedded structures is very similar with final and fronted finite verbs: in non-public texts, embedded structures are found in (30.3+1=) 31.3% of the clauses with final verbs and in (30.4+1.8=) 32.2% of the clauses with fronted verbs; in public texts, embedded structures appear in (37.7+4.4=) 42.1% of the tokens with finite verbs

and (39.0+4.1+0.2=) 43.3% of the tokens with fronted verbs. There is a difference between registers (which is dealt with in 5.2), but the position of the finite verb does not seem to affect the depth of embedding.

Table 5. Position of the finite verb in V-projections.

| | | non-public | | | | public | | | | total | | | |
| | | final | | fronted | | final | | fronted | | final | | fronted | |
finite V		n	%	n	%	n	%	n	%	n	%	n	%
	1	265	68.7	740	67.7	278	57.9	512	56.7	543	62.7	1252	62.7
n of verbal heads	2	117	30.3	333	30.4	181	37.7	352	39.0	298	34.4	685	34.3
	3	4	1.0	20	1.8	21	4.4	37	4.1	25	2.	57	2.9
	4	0	0	0	0	0	0	2	0.2	0	0.0	2	0.1
	total	386	100	1093	100	480	100	903	100	866	100	1996	100

5.2   Depth of embedding

The depth of embedding relates to all (finite and non-finite) verbal heads, comprising lexical and functional verbs. Non-verbal predicates are excluded from the analysis, e.g. … *haben die Verpflichtung, die Kinder zu erziehen* '…have the obligation to educate the children'. We adopt the CP-layer as upper bound for embedding structures, i.e., we ignore VPs embedded in a CP that is itself embedded within a higher VP. The examined (public, non-public) sample contains 2862 (simple or complex) VPs in total; see Appendix, Table 11. Beyond VPs without embedding structures (*n*=1795, 62.7%; depth = 1), the corpus contains 983 (34.3%) VPs with a single embedded VP as shown in (9) (depth = 2), 82 (2.9%) VPs with a twofold embedding as in (11) (depth = 3), and 2 instances of threefold embedding, as in (14) and (13) above (depth = 4).

(14)   *damit   würden*$_i$ … [$_{VP1}$ [$_{VP2}$ [$_{VP3}$ [$_{VP4}$ *Gegengründe …   mobilisiert*]

thereby   would   counter-arguments   activated

werden]   *können*] t$_i$] (FR213)

be   can(INF)

'thereby counter-arguments could be activated'

The complexity of V-projections differs between public and non-public texts; see Figure 2.
Complex VPs (i.e. VPs with more than one verbal head) constitute 32% of the non-public and
42.9% of the public data; see Appendix, Table 11. The density plot in Figure 2. (right panel)
shows that most speakers of the public subcorpus have an average depth of 1.47 in verbal
structures while the average embedding per speaker is lower in the non-public register (1.34).
Similar to the nominal projections (see Section 4.2), the density plot of public texts reveals a
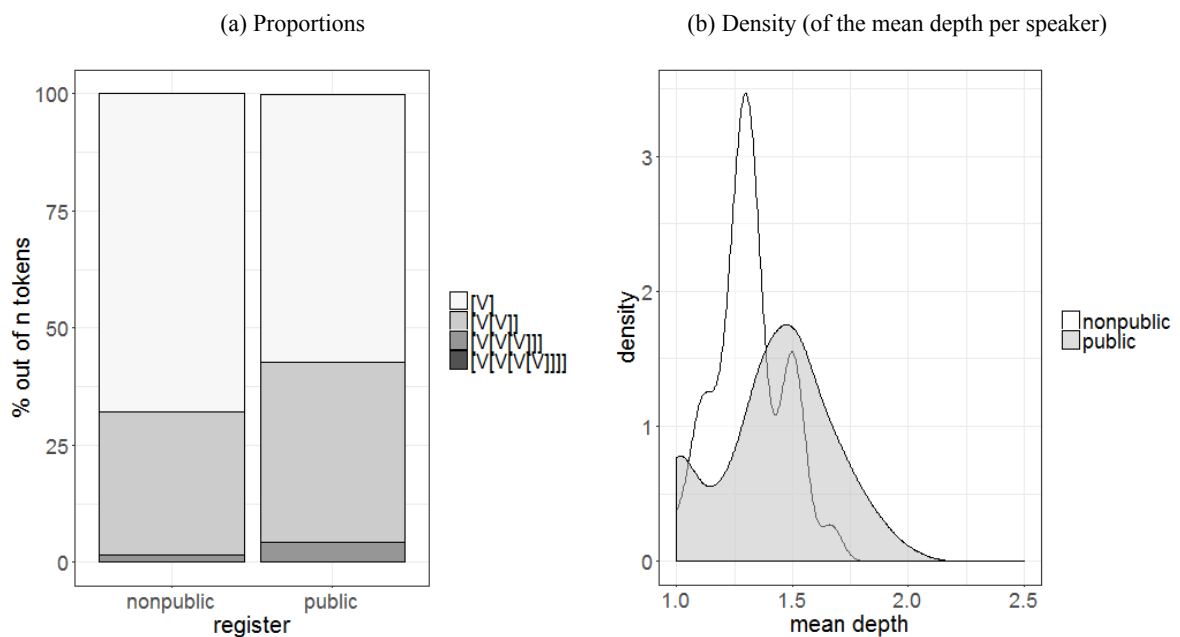greater variability (reflected in the larger spread of the graph).

Figure 2. Depth of embedding verbal projections.

A Log-Likelihood Test reveals that the effect of REGISTER on the depth of embedding in
V-projections is significant; $\chi^2(1) = 9.5$, $p < .01$. Including REGISTER into the model results in
a better model fit (model with REGISTER: AIC= 6977; model without REGISTER: AIC = 6985).
The parameters of the model of maximal fit are given in Table 6, which confirms that the ratio

of the effect (β) and its standard error (SE) corresponds to a significant p-value in the z-distribution.

Table 6. Generalized linear mixed-effects model on the depth of V-projections (Poisson distribution; random factor: speaker).

fixed factor	β	SE	z	p
INTERCEPT	.29	.02	12.9	< .001
REGISTER	.1	.03	3.1	< .01

6 Clausal projections

6.1 Overview of the data

Embedded clauses come in three types: (a) complement clauses ($n = 291+319=610$), (b) adverbial clauses ($n = 113+171=284$), and (c) relative clauses ($n = 68+121=189$); see counts in Table 7 (sums of 'non-public' and 'public' in column 1). The registers differ according to clause type: the relative frequency of complement clauses increases in non-public registers while the relative frequency of both adverbial and relative clauses correspondingly decreases. This difference across registers has also been observed by Biber & Gray (2010) (when comparing written and oral speech) and may be due to the frequent use of saying/thinking verbs in more spontaneous types of communication.

A subordinate clause may be further embedded in a clause of the same type, as illustrated in the examples in (15): in (15a) a complement clause CP_3 is embedded in another complement clause; in (15b) a relative clause CP_3 is embedded in another relative clause, in (15c) an adverbial clause CP_3 is embedded in another adverbial clause.

(15)   a.   [$_{CP1}$ *ja,   nun, ich   wollte vorhin   schon einmal sagen …*

          yes   well I   wanted earlier yet   once   say(INF)

          [$_{CP2}$ *daß   vor   einmal die Verwaltung   angesprochen worden ist …*

that   earlier   once   the administration   addressed   being   was

[CP3 *daß sie den Problemen   nicht gewachsen ist.*]]] (FR205)

that   she   the problems(DAT)   not   up   is

'Yes, well, I wanted to say earlier that the administration has been told before

…, that it is not up to the problems.'

b.   [CP1 *wer haftet   gegenüber der   stadt*

who guarantees vis-à-vis   the   city

[CP2 *die   diese   minare mineralbäder betreibt*

who those   [….]   mineral.bath runs

[CP3 *die hier auch n n wirtschaftlichen erfolg   damit   erzielten…*]]]

who here too a a economic   success therewith achieved …

'Who is liable to the city that runs those mineral baths, which also achieved

economic success with it…' (FOLK69)

c.   [CP1 *der Pfarrer zum Beispiel   muß*

the   priest   for   example   must

[CP2 *wenn er die   Taufe   eines Kindes nicht vollziehen möchte*

if   he the   baptism of.a child   not perform   wants

[CP3 *weil   die   Eltern sich   absolut   antikirchlich zeigen ...*]]

because the   parents themselves absolutely anti-church   show

das   tun](FR199)

that   do

'the priest, for example, has to do it if he does not want to perform the baptism

of a child because the parents appear absolutely anti-church ...'

Self-embedding the clause of a certain type into a clause of the same type is less likely than

self-embedding in clauses of different types, as shown in Table 7. The grey cells in the columns

3-8 highlight the cases in which superordinate and subordinate clause are of the same type. These percentages are generally lower than the percentages that embeddings of a given clause type have in the entire corpus (Column 2). The only exception are relative clauses within relative clauses in the public texts (20% vs. 19.8% overall), although that only relates to a small number of observations ($n = 10$).

Table 7. Frequencies of embedded clause types.

		all		within adverbial		within complement		within relative	
		n	$\%$	n	$\%$	n	$\%$	n	$\%$
non-public	adverbial	113	23.9	4	16.7	17	23.3	1	33.3
	complement	291	61.7	16	66.7	36	49.3	2	66.7
	relative	68	14.4	4	16.7	20	27.4	0	0.0
	Total	472	100.0	24	100.0	73	100.0	3	100.0
public	adverbial	171	28.0	7	23.3	38	34.5	5	50.0
	complement	319	52.2	18	60.0	38	34.5	3	30.0
	relative	121	19.8	5	16.7	34	30.9	2	20.0
	Total	611	100.0	30	100.0	110	100.0	10	100.0

There are three possibilities with respect to the location of CP-embedding. The most frequent case is final embedding, i.e., a CP occurs at the right side of another CP; see (15a-b). Final embedding as a means to postpone heavy components is a common strategy in both registers, but non-public conversations rely on it more, probably because of its advantages in processing (Wasow 1997: 94); see counts in Table 8. Alternatively, a CP may be embedded at the left side of other CPs: see CP_2 in CP_1 in (16a). Beyond the (right/left) peripheral options, CPs may be center-embedded within other CPs, as CP_3 in CP_2 in (16b). Center embedding has a disadvantage in terms of processing difficulty (Gibson 1998) and is avoided in those registers that avoid structural complexity (Karlsson 2007). The frequencies in the corpus confirm the

influence of register in spoken data: the percentages of center embedding are lower in non-public texts than in public texts; see Table 8.

(16)   a.   [CP1 [CP2 *ließen*   *wir jetzt das Gewehr fallen* …]

                let(SBJV) we now the gun   fall

                dann hätten   *wir überhaupt nichts zu diskutieren*] (FR16)

                then would.have we at.all   nothing to discuss(INF)

                'If we were to drop the gun now, we would have nothing to discuss at all'

    b.   [CP1 *dann weiß isch nischt* [CP2 *ob   man leute*

                then know I not   whether one people

                [CP3 *die gegen die landesverfassung verstoßen*]

                who against the constitution   violate

                noch wählen kann]]. (FOLK 126)

                still elect   can

                'then I do not know whether one can still vote for people who violate the constitution.'

Table 8. Branching of C-projections.

	non-public		public		total	
	n	*%*	*n*	*%*	*n*	*%*
[CP1 [CP2 …] …]	*39*	8.3	*50*	8.2	*97*	8.9
[CP1 …[CP2 …]]	*408*	86.4	*503*	82.3	*916*	83.6
[CP1 …[CP2 …] …]	*25*	5.3	*58*	9.5	*83*	7.6
Total	472	100	611	100	1096	100

6.2 Depth of embedding

In order to estimate the limits of complexity in C-projections, we considered all types of subordinate clauses (relative clauses, complement clauses, adverbial clauses). This includes clauses with subordinating conjunctions and subordinate clauses with verb-first structure; see CP2 in (16a). Furthermore, our counts considered root clauses embedded in verbs of saying, e.g., CP2 in (17a). We only counted clausal constituents with a C-layer, which excludes lower clausal constituents, in particular constituents with non-finite verbs (e.g. infinitival clauses introduced with *um zu* 'in order to' as in *er ging in die Küche, um sich ein Brot zu machen* 'he went to the kitchen in order to prepare a sandwich').

   The examples that reached the highest depths of embedding are illustrated in (17). In German root clauses, the C-head is occupied by the finite verb while in subordinate clauses, the C-head is occupied by the subordinating conjunction. Embedded root clauses as in (17a)/CP2 do not contain a subordinating conjunction, but a fronted finite verb. In spontaneous speech, we also find cases with a *dass*-clause and verb second, as in (16b)/CP3.

(17) a. depth = 6

[CP1 *ich mein* [CP2 *das fand ich jetz ganz intressant äh*

I think that found I now quite interesting uhm

[CP3 … *wie diese ärzte beschrieben haben äh*

how these doctors described have uhm

[CP4 *dass die sich des gesicht genau angucken*

that they themselves the face closely look.at

un dann genau gucken

and then closely look

[CP5 *wo muss was rein*

where must what in

[CP6 *damit   das   sich   hebt und   so*]]]]]] (FOLK80)

     so.that   that   itself   lifts   and   so

'I'm just saying I found it quite interesting uhm how these doctors described that they will look closely at the face and then they determine where to put what so that it lifts and so'

b.   depth = 7

[CP1 *ich   will   noch mal   darauf … hinweisen* [CP2 *dass man   sehen muss*

     I   want still once at …   point.out   that   one   see   must

[CP3 *dass* [CP4 *selbscht … we   man in stufen   vorgeht*]

     that   even …   if   one in steps   proceeds

muss man   vorher   sagen

must one   before   say

[CP4 *ob man alle stufen   realisieren will*]

     if one   all steps   realize   wants

[CP4 *weil   man dann entlang dieser   stufen   ein*

     because one   then along   of.these steps   a

gesamtplanungsverfahren …   machen muss

process.of.overall.planning   do(INF) must

[CP5 *weil   eben die   frag   isch* [CP6 *wenn die   stufe kommt*

     because …   the   question is   if   the   step comes

[CP7 *in dem   ich ganz   plötzlich*

     in which   I   entirely sudden

eben neue bauwerke   mache …]]]]]]] (FOLK 68)

    …   new bildings   construct

'I would like to highlight again that one has to see that even if one proceeds in steps one has to say beforehand if one wants to realize all steps because one has to do a process of overall planning because the question arises when the step comes, in which I suddenly construct new buildings …'

The examined (public, non-public) sample contains 2010 (simple or complex) CPs in total; see counts in the Appendix, Table 12. Beyond CPs without embedding structures (depth = 1), the corpus contains 595 (29.5%) CPs with a single embedded CP as shown in (16a) (depth = 2), 171 (8.5%) CPs with twofold embedding as in (16b) (depth = 3), and 45 (2.2%) instances of threefold embedding (depth = 4). Furthermore, the corpus contains one instance of fivefold embedding (depth = 6; see (17a)) and one of sixfold embedding (depth = 7; see (17b)).

Figure 3 demonstrates that the embedding depth of the C-projections differs between public and non-public texts. Complex CPs (i.e., CPs containing one or more embedded CPs) represent 32,8% of the non-public and 49.7% of the public data; see Appendix, Table 12. The density plot in Figure 3 (right panel) shows that most texts of the public subcorpus have an average depth 1.67 in clausal structures; the average embedding per text in the non-public register is lower (1.43). Similar to the nominal and verbal projections, the density plot for the public texts reveals a greater variability (reflected in the larger spread of the graph).
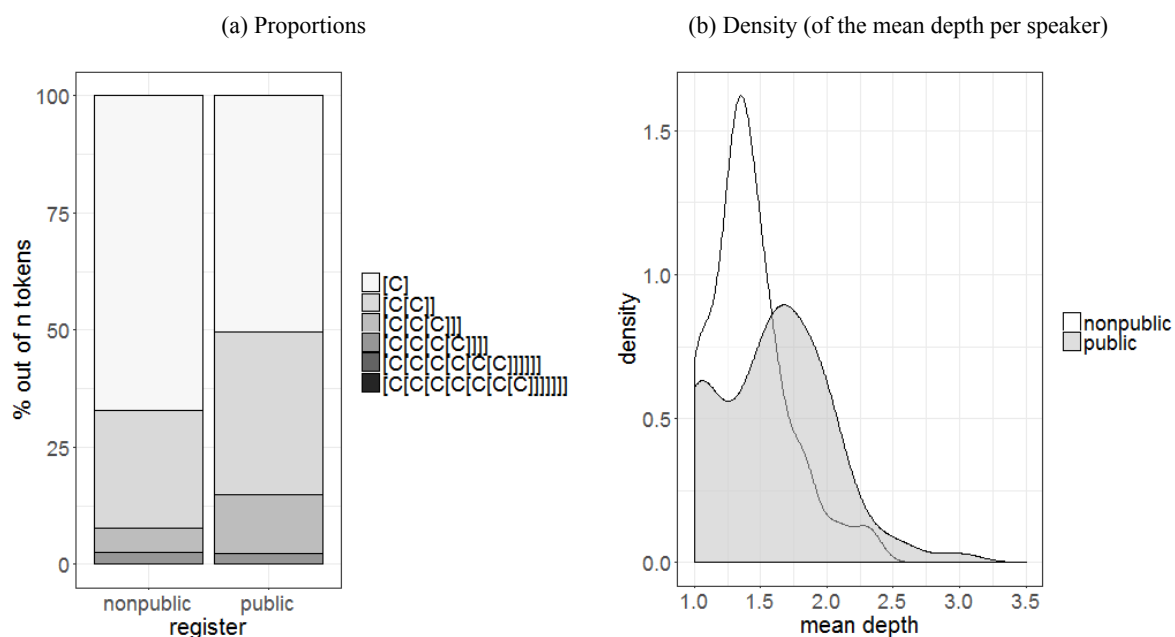
(a) Proportions   (b) Density (of the mean depth per speaker)

Figure 3. Depth of embedding clausal projections.

The parameters of the generalized linear mixed model (Poisson distribution) are listed in Table 9. The model with the factor REGISTER has a better fit (AIC= 5261) than the corresponding model without this factor (AIC = 5274). A Log-Likelihood Test reveals that the effect of REGISTER on the depth of embedding in C-projections is significant; $\chi^2(1) = 15.1$, $p < .001$.

Table 9. Generalized linear mixed-effects model on the depth of C-projections (Poisson distribution; random factor: speaker).

fixed factor	β	SE	z	p
INTERCEPT	.36	.03	14.1	$< .001$
REGISTER	.15	.04	4.3	$< .001$

7 Between projections

The corpus data confirm that register variation has an impact on the depth of embedding in three different types of syntactic projections: nominal, verbal and clausal (see Sections 4-6). A further question is whether the variation of these projections is correlated, i.e., do speakers at some level of speech planning opt for more or less complex syntactic configurations, which is

then reflected in all types of projections? To confirm this possibility the observed depths of embedding in different syntactic projections would have to correlate within the textual units of our sample. These units reflect the speech production of individual speakers in public and non-public registers, whereby individual variation is nested within register variation (since there is no data of the same speaker in different registers).

The results per speaker are plotted in Figure 4. The dots present the average depths of embedding in the texts of different individuals (white dots = non-public texts; grey dots = public texts). Descriptively, the permutations of syntactic projections (C-to-V, V-to-N, N-to-C) correlate positively, i.e., an increase of the average depth of any projection correlates with an increase in the average depth of a different type of projection. Pearson's product-moment correlation r indicates that the correlations between CPs and DPs and between VPs and DPs are very weak (close to zero). The results of the linear regressions (with the projection in the x-axis as a predictor and the projection in the y-axis as a dependent variable) show the coefficients of the regression line in the graph (intercept and slope) and reveal that the only slope that is associated with a significant value is found in the correlation between CPs and VPs. Hence, there is evidence that the average depth of embedding of verbal and clausal projections depend on each other (only), while there is no corresponding evidence from their correlation with nominal projections.

(a) CP and VP	(b) CP and DP	(c) VP and DP
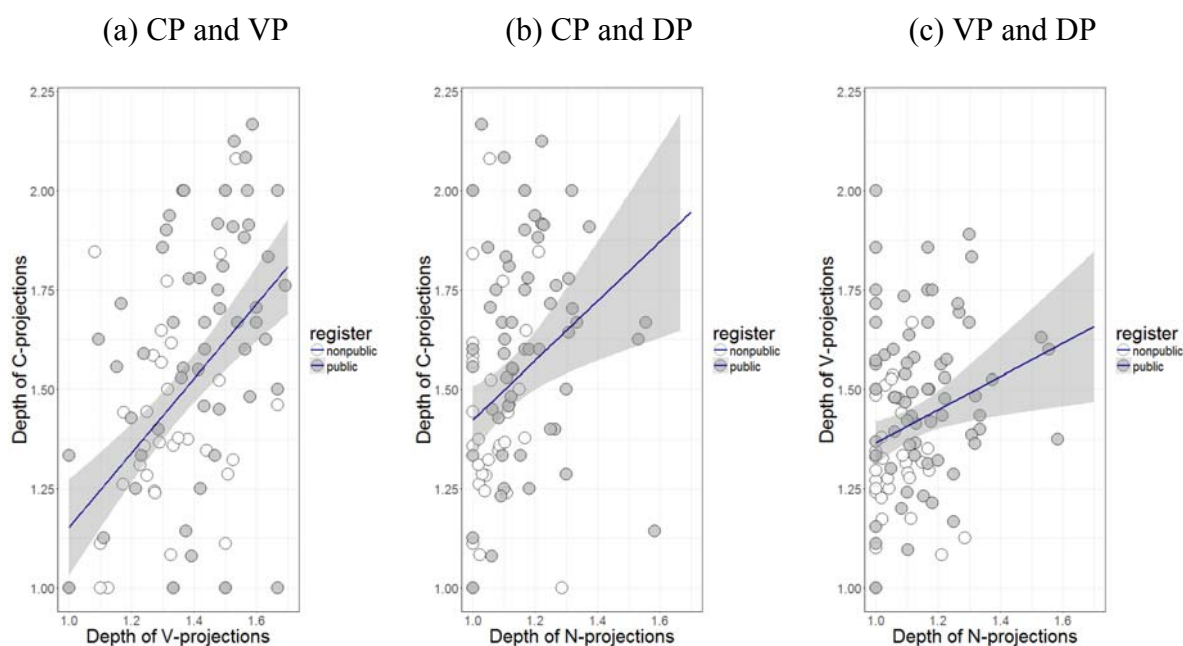
Pearson's product-moment correlation

$r = .31$	$r = .07$	$r = .09$

Linear regression

intercept = 1.12	*intercept* = 1.08	*intercept* = 1.25
slope = .18, *t* = 3.5, *p* < .001	*slope* = .03, *t* = .8, *p* = .4	*slope* = .14, *t* = 1.1, *p* = .3

Figure 4. Correlation between projections.

8   Discussion

The preceding results and analyses of the corpus data have shown that register significantly influences the depth of structures of embedding in all three investigated projections, i.e. C, V, and N projections. This result confirms the expectation that speakers expend more effort in public registers compared to non-public registers, which manifests itself in the use of more complex structures in the sense of (4a) in the former compared to the latter (see Section 2). The variation in left, right and center embedding in (4b) shows a more differentiated picture, which is discussed in the following separately for each syntactic projection.

The corpus results indicate that the depth of DP self-embedding is influenced by register (see Table 3). The influence of register is clearly manifested in the frequency of structures that

involve more than a single instance of embedding (see Appendix, Table 10): such structures are attested in 5 out of 1252 tokens in non-public texts and in (30+2=)32 out of 1662 tokens in public texts, i.e., multiple DP embedding is ((32/1662)/(5/1252)=) 4.8 times more frequent in the public register than in the non-public register. These results comply with previous studies on register variation: see Karlsson (2010b) who reports higher levels of depth in written language than in spoken language (for English, Finnish and Swedish).

The corpus results show an overwhelming predominance of right DP embedding (see Table 2): left embedding only occurs 4 times (in both registers) and center embedding 5 times in the public register. This asymmetry is already predicted by the grammatical properties of German DPs, since left embedding is generally restricted to genitives that replace the determiner and center embedding is only possible with embedded adjective phrases. This means that right, left, and center embedding in German are not variants that can be selected for stylistic purposes. Their alternation is restricted to certain structural configurations. The predominance of right embedding is informative for the occurrence of different types of DP embedding in language use in general, but not for the possible influence of register. A similar asymmetry between left and right embedding is reported by Karlsson (2010b): while left embedded genitives are very rarely embedded in other genitives (maximal depth =3) in English written language, right embedding is less restricted, for depths of up to 9 self-embedded DPs are attested in written corpora, although embedded structures beyond the depth of 6 are rare.[5] The spoken data examined in our study does not reach the depths reported from written corpora: self-embedding

---

[5] Some authors, e.g. Karlsson (2010b), Bach et al. (1994), use a different notion of 'depth' counting the *n* of embedded projections, i.e., simple DPs have depth 0, a single embedding has depth 1, etc. The depths mentioned in the discussion are based on the definition of depth in the present study.

of DPs (i.e. without the mediation of a prepositional projection) reaches the depth of 3 and embedding of DPs with the mediation of PPs reaches the depth of 4.

Finally, our data offers empirical support to previous observations about replacing the genitive with *von*-phrases in colloquial registers. In our corpus, *von*-phrases are more frequent than genitives in non-public texts. Again, the alternation between embedded DPs and PPs cannot be reduced to a stylistic choice: e.g., PPs appear with thematic roles such as place, goal, origin, comitative, etc., as such they cannot be replaced by a genitive DP. Hence, the crucial finding in our data is not the increase of genitive DPs as such, but the enhancement of the ratio between genitives and possessor *von*-phrases in public texts (see 4.1). The asymmetry between these structures is observed in several types of data. Acquisition studies for English show that self-embedding of PPs in nominal projections is easier to acquire than self-embedding of genitive phrases (e.g. Pérez-Leroux et al. 2012). In acceptability studies, structures with multiple embedded genitives are more often rejected than those with multiple embedded PPs (Christianson & MacDonald 2009).

Embedding occurs more frequently with verbal projections; compare frequencies of embedded structures in Table 10 (DPs) and Table 11 (VPs). The significant impact of register confirms the expectation that more complex structures are more likely in public texts corresponding to (4a) (see Table 6). Next to this expected result there are two general issues to be discussed related to the upper limits of the depth of embedding and to the role of the order of the verbal heads. First, it was shown that the overall data displays a ceiling effect of embedding depth of 3 in V-projections which is independent of the specific type of verb (function verbs vs. lexical verb). This means with respect to the investigated data that speakers generally do not produce structures with more than two (bare) infinitives. This holds across register; there were only two instances with embedding depth 4 in the public part of the corpus, both of which did not show more than two bare infinitives. This result is in line with results from psycholinguistic studies

on the comprehension of several layers of verbal center embedding (Bach et al. 1984; de Vries et al. 2011), which generally demonstrate that comprehension difficulties start with embedding depth 4. Bach et al. (1984: 255ff) report that acceptability and comprehension of multiple center embedding with German verb clusters (both those involving infinitives and those involving participles) steeply decline from embedding level 3 to level 4. The spoken data in our study confirms these results: verbal center embeddings of level 4 do not occur since they are associated with considerable processing difficulty.

The order of the verbal heads is informative for the asymmetries stated in (4a). As discussed in Section 5.1, the constituent structure of German VPs always involves leftwards embedding. V-fronting in main clauses results into a linearization in which the embedded VP follows the head. Since V-fronting is determined by clause type, it is not subject to stylistic variation. Our data reveals that the position of the finite verb (final vs. verb-second) does not interfere with the depth of embedding (see Table 5). Hence, it seems to be the case that verb clusters do not involve a constraint against stacking verbal heads on the right edge of the clause. The register difference with respect to the depth of embedding does not interact with the differences in linearization.

C projections differ from nominal and verbal projections in that they exploit a higher range of embedding depths. Register has a significant influence on the preferences of embedding with these projections, too: the depth of CP embeddings is significantly higher in public speech than in non-public speech (Table 9). Beyond the register difference, the position of embedding plays a crucial role. Across registers right embedding (83.6%) is by far more frequent than left embedding (8.9%) and central embedding (7.6%), complying with the complexity scale in (4b). Center embedding occurs more often in the public register (see Table 8). This distribution reflects the assumed differences in processing complexity (Gibson 1998) and the preference for higher complexity in public registers, as outlined in Section 2. Similar distributions of final vs.

non-final embedding are reported for several languages (English, Swedish, Finnish) together with observations that the percentages of initial and center embeddings increase in more complex registers (as e.g. written legal language) (Karlsson 2010a: 95).

Karlsson's observations on embedding depth are largely confirmed by our data (see Karlsson 2007a; 2007b; 2010a): multiple final embedding occurs up to depth 4 in colloquial registers and up to depth 6 in varieties favoring complexity; multiple initial embeddings (of depth 3, i.e. two embeddings) only occur in written language; multiple center embeddings very rarely reach a depth level of 4 in written and 3 in spoken language, where it is "close to non-existent" (Karlsson 2007b: 387). In our corpus, apart from the two instances of embedding depth 6 and 7 (see Appendix, Table 12), the deepest embedding level with palpable frequency is level 4 occurring in both registers with similar frequency (overall 45 tokens, 2.2%). These are mainly instances of multiple final embeddings. The data contains only a few instances of initial or center embedding at any level. At depth 4 and 3, there are only two tokens of multiple initial embedding in the non-public subcorpus.

The frequencies relating to self-embedding of a given clause type lead to an important observation; see Table 7. The relative frequencies of self-embedding within clauses of the same subtype (i.e. relative clauses within relative clauses, adverbial clauses within adverbial clauses, complement clauses within complement clauses) are lower than the overall relative frequencies for embedding of the corresponding clause type. This means that there is a preference to vary the type of subordinate clause in multiple embedding (see Karlsson 2010a: 94 for a related observation with regard to multiple final clause embedding in English corpora).

Finally, Section 7 has shown that the embedding depth of the V projections correlates with that of the C projections, however, there is no sufficient evidence for a correlation of either type of projection with the nominal projections. The significant correlation of different projections (V and C) is evidence that at some level of planning speakers decide to produce less or more

complex syntactic projections, influenced by stylistic factors. Hence, the significant result indicates that self-embedding exists in linguistic reality independently of a particular projection. That this prediction is not confirmed for nominal projections requires a closer look. Embedding is generally less frequent with nominal projections in our corpus (% of embedded N-structures: 11.5, V-structures: 37.3; C-structures: 40.5), which implies that the variation between speakers with respect to N projections is less informative (note that the texts of most speakers have an average depth of embedding between 1 and 1.5; see *Figure 1*b). Nominal embeddings are rare overall in spoken texts, being instead a characteristic of written language (see Miller & Weinert 1998; Biber & Gray 2010; Neumann 2014: 77). Spoken narrations and conversation show a higher complexity in clausal embedding. There is no obvious structural reason for this difference, i.e., complexity in the nominal domain is manifested with the same type of structural operations as in the verbal and clausal domains. The differences in language use may be grounded in the different functions that are prototypically fulfilled by the different projections. With regard to the communicative goals of spoken language, complexity at the predicate level (either through embedding clauses or verbal phrases) serves the aim of elaborating on the propositional content. A structural elaboration at the NP level generally provides an identification of referents and makes the information to be processed online denser, which might be a disadvantage in oral exchange.

9   Conclusions

The present study contributes to the research on variation in syntactic complexity dependent on register, comparing public and non-public spoken data. The findings of the corpus study confirm the preference for structures of higher complexity in registers that are associated with higher social prestige, i.e. in public texts. This insight extends the generalizability of previous findings from spoken and written data: the difference in complexity cannot be reduced to the

larger time window for sentence planning and processing that is available in written communication. Differences in complexity appear also between different registers of spoken communication, which shows that structural complexity belongs to the strategies speakers select in order to manipulate the manner of speaking in accordance with the situational context.

Structural complexity was examined in three types of projections: nominal, verbal and clausal. Register had significant effects in all three syntactic projections. Furthermore, the depths of embedding in the text sample are correlated for V and C projections, which suggests that the difference between registers is part of a general strategy (for or against complexity) depending on stylistic factors, i.e., it is not restricted to a particular type of syntactic projection. Nominal projections generally show a lower depth of embedding in spoken data; the observed correlations of N projections with the V and C projections were weak and could not be confirmed statistically.

The present study examined previous assumptions about the differences between right, left and center embedding. A closer inspection of the constituent structures of nominal and verbal projections in German reveals that the related structural options are not subject to stylistic variation and hence are not influenced by register. The assumptions about the right>left asymmetry as well as the asymmetry between peripheral>center embedding can be tested in those structures in which these structural options are true variants. This applies to embedding in the C projections. Here, our data confirms a preference against center embedding in non-public texts, which reflects the complexity of these structures and the decline of complexity in non-public spoken data.

Abbreviations

DAT = dative, GEN = genitive, INF = infinitive, PTCP = participle, SBJV = subjunctive

Competing interests

The authors declare that they have no competing interests.

References

Albert, Stefanie, Jan Anderssen, Regine Bader, Stephanie Becker, Tobias Bracht, Sabine Brants, Thorsten Brants, Vera Demberg, Stefanie Dipper, Peter Eisenberg, Silvia Hansen, Hagen Hirschmann, Juliane Janitzek, Carolin Kirstein, Robert Langner, Lukas Michelbacher, Oliver Plaehn, Cordula Preis, Marcus Pußel, Marco Rower, Bettina Schrader, Anne Schwartz, George Smith & Hans Uszkoreit. 2003. *TIGER Annotationsschema*. Saarland/Stuttgart/Potsdam: Universität des Saarlandes/Universität Stuttgart/Universität Potsdam.

Arsenijević, Boban & Wolfram Hinzen. 2012. On the absence of X-within-X recursion in human grammar. *Linguistic Inquiry* 43(3). 423-440.

Bach, Emmon, Colin Brown & William Marslen-Wilson. 1986. Crossed and nested dependencies in German and Dutch: A psycholinguistic study. *Language and Cognitive Processes* 1(4). 249-262.

Beaman, Karen. 1984. Coordination and subordination revisited: Syntactic complexity in spoken and written narrative discourse. In Deborah Tannen (ed.), *Coherence in spoken and written discourse: Advances in discourse processes*, 45-80. Norwood, N.J.: ABLEX Pub. Corp.

Besnier, Niko. 1988. The linguistic relationships of spoken and written Nukulaelae registers. *Language* 64. 707-736. doi:10.2307/414565

Biber, Douglas. 1995. *Dimensions of register variation: A cross-linguistic comparison*. Cambridge: CUP. doi:10.1017/CBO9780511519871

Biber, Douglas. 1999. A register perspective on grammar and discourse: variability in the form and use of English complement clauses. *Discourse Studies* 1. 131-150.

Biber, Douglas. 2008. Multidimensional Approaches. In Anke Lüdeling & Merja Kytö (eds.), *Corpus Linguistics*, 822–855. Berlin: Mouton de Gruyter.

Biber, Douglas & Susan Conrad. 2009. *Register, genre, and style*. Cambridge: Cambridge University Press.

Biber, Douglas & Bethany Gray. 2010. Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes* 9. 2-20. doi:10.1016/j.jeap.2010.01.001

Chafe, Wallace. & Deborah Tannen. 1987. The relation between written and spoken language. *Annual Review of Anthropology* 16. 383-407.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, Mass.: MIT Press.

Christianson, Morten, H. & Maryellen C. MacDonald. 2009. A usage-based approach to recursion in sentence processing. *Language Learning* 59 Suppl. 1. 126-161.

CLARIN-D/SfS-Uni. Tübingen. 2012. *WebLicht: Web-Based Linguistic Chaining Tool*. Online. (https://weblicht.sfs.uni-tuebingen.de/) (Accessed 2016-08-17)

den Besten, Hans. 1989. *Studies in West Germanic syntax*. Amsterdam: Atlanta.

Demberg, Vera & Frank Keller. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition* 109. 193-210. doi:10.1016/j.cognition.2008.07.008

de Vries, Meinou H., Morten H. Christiansen & Karl Magnus Petersson. 2011. Learning recursion: multiple nested and crossed dependencies. *Biolinguistics* 5(1-2). 10-35.

Eckart de Castilho, Richard, Chris Biemann, Iryna Gurevych & Seid M. Yimam. 2014. WebAnno: a flexible, web-based annotation tool for CLARIN. In *Proceedings of the CLARIN Annual Conference* (CAC) 2014, Soesterberg, Netherlands.

Eggins, Suzanne & Diana Slade (eds.). 2005. *Analysing casual conversation*. London: Equinox Publishing Ltd.

Everett, Daniel L. 2005. Cultural Constraints on Grammar and Cognition in Pirahã: Another Look at the Design Features of Human Language. *Current Anthropology* 46. 621-646. doi:10.1086/431525

Frazier, Lyn. 1985. Syntactic complexity. In David R. Dowty, Lauri Kartunnen & Arnold M. Zwicky (eds.), *Natural Language Parsing*. Cambridge: Cambridge University Press. 129-189.

Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68. 1-76. doi:10.1016/S0010-0277(98)00034-1.

Gibson, Edward & James Thomas. 1996. The processing complexity of English center-embedded and self-embedded structures. In Carson T. Schütze (ed.), *Proceedings of the NELS 26 Sentence Processing Workshop, MIT Occasional Papers in Linguistics 9*, 45-71. Cambridge, MA.

Givón, Talmy. 1979. From discourse to syntax: Grammar as a processing strategy. In Talmy Givón (ed.), *Discourse and Syntax*, 81-112. New York: Academic Press.

Gussenhoven, Carlos. 2004. *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Hall, Johan, Jens Nilsson & Joakim Nivre. 2009. *Homepage of MaltParser*. (http://maltparser.org)

Halliday, Michael A. K. 1979. Differences between spoken and written language. In Glenda Page, John Elkins & Barrie O'Connor, Australian Reading Association (eds.), *Communication through Reading: Proceedings of Fourth Australian Reading Conference*, Brisbane, August, 25-27th, 1978, 37-52. Adelaide: Australian Reading Association.

Halliday, Michael A. K. & Christian Matthiessen. 2013. *Halliday's Introduction to Functional Grammar*, 4th edition. Abingdon, Oxon.: Routledge.

Haider, Hubert. 1988. Die Struktur der deutschen Nominalphrase. *Zeitschrift für Sprachwissenschaft* 7(1). 32-59.

Hauser, Marc D., Noam Chomsky & W. Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298. 1569-1579.

Hinrichs, Erhard W., Marie Hinrichs & Thomas Zastrow. 2010. WebLicht: Web-Based LRT Services for German. In *Proceedings of the ACL 2010 System Demonstrations*. 25-29.

IDS, Mannheim. 2015. *Datenbank für Gesprochenes Deutsch (DGD)*. (http://dgd.ids-mannheim.de)

Karlsson, Fred. 2007a. Constraints on multiple initial embedding of clauses. *International Journal of Corpus Linguistics* 12(1). 108-118.

Karlsson, Fred. 2007b. Constraints on multiple center-embedding of clauses. *Journal of Linguistics* 43. 365-392.

Karlsson, Fred. 2009. Origin and maintenance of clausal embedding complexity. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language complexity as an evolving variable*, 192-202. Oxford: Oxford University Press.

Karlsson, Fred. 2010a. Multiple final embedding of clauses. *International Journal of Corpus Linguistics* 15(1). 88-105.

Karlsson, Fred. 2010b. Syntactic recursion and iteration. In Harry van der Hulst (ed.), *Recursion and Human Language*, 43-67. Berlin, New York: de Gruyter.

Kornai, András. 2014. Resolving the infinitude controversy. *Journal of Logic, Language and Information* 23. 481-492. doi:10.1007/s10849-014-9203-2

Kubczak, Jacqueline. 2016. E-VALBU. *Grammis II: Grammatische Datenbanken und Informationssysteme*. (http://hypermedia.ids-mannheim.de/evalbu/index.html)

Kunz, Kerstin A. 2010. *Variation in English and German nominal coreference: A study of political essays*. Frankfurt am Main; New York: Peter Lang.

Labov, William. 1991. *Sociolinguistic patterns*, 11. print. ed, Philadelphia: Univ. of Pennsylvania Press.

Luuk, Erkki & Hendrik Luuk. 2011. The redundancy of recursion and infinity for natural language. *Cognitive Processing* 12(1). 1-11.

Machicao y Priemer, Antonio. 2017. *NP-arguments in NPs: An analysis of German and Spanish noun phrases in Head-Driven Phrase Structure Grammar*. Humboldt-Universität zu Berlin: PhD thesis.

Michaels, Sarah & James Collins. 1984. Oral discourse styles: Classroom interaction and the acquisition of literacy. In Deborah Tannen. (ed.), *Coherence in spoken and written discourse: Advances in discourse processes*. Norwood, N.J.: ABLEX Pub. Corp.

Miller, George A. & Noam Chomsky. 1963. Finitary models of language users. In Duncan Luce et al. (eds.), *Handbook of mathematical psychology*, 419-491. New York: Wiley.

Miller, George A. & Stephen Isard. 1964. Free recall of self-embedded English sentences. *Information and Control* 7. 292-303. doi:10.1016/S0019-9958(64)90310-9

Miller, Jim & Regina Weinert. 1998. *Spontaneous spoken language: Syntax and discourse*. Oxford: Oxford University Press.

Nakatani, Kentaro & Edward Gibson. 2010. An on-line study of Japanese nesting complexity. *Cognitive Science* 34. 94-112.

Neumann, Stella. 2014. *Contrastive register variation: A quantitative approach to the comparison of English and German*. Berlin: de Gruyter.

Nevins, Andrew, David Pesetsky & Cilene Rodrigues. 2009. Pirahã exceptionality: A reassessment. *Language* 85. 355-404.

Ochs, Elinor. 1979. Planned and unplanned discourse. In Talmy Givón (ed.) *Discourse and syntax*, 51-80. New York: Academic Press.

Ong, Walter J. 1982. *Orality and literacy.* London: Methuen.

Pérez-Leroux, Ana T., Anny P. Castilla-Earls, Susana Bjar & Diane Massam. 2012. Elmo's sister's ball: The problem of acquiring nominal recursion. *Language Acquisition* 19. 301-311.

Poalillo, John C. 2000. Formalizing formality: An analysis of register variation in Sinhala. *Journal of Linguistics* 36(2). 215-259.

R Development Core Team. 2008. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. (http://www.R-project.org)

Reznicek, Marc. 2013. Linguistische Annotation von Nichtstandardvarietäten – Guidelines und „Best Practices": Guidelines Vorverarbeitung, Technische Berichte. Humboldt-Universität zu Berlin, Berlin.

Reznicek, Marc & Burkhard Dietterle. 2013–2014. *NoSta-D-Annotationsschema (versch. Vers.)*. Technische Berichte: Humboldt-Universität zu Berlin.

Roeper, Thomas. 2011. The acquisition of recursion: How formalism articulates the child's path. *Biolinguistics* 5(1/2). 57-86.

Roeper, Thomas & Speas, Margaret (eds.). 2014. *Recursion: Complexity in cognition*. Cham: Springer. doi:10.1007/978-3-319-05086-7

Sakel, Jeanette & Eugenie Stapert. 2010. Pirahã – in need of recursive syntax? In Harry van der Hulst (ed.), *Recursion and human language*, 1-16. Berlin, New York: Mouton De Gruyter. doi:10.1515/9783110219258.1

Salzmann, Martin. 2013. Rule ordering and verb cluster formation: On the extraposition paradox and the placement of the infinitival particle *te*/*zu*. In Fabian Heck & Anke Assmann (eds.), *Rule interaction in grammar*, Linguistische Arbeitsberichte 90. 65-121.

Schiller, Anne, Simone Teufel, Christine Stöckert & Christine Thielen. 1999. *Guidelines für das Tagging deutscher Textkorpora mit STTS*. Technischer Bericht. Stuttgart/Tübingen: Universität Stuttgart/ Universität Tübingen.

Schikorsky, Isa. 1990. *Private Schriftlichkeit im 19. Jahrhundert: Untersuchungen zur Geschichte des alltäglichen Sprachverhaltens kleiner Leute*. Tübingen: Niemeyer.

Schmidt, Tanja & Ralf Vogel. 2004. Dialectal variation in German 3-verb clusters: A surface oriented optimality theoretic account. *Journal of Comparative Germanic Linguistics* 7. 235-274.

Scott, Alan K. 2014. *The genitive phrase in Dutch and German. A study of morphosyntactic change in codified languages*. Leiden, Boston: Brill.

Smith, George. 2003. On the distribution of the genitive attribute and its prepositional counterpart in Modern Standard German. *Proceedings of the 25th Annual Penn Linguistics Colloquium,* Volume 8(1). 173-186.

Sternefeld, Wolfgang. 2006. *Syntax: Eine morphologisch motivierte generative Beschreibung des Deutschen.* Tübingen: Stauffenburg Verlag.

Telljohann, Heike, Erhard W. Hinrichs, Sandra Kübler, Heike Zinsmeister & Kathrin Beck. 2012. *Stylebook for the Tübingen treebank of written German (TüBa-D/Z)*. Technischer Bericht: Universität Tübingen.

Thiersch, Craig L. 1978. *Topics in German syntax*. MIT: PhD thesis.

Trotzke, Andreas & Joseph Bayer (eds.). 2015. *Syntactic complexity across interfaces.* Berlin: de Gruyter.

Trudgill, Peter. 2011. *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford; New York: OUP.

Ueno, Mieko & Maria Polinsky. 2009. Does headedness affect processing? A new look at the VO-OV contrast. *Journal of Linguistics* 45. 675-710.

van der Hulst, Harry (ed.). 2010. *Recursion and human language*. Berlin/New York: Mouton de Gruyter.

Vater, Heinz. 1986. Zur NP-Struktur im Deutschen. In Heinz Vater (ed.), *Zur Syntax der Determinantien*, 123-145. Tübingen: Gunter Narr.

Zāhāvî, Amotz & Zāhāvî, Avishag. 1997. *The handicap principle: A missing piece of Darwin's puzzle*. New York; Oxford: OUP.

Appendix

Table 10. Frequencies of N-structures.

structure	non-public		public		total	
	n	*%*	*n*	*%*	*n*	*%*
[DP1…]	*1171*	93.5	*1409*	84.8	*2580*	88.5
[DP1…[DP2…]]	*76*	6.1	*221*	13.3	*297*	10.2
[DP1…[DP2…[DP3…]]]	*5*	0.4	*30*	1.8	*35*	1.2
[DP1…[DP2…[DP3…[DP4…]]]]	*0*	0.0	*2*	0.1	*2*	0.1
total	*1252*	100.0	*1662*	100.0	*2914*	100.0

Table 11. Frequencies of V-structures.

structure	non-public		public		total	
	n	*%*	*n*	*%*	*n*	*%*
[VP1…]	*1005*	68.0	*790*	57.1	*1795*	62.7
[VP1…[VP2…]]	*450*	30.4	*533*	38.5	*983*	34.3
[VP1…[VP2…[VP3…]]]	*24*	1.6	*58*	4.2	*82*	2.9
[VP1…[VP2…[VP3…[VP4…]]]]	*0*	0.0	*2*	0.1	*2*	.1
total	*1479*	100.0	*1383*	100.0	*2862*	100.0

Table 12. Frequencies of C-structures.

structure	non-public		public		total	
	n	*%*	*n*	*%*	*n*	*%*
[CP1…]	*739*	67.2	*458*	50.3	*1197*	59.5
[CP1…[CP2…]]	*277*	25.2	*318*	34.9	*595*	29.5
[CP1…[CP2…[CP3…]]]	*56*	5.1	*115*	12.6	*171*	8.5
[CP1…[CP2…[CP3…[CP4…]]]]	*26*	2.4	*19*	2.1	*45*	2.2
[CP1…[CP2…[CP3…[CP4…[CP5…[CP6…]]]]]]	*1*	.1	-	-	*1*	.1
[CP1…[CP2…[CP3…[CP4…[CP5…[CP6…[CP7…]]]]]]]	-	-	*1*	.1	*1*	.1
total	*1099*	100.0	*911*	100.0	*2010*	100.0