

MARGIT BRECKLE (Vilnius Pedagogical University)
HEIKE ZINSMEISTER (University of Konstanz)¹

A corpus-based contrastive analysis of local coherence in L1 and L2 German

Margit Breckle
Vilnius Pedagogical University
Department of German Philology and Didactics
Faculty of Philology
Studentų g. 39
08106 Vilnius
Lithuania
margit.breckle@gmx.de

Heike Zinsmeister
University of Konstanz
Linguistics Department
Box 185
78457 Konstanz
Germany
heike.zinsmeister@uni-konstanz.de

Abstract

The sentence-initial constituent of declarative main clauses in German that immediately precedes the finite verb (the *Vorfeld*, ‘prefield’) is not grammatically determined, but is rather influenced by information structure i.e. the presentation of old and new information in a sentence. It contributes to local coherence – the fluency of a text on a sentence-by-sentence basis – by linking a sentence to its textual context, e.g. by coreferring elements, and other pragmatic means. In this study we compare the use of the *Vorfeld* constituent in texts of the learner corpus ALeSKo which consists of essays by Chinese learners of German (level: ~B2) and comparable essays by German native speakers. The results indicate that the learners have acquired the German *Vorfeld* structure and that they use a range of different grammatical functions in this position almost in a target-language-like way. However, with respect to pragmatic functions they tend to overuse certain coreferring elements, which we interpret as a transfer of pragmatic properties of their native language, Chinese, to their second language, German.

Keywords

L2 German, learner corpus, *Vorfeld*, pragmatic functions

1 Introduction

According to the model of topological fields (e.g. Höhle 1986), declarative main clauses in German are organised in different fields as shown in table 1:

¹ Heike Zinsmeister’s research was financed by the Europäischer Sozialfonds in Baden-Württemberg. We would like to thank our student annotators Sonja Froitzheim, Christine Rieger and Marina Schmieder.

Vorfeld (‘prefield’)	Linke Satzklammer (‘left sentence bracket’) finite verb	Mittelfeld (‘middle field’)	Rechte Satzklammer (‘right sentence bracket’) verbal complex	Nachfeld (‘post field’)
-------------------------	---	--------------------------------	--	----------------------------

Table 1: Topological field model of a German main clause

As stated in the model, the *Vorfeld* is the constituent that precedes the finite verb.^{2,3} In example (1), the *Vorfeld* is realised by the constituent *Auf dieser Reise* (‘During this trip’) – an adverbial prepositional phrase:

(1)	Auf dieser Reise	haben	wir viel	gelernt.	-
	On this travel	have	we a lot	learnt	-
	‘During this trip we have learnt a great deal’				
	<i>Vorfeld</i>	finite verb	middle field	verbal complex	post field

The choice of the *Vorfeld* element in German is not restricted by grammatical functions such as subject or object. As a consequence, the *Vorfeld* can be filled with almost any constituent.

Local coherence describes the fluency of a text on a sentence-by-sentence basis. In German it is assumed that the *Vorfeld* contributes to local coherence by preferably hosting elements that are referentially linked to the preceding context (cf. e.g. Molnár 1991, Reis 2000 as discussed in Speyer 2007). Speyer (2005, 2007, 2008) investigates the pragmatic use of the *Vorfeld* constituent in German in a corpus of different text types. He distinguishes four different pragmatic functions and evaluates which one of them tends to occur in the *Vorfeld* if a sentence contains more than one of these functions. In the present study, we use the pragmatic functions defined by Speyer to find out whether a group of Chinese learners of German (level: ~B2), who have acquired the syntactic structure of German, have also learned the pragmatic use of the *Vorfeld* constituent in a target-language-like way. This is motivated by the fact that learner texts often not sound natural even though individual sentences are perfectly grammatical. This kind of awkwardness is due to an observable overuse or underuse of certain words, functions or constructions in comparison to the target language. In the literature, such proportional misuse with respect to the *Vorfeld* in German is explained inter alia by transfer of pragmatic preferences from the native language of the learner (or from some other secondary language) to the target language, e.g., with respect to subject realisation by advanced English learners of German (von Stutterheim and Carroll 2005) or with respect to information structure by Swedish learners of German (Bohnacker and Rosén 2007).

The aim of our study is to explore to what extent the realisation of the *Vorfeld* in L2-texts corresponds to those of L1-texts and, if applicable, what kind of divergences concerning the pragmatic *Vorfeld* use can be observed.

The paper is organised as follows. In section 2, we will present the corpus-based study including the methodology, the data (ALeSKo corpus), the theoretical background as well as our hypotheses. After reporting on the results on our study in section 3, we draw a number of conclusions (section 4) and give some points for future research (section 5).

² Usually, the *Vorfeld* contains only one constituent. For target-language exceptions that show multiple constituents in the *Vorfeld* see Müller (2005).

³ The *Vorfeld* is not necessarily the first element in all declarative main clauses. It may be preceded by an additional constituent, the *Vor-Vorfeld* (‘pre-prefield’), see also footnote 12 in section 3. For a contrastive analysis of the *Vorfeld* use by Japanese learners of German, which takes the *Vor-Vorfeld* into account see Lipsky (2010).

2 Corpus-based study

2.1 Methodology

Contrastive interlanguage analysis assumes that learners speak an ‘interlanguage’ that systematically differs from the target language. It compares either learner texts with texts of native speakers or data of different learner groups. This method allows us to study the overuse and underuse of certain words and structures by learners which “contribute to the foreign-soundingness of perhaps otherwise error-free advanced interlanguage” (Granger 2008: 267). The basis of our contrastive analysis is the learner corpus ALeSKo which is introduced in the next section.⁴ We compare the *Vorfeld* use in L2 and L1 texts on the basis of manual annotation of grammatical properties and pragmatic functions that are assigned to the *Vorfeld* constituents.

2.2 Corpus

The ALeSKo corpus⁵ (Breckle and Zinsmeister 2010; Zinsmeister and Breckle 2010a, 2010b) has been built since 2009. It consists of argumentative essays in which Chinese L2 learners of German discuss the pros and cons of a given thesis and conclude with their own opinion on the topic. The learner texts are complemented with analogous texts of German native speakers, which have been collected as part of the Falko corpus (Lüdeling et al. 2008). All texts are described by metadata: an ID for the author, native language, year of birth, gender, study programme, foreign language(s), length of L2 exposure and essay topic. Furthermore, all texts are annotated with linguistic information such as part of speech of individual words, topological fields of clauses, and grammatical and pragmatic functions of constituents (for the latter see section 2.3). Table 2 summarises the texts collected in the ALeSKo corpus. The hand-written L2 learner texts were transcribed by two Chinese student assistants in Word documents. The transcriptions were checked independently and corrected necessary.

Subcorpus	Authors	Size	Topic	Conditions
wdt07 ⁶	Chinese L2 learners of German: students at HTGW Konstanz in the BA program ‘Business German and Tourism Management’ in 2007 and 2008, level: ~B2.	25 texts, 6,902 tokens, Ø 23.2 sentences/text	Are holidays an unsuccessful escape from everyday life?	30–45 min, (hand-) written exam, no aids.
wdt08	German native speakers: Berlin high-school students, age 16–19.	All: 39 texts, 34,155 tokens Used in the current study: 24 texts, 19,587 tokens, Ø 48.5 sentences/text	Different topics: ⁸ Women are harmed by feminism more than they benefit from it. Crime is not worth it. People should earn as much money as they contribute to society.	90 min., in-class task, dictionary permitted. 90 min, in-class task, typed in Notepad, no internet access, no spell-checker.

Table 2: Subcorpora of the ALeSKo corpus

⁴ Granger (2008) gives an overview of the creation and use of learner corpora. A comprehensive list of learner corpora is provided at <www.uclouvain.be/en-cecl-lcWorld.html> (31.08.2010).

⁵ ALeSKo is an abbreviation for *Annotiertes Lernersprachenkorpus* (‘annotated learner language corpus’). Its URL is <ling.uni-konstanz.de/pages/home/zinsmeister/alesko.html> (31.08.2010).

⁶ ‘wdt’ is the label for L2 texts collected in the BA program ‘Business German and Tourism Management’ at HTWG Konstanz (German: *Wirtschaftssprache Deutsch und Tourismusmanagement*).

⁷ ‘dhw’ is the label for L1 texts collected at the Highschool ‘Evangelisches Gymnasium Hermannswerder, Deutschland’.

⁸ The translation of the topics is ours. See Falko Spezifikation (2010) for more details on the L1 subcorpus.

The annotation of the current study is based on annotation guidelines (Breckle and Zinsmeister 2009) that define the annotation labels, give examples and discuss problematic cases. The texts were converted into XML to be used by the annotation tool MMAAX2 (Müller and Strube 2006), which we use for the manual annotation task. The annotation process is described in Zinsmeister and Breckle (2010b) in more detail.

2.3 Pragmatic annotation

In our study of the *Vorfeld* in L1- and L2-texts, we focus on pragmatic *Vorfeld* functions. Following Speyer (2005: 79ff.; 2007: 104ff.), we distinguish the following four functions:⁹

- backward-looking center ('familiarity topic')
- brand-new
- poset ('partly ordered set')
- (clause internal) frame-setting

Subsections 2.3.1 to 2.3.4 present these pragmatic functions in detail.

2.3.1 Backward-looking center

Speyer (2007) defines backward-looking center as an approximation to the aboutness and familiarity topic of a sentence (cf. Reinhart 1982). This center concept is part of the Centering Theory (Grosz et al. 1995), the aim of which is to trace local coherence in discourse.

In a discourse, each utterance refers to different entities. These referential expressions are called forward-looking center (ibid.). The Centering Theory assumes that all forward-looking centers are hierarchically organised. The standard saliency hierarchy for German – as well as for English – is assumed to be as follows:

subject > object(s) > others (e.g. adverbials)

In the Centering Theory, the entity which creates coherence is called backward-looking center. It is a constituent that is referentially linked to a salient element in the previous sentence. We adopt the following assumptions for identifying backward-looking centers (cf. Speyer 2007: 92ff.):¹⁰

- *Isolation of the center*: Even referential expressions, which are embedded in other structures, can be considered to be centers, i.e. NPs which are embedded in PPs.
- *Category of the center*: Not only NPs but also certain adverbial expressions can be classified as centers, i.e. referential adverbials or pronominal adverbs such as *dann* 'then', *hier* 'here', *damit* 'with that, thereby', *dafür* 'for that', *davon* 'from that'.
- *Locality of the center*: In cases where the previous clause serves as a thematic insertion (cf. Speyer 2007: 94f.), it is possible to trace coherence not in the preceding clause but in the clause before that.
- *Verbatim quotation of the center*: The backward-looking center has to refer to the same concept as a forward-looking center in the previous clause. However, it does not have to be a strictly verbatim correspondence, e.g. *Bundeswehr* 'German armed forces' vs. *Streitkräfte* 'armed forces' (Speyer 2007: 96).
- *Concept-boundness of the center*: Not only a single concept (phrase) but also a

⁹ In addition, the corpus contains annotations of discourse relations of *Vorfeld* constituents such as *comparison* or *expansion*, which are not evaluated in the current study.

¹⁰ Speyer (2007) has an additional criterion for implicit centers which we do not employ.

whole situation, i.e. a proposition, can serve as antecedent of a backward-looking center.

In order to decide whether a phrase functions as backward-looking center, all referential expressions in the current clause have to be compared with those in the previous one. The antecedent, which is highest on a saliency hierarchy, i.e. in comparison to other potential antecedents, is the backward-looking center. Example (2) from wdt07_22 shows a backward-looking center:

- (2) Durch Reisen können [sie]_1 auch andere Kultur und Lebensstile kennenlernen.
by travel can they also other culture and lifestyles get_to_know
[Sie]_1 können auch ihre Kenntnisse durch Reisen erweitern,
they can also their knowledge by travel broaden
'By travelling, they can also get to know other culture and lifestyles.
They can also broaden their knowledge by travelling.'

In example (2), the referential expression *Sie* ('they') in the second clause *Sie können auch...* ('They can also...') corefers with the expression *sie* ('they') in the previous sentence *Durch Reisen können sie...* ('By travelling, they can...'). The antecedent is highest on a saliency hierarchy and, therefore, the second *Sie* in the Vorfeld is annotated as backward-looking center.

2.3.2 Brand-new

The next pragmatic function does not link the sentence to its preceding context but is part of the information-structural organisation of the sentence. While hearing or reading a text, the hearers/readers create a discourse model in their minds (Webber 1988). This model represents all referents, which are introduced in the text. A referential constituent may either introduce a new discourse referent or refer to some already introduced referent in the model.

The concept of brand-new is defined in Prince (1981: 237). She suggests a "taxonomy of the values of assumed familiarity [of discourse referents]", which distinguishes three major types of familiarity (*new*, *inferrable*, *evoked*) with further subdivisions (e.g. *new* is divided into *brand-new* and *unused*). In our study, the concept of *brand-new* corresponds to Prince's categories *brand-new (unanchored)* und *unused* (ibid.). We refer to *brand-new* when the constituent is not referentially linked to the previous discourse. It has neither been mentioned before nor is it anchored by an indirect reference in the text. Entities which are supposed to be known by the hearer/reader on basis of their general knowledge (e.g. *Barack Obama*) are also annotated as brand-new. Example (3) from wdt07_04 shows a brand-new constituent in the *Vorfeld*:

- (3) [Die Leute, die viele Reise machen,] haben immer mehr Geld als die, die selten reisen.
the people that many travels do have always more money than those that seldom travel
'Frequent travellers have more money than those who travel less.'

The term *Leute* ('people in general') was mentioned in the text before. The term *Die Leute, die viele Reise machen* ('people, who travel a lot') in this example, however, differs from *Leute* ('people in general'), because it refers to a more closely specified group of people. In this instance, the constituent *Die Leute, die viele Reise machen* is introduced to the discourse for the first time and it is therefore annotated as brand-new.

2.3.3 Poset

Another pragmatic *Vorfeld* function is *poset* i.e. *partly ordered set* (cf. Speyer 2007: 104–106). This means that a new entity is introduced in the discourse model, having a

specific relation to another entity, which has already been introduced (cf. Prince 1999: 7). A poset relation can be regarded as a special kind of contrast (cf. Vallduví and Vilkkuna 1998) between two elements that belong to a common set of entities. Only if the second element is introduced the recipient becomes aware of the poset relation. For ease of recognition we used a simplified heuristics to identify poset relations realised by lexical noun phrases: the lemma of the newly introduced discourse referent in the *Vorfeld* is a co-hyponym of the lemma of the previously introduced one. On the lexical level, both words shared an implicit hypernym.

As shown in example (4) from wdt07_04, poset is a constituent, which belongs to a partly ordered set of which another element is already introduced:

- (4) [Jeden Morgen] stehen wir auf, um pünktlich zur Arbeit zu sein. (...)
 Every morning stand we up in order to punctually to the work to be. (...)
 [Jeden Abend] bleiben wir zu Hause, sehe die sinnlosen Serien im Fernseh.
 Every evening stay we at home, watch the senseless series in television.
 ‘Every morning, we get up in order to be at work in time. (...)
 Every evening, we stay at home and watch senseless series on TV.’

In example (4), there are two co-hyponyms – *jeden Morgen* (‘every morning’) and *jeden Abend* (‘every evening’) – the implicit hypernym is *Tageszeiten* (‘times of the day’). Only when the second part of the set is mentioned, it is annotated as poset – therefore, *jeden Abend* (‘every evening’) is annotated as poset.

2.3.4 Frame-setting

Another option for pragmatic *Vorfeld* elements is frame-setting. Jacobs (2001: 656) defines frame-setting as follows: “In (X Y), X is the *frame* for Y if X specifies a domain of (possible) reality to which the proposition expressed by Y is restricted.” This means that one element in the sentence – the frame – specifies a situation in which the remaining clause is interpreted and is said to be true. The concept of frame-setting is similarly defined by Speyer (2008: 280) who points out that frame-setting is “an expression that names a crucial restriction on the situation (such as: the place, the time, etc...) in which the proposition is true”.

Another point to which one has to pay attention is the distinction between frame-setting elements and temporal discourse relations in the *Vorfeld*. A frame-setting element sets the rest of the clause in a temporal frame in which it is said to be true, whereas a temporal discourse relation such as *gleichzeitig* (‘at the same time’) also creates a temporal relation between different clauses.

Example (5) from wdt07_02 shows a temporal frame-setting element:

- (5) [Heutzutage] gelangt es in hoher Konjunktur, einen Urlaub zu machen.
 these_days get it in high trend a holiday to make
 ‘Nowadays, it is more popular to go on holiday.’

The *Vorfeld* is filled with the adverbial *Heutzutage* (‘these days’), which sets a temporal frame in which the rest of the clause *gelangt es in hoher Konjunktur, einen Urlaub zu machen* (‘it becomes very popular to go on holidays’) is interpreted. In this context, it is important to note that the adverbial *Heutzutage* does not refer to another clause.

A locative frame-setting element is realised in example (6) from wdt07_17:

- (6) [In der Natur] genießt man nur die schöne Landschaft.
 in the nature enjoys one only the nice landscape
 ‘In nature, one just enjoys the nice landscape.’

In example (6), the *Vorfeld* constituent *In der Natur* ('In nature) serves as a locative frame in which the remaining sentence is interpreted. As in example (5), the frame setting element does not refer to other sentences but it sets a scene for the current sentence.

2.4 Preferred beginning of the sentence

For German, Speyer (2007: 111) argues for a preference hierarchy of the pragmatic *Vorfeld* functions we have just presented. His findings are derived from a corpus study of different sorts of German texts in which he evaluates which of these functions preferably occurs in the *Vorfeld* if a sentence contains more than one of them. The hierarchy predicts which type of pragmatic function is preferred over other functions to occur in the *Vorfeld*:

frame-setting / brand-new > poset > backward-looking center

Frame-setting and brand-new elements are preferred over both poset and backward-looking center, and poset is in turn preferred over backward-looking center. It is important to note that this hierarchy does not measure the actual frequencies of proportional occurrences of these functions in the *Vorfeld*.

Chinese is a topic-prominent language, which means that "[t]he topic always comes first in a sentence." (Li and Thompson 1989: 15, 85f., 94f.). According to Li and Thompson (*ibid.*), the topic can either be a familiar referent, i.e. a referent which is known from the preceding discourse, or a time or locative phrase (see also Molnár 1991: 183).

Against this background, our working hypothesis is that Chinese topics correspond to both frame-setting elements and backward-looking centers. We therefore assume that the hierarchy for the preferred beginning of a sentence in Chinese is the following:

frame-setting / backward-looking center > others

2.5 Hypotheses

Based on the theoretical background presented in 2.3 and 2.4 and in accordance with findings in related works (cf. Bohnacker and Rosén 2007, Carroll and von Stutterheim 2005), we assume that Chinese learners of German transfer the pragmatic use of the *Vorfeld* constituents from the preferences for the sentence beginning of their L1 into their L2. For our study, we therefore make the following predictions for the use of pragmatic *Vorfeld* elements:

- Backward-looking centers will be overused because they are higher up in the Chinese preference hierarchy than in the German one.
- For frame-setting, we cannot make any prediction, since it ranked highest both in the German and Chinese preference hierarchy.
- For both brand-new elements and poset, we predict an underuse, because they are higher up in the German preference hierarchy than in the Chinese one.

3 Results

For this study we evaluate all 43 L2 texts of the ALeSKo corpus and a subcorpus of 24 texts of the L1 essays, cf. table 2. All declarative sentences that feature a *Vorfeld* are taken into account except for sentences that were marked *erroneous*.¹¹ We also excluded sentences that are marked with *complex Vorfeld*. This label refers to grammatical sentences in which the *Vorfeld* is preceded by additional material (*Vor-Vorfeld* ‘pre-prefield’ and coordination field, respectively, in the topological field model).¹² Given these restrictions, the evaluation compares 884 *Vorfeldd*s in L2 texts with 764 *Vorfeldd*s in L1 texts.

Before analysing the use of pragmatic functions in subsection 3.3, we compare the frequencies of syntactic categories and grammatical functions in the *Vorfeld* in subsections 3.1 and 3.2 to exclude the possibility that observed divergences of the learner language are due to mere syntactic preferences.

3.1 Grammatical function

Table 3 shows that the proportional use of grammatical functions is almost evenly distributed in the L2 texts compared with the L1 texts: most *Vorfeldd*s either host a subject or an adverbial (93% in L2 texts, 87.2% in L1 texts). The L2 texts feature slightly more adverbials in the *Vorfeld* than the L1 texts (42% vs. 36%). This proportional difference is statistically significant, i.e. it cannot be explained just by chance variation ($\chi^2=4.3121$, $df=1$, $p<0.05$). Prepositional object, predicative, accusative object and replacement *es* (‘it’)¹³ are all used in about 2%–3% of *Vorfeldd*s in L1 texts and slightly less frequently in L2 texts except for the prepositional objects, which have a slightly bigger proportion in L2 texts. However, the differences are not statistically significant given the sample sizes. The category *others* subsumes *Vorfeld es* (see footnote 13), dative object and other functions such as a preposed particle of a split particle verb.¹⁴ The L2 texts feature a significant lower number of these functions in the *Vorfeld* than the L1 texts.

Type	Subject	Adverbial	Prepositional object	Predicative	Accusative object	Replacement <i>es</i>	Others	Sum
L2	51.5% (455)	41.5% (367)	3.3% (29)	1.7% (15)	1.0% (9)	0.8% (7)	0.2% (2)	100% (884)
L1	50.8% (388)	36.4% (278)	3.0% (23)	2.7% (21)	2.6% (20)	1.8% (14)	2.6% (20)	100% (764)

Table 3: Grammatical function of the *Vorfeld* constituent (significant differences are typed in bold)

¹¹ There are ten erroneous sentences in the L2 texts: seven cases in which erroneously more than one constituent precedes the finite verb (cf. (i)) and three cases of main-clause structure in embedded sentences which required the finite verb in final position.

(i) **Vielleicht vorhin glaubst du, dass ...* (literal: Maybe before believe you that ..., wdt08_13).

¹² An example for a grammatical *complex Vorfeld* is given in (i) starting with a coordinating *Und*:

(i) *Und Reise macht Menschen auch müde* (‘And (the) trip makes one tired’, wdt07_03).

¹³ We distinguish between *Replacement es*, which substitutes a dislocated sentential constituent, cf.(i) in which a subject clause is extraposed, and *Vorfeld es*, which is a mere filler of the *Vorfeld* and does not substitute any grammatical function in the sentence, cf. (ii), in which the indefinite subject *etwas* is middle field-final:
(i) *Es ist nicht einfach zu beurteilen, ob es von Vorteil ist.* (‘It is not easy to judge whether it is advantageous.’, wdt07_02).

(ii) *Es wird sich nie etwas ändern, ...* (‘Nothing will change...’, dhw004).

¹⁴ An example for verb particle in the *Vorfeld* is (i), with the split particle verb *hinzukommen* (‘is added’).

(i) *Hinzu kommt, dass ...* (‘That ... is added.’ dhw003).

3.2 Syntactic category

The proportional use of syntactic categories in both text types is very similar. Table 4 shows that the only significant difference is the relative preference for prepositional phrases in L2 texts (16.4% vs. 10.1%, $\chi^2 = 13.5244$, $df = 1$, $p < 0.001$), which might correlate with the preference of adverbials described in section 3.1. The observed differences in the other categories are likely to be due to chance.¹⁵

Type	Noun phrase	Adjective/adverb phrase	Prepositional phrase	Sentence	Others	Sum
L2	54.8% (484)	23.4% (207)	16.4% (145)	3.7% (33)	1.7% (15)	100% (884)
L1	57.9% (442)	24.1% (184)	10.1% (77)	5.8% (44)	2.2% (17)	100% (764)

Table 4: Syntactic category of the *Vorfeld* constituent (text type-wise; significant differences are typed in bold)

3.3 Pragmatic functions

With respect to the proportional *Vorfeld* use of the pragmatic functions described in subsection 2.3, we observe two statistically significant divergences between the L2 texts and the L1 texts. Backward-looking centers start a sentence significantly more often in L2 texts than in L1 texts (32.6% vs. 27.1%, $\chi^2 = 5.6089$, $df = 1$, $p < 0.05$). The same seems to hold for brand-new elements (21.3% vs. 16.9%, $\chi^2 = 4.7876$, $df = 1$, $p < 0.05$). Table 5 presents the results of comparing the pragmatic functions on all *Vorfeldd*s in the corpus grouped together (i.e. text type-wise).

Type	Backward-looking center	Brand-new	Partly ordered set (poset)	Frame-setting	None	Sum
L2	32.6% (288)	21.3% (188)	1.9% (17)	7.5% (66)	36.8% (325)	100% (884)
L1	27.1% (207)	16.9% (129)	2.4% (18)	7.3% (56)	46.3% (354)	100% (764)

Table 5: Pragmatic function of the *Vorfeld* constituent (text type-wise; significant differences are typed in bold)

However, taking into account that our corpus consists of more L2 texts than L1 texts, cf. table 2, and that each individual text tends to introduce some brand-new elements¹⁶ the L2 subcorpus is likely to feature more brand-new elements for the simple fact of containing a higher number of text beginnings. Figure 1 illustrates the textwise evaluation by means of ‘box plots’. The boxes visualise how the proportion of backward-looking centers and brand-new elements are distributed across the texts. The crosses inside the boxes mark the means of the texts in each subcorpus (and the vertical lines the statistical medians). Note, that these values differ from the text type-wise evaluation in table 5. The boxes themselves correspond to the values that 50 % of the texts take. A large box indicates a lot of variation in the data, whereas a flat box means that the values do not differ a lot. The dashed lines with vertically marked ends are called ‘whiskers’. They delimit the statistical bound of ‘even’ distributions. Observations that fall outside these bounds are ‘outliers’ that have extreme values with respect to the rest of the data. In the case of the proportional use of brand-new elements, there are three outliers, i.e., three texts that differ greatly in their proportion of brand-new elements in the *Vorfeld* in comparison to the other texts. We interpret

¹⁵ The category *Others* subsumes free-relative clauses, infinitives, particles etc.

¹⁶ There is in fact one short L2 text that does not introduce any brand-new element in the *Vorfeld* (wdt07_01).

the results as follows: The mean proportional use of backward-looking centers in L2 texts is higher than in L1 texts.¹⁷

With respect to brand-new elements, the difference between the means in L2 texts and L1 texts becomes insignificant when the three outliers are excluded from the evaluation (cf. ‘L2: brand-new (clean)’). This means that the apparent overuse of brand-new elements by L2 learners is not confirmed in the textwise evaluation.

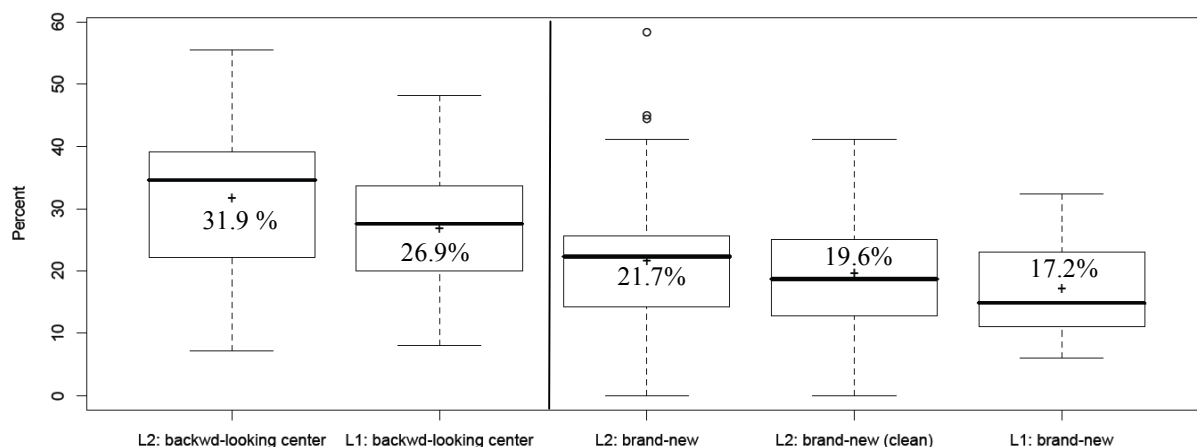


Figure 1: The means of textwise distribution of backward-looking centers and brand-new elements in the *Vorfeld*

4 Conclusion

The Chinese Learners of German (level: ~B2) have acquired the German *Vorfeld* structure, i.e., they regularly use one constituent preceding the finite verb in declarative main clauses. They also use different syntactic categories and grammatical functions in a target-language-like way – there is only a slight overuse of prepositional phrases and adverbials, respectively, and an underuse of less common functions such as preposed verb particles or the *Vorfeld* es, cf. sections 3.1 and 3.2. We hypothesised in section 2.5 that the learners would transfer the pragmatic functions at the beginning of a sentence in Chinese to German. The study shows mixed results with respect to our predictions: As predicted backward-looking centers are slightly overused and there is no significant difference in relation to frame-setting elements. However, the underuse of brand-new elements and poset relations could not be confirmed in the data. The overall evaluation also indicates an overuse of brand-new elements, which, however, disappears in the textwise evaluation. We conclude that regarding the proportional use of pragmatic functions in the *Vorfeld*, Chinese learners of German show a preference for familiar referents which correspond to topics that start a sentence in Chinese, thus

¹⁷ The footnote presents the statistics for the *textwise* evaluation including the mean (‘ \bar{x} ’) as well as the lower bound (‘lo’) and the higher bound (‘hi’) of the respective 95% confidence interval.

Proportional use of backward-looking center per text:

- L2: \bar{x} 31.9%, lo: 28.1%, hi: 35.8%
- L1: \bar{x} 26.9%, lo: 23.1%, hi: 30.7%.

The means are significantly different according to a one-sample t-test – only a one-sample t-test is possible since the variances of the two distributions are not homogeneous. However, given the variance in the data, the 95% confidence intervals overlap so that the overall textwise difference is not significant.

Proportional use of brand-new elements per text:

- L2 without outliers: \bar{x} 19.6%, lo: 16.7%, hi: 22.5% (L2 with outliers: \bar{x} 21.7%, lo: 18.1%, hi: 25.3%)
- L1: \bar{x} 17.2%, lo: 13.9%, hi: 20.5%.

The textwise difference is not significant since the 95% confidence intervals overlap. If the three outliers are not taken into account, the means of the distributions are even included in each other’s confidence intervals.

suggesting that the learners indeed transfer pragmatical preferences of their native language to their second language.

5 Future work

In order to make full use of the predictions by Speyer (2005, 2007), we need to investigate the pragmatic functions of all constituents in a sentence, not just those of the *Vorfeld* element. This requires further annotation of the corpus. In addition, we plan to investigate the influence of *Vor-Vorfeld* elements and the effect they have on local coherence as well as the marking of discourse relations such as *contingency* and *expansion*. The current study is only an approximation to local coherence. With the aim to evaluating the correlation between *Vorfeld* use and local coherence, we plan to perform a rating and rewriting experiment on the texts (cf. Rosén 2006).

6 References

- Bohnacker, Ute, Christina Rosén (2007). How to start a V2 declarative clause: Transfer of syntax vs. information structure in L2 German. Anderssen, Merete, Westergaard, Marit, eds. *Nordlyd: Papers from the Language Acquisition Workshop, SCL 2006*. Tromsø: CASTL. 29–56.
- Breckle, Margit, Heike Zinsmeister (2009). Annotationsrichtlinien Funktion des Vorfelds. Manuscript. December 2009. Pedagogical University Vilnius and University of Konstanz.
- Breckle, Margit, Heike Zinsmeister (2010). Zur lernersprachlichen Generierung referierender Ausdrücke in argumentativen Texten. In Dirk Skiba (ed.) *Textmuster: schulisch-universitär-kulturkontrastiv*. 79–101. Frankfurt/Main: Peter Lang.
- Falko Spezifikation (2010). Spezifikationen des Falko Korpus 2.0. Version 1.0. Manuscript. Humboldt Universität zu Berlin.
- Granger, Sylviane (2008). Learner corpora. Lüdeling, Anke, Merja Kytö, eds. *Corpus Linguistics. An International Handbook*. (Handbooks of Linguistics and Communication Science 29). Berlin et al.: de Gruyter, 259–275.
- Grosz, Barbara, Aravind Joshi, Scott Weinstein (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* 21: 203–225.
- Höhle, Tilman (1986). Der Begriff ‚Mittelfeld‘. Anmerkungen über die Theorie der topologischen Felder. Weiss, Walter, Herbert Ernst Wiegand, Marga Reis, eds. *Textlinguistik contra Stilistik? – Wortschatz und Wörterbuch – Grammatische oder pragmatische Organisation von Rede?* Tübingen: Niemeyer, 329–340.
- Jacobs, Joachim (2001). The dimensions of topic-comment. *Linguistics* 39: 641–681.
- Li, Charles N., Sandra A. Thompson (1989). *Mandarin Chinese: A Functional Reference Grammar*. Berkeley, Los Angeles: University of California Press.
- Lipsky, Angela (2010). Vorfeldfehler in Texten von Deutschlernern. *Deutsch als Fremdsprache* 2 (2010): 70–76.
- Lüdeling, Anke, Seanna Doolittle, Hagen Hirschmann, Karin Schmidt, Maik Walter (2008). Das Lernerkorpus Falko. *Deutsch als Fremdsprache* 2 (2008): 67–73.
- Molnár, Valéria (1991). *Das TOPIK im Deutschen und im Ungarischen*. Stockholm: Almqvist & Wiksell.
- Müller, Christoph, Michael Strube (2006). Multi-level annotation of linguistic data with MMAX2. Braun, Sabine, Kurt Kohn, Joybrato Mukherjee, eds. *Corpus Technology and Language Pedagogy. New Resources, New Tools, New Methods*. Frankfurt/Main: Lang, 197–214.
- Müller, Stefan (2005). Zur Analyse der scheinbar mehrfachen Vorfeldbesetzung. *Linguistische Berichte* 203, 29–62.

- Prince, Ellen F. (1981). Toward a taxonomy of given-new information. Cole, Peter, ed. *Radical Pragmatics*. New York: Academic Press, 223–255.
- Prince, Ellen F. (1999). How not to mark topics: ‘Topicalization’ in English and Yiddish. *Texas Linguistics Forum*, chapter 8. Austin: University of Texas.
<ftp://babel.ling.upenn.edu/papers/faculty/ellen_prince/texas.ps> (30.6.2010)
- Reinhart, Tanja (1982). Pragmatics and linguistics. An analysis of sentence topics. Reprint by the Indiana University Linguistics Club (originally published 1981 in *Philosophica* 27: 53–94).
- Reis, Marga (2000). Anmerkungen zu Verb-erst-Satz-Typen im Deutschen. Thieroff, Rolf, Matthias Tamrat, Nanna Fuhrhop, Oliver Teuber, eds. *Deutsche Grammatik in Theorie und Praxis* 215–227. Tübingen: Niemeyer.
- Rosén, Christina (2006). *Warum klingt das nicht deutsch? Probleme der Informationsstrukturierung in deutschen Texten schwedischer Schüler und Studenten*. Stockholm: Almqvist & Wiksell International.
- Speyer, Augustin (2005). Competing Constraints on Vorfelddbesetzung in German. Sassen, Claudia, Anton Benz, Peter Kühnlein, eds. *Proceedings of the Constraints in Discourse Workshop*, Universität Dortmund, 79–87.
- Speyer, Augustin (2007). Die Bedeutung der Centering Theory für Fragen der Vorfelddbesetzung im Deutschen. *Zeitschrift für Sprachwissenschaft* 26: 83–115.
- Speyer, Augustin (2008). German Vorfeldd-filling as constraint interaction. Benz, Anton, Peter Kühnlein, eds. *Constraints in Discourse*. (Pragmatics & Beyond New Series 172). Amsterdam, Philadelphia, PA: John Benjamins, 267–290.
- Stutterheim, Christiane von, Mary Carroll (2005). Subjektwahl und Topikkontinuität im Deutschen und Englischen. *Zeitschrift für Literaturwissenschaft und Linguistik* 139: 7–27.
- Vallduví, Eric, Maria Vilkuña (1998). On rheme and kontrast. Culicover, Peter, Louise McNally, eds. *Syntax and Semantics 29: The Limits of Syntax*. New York: Academic Press, 79–108.
- Zinsmeister, Heike, Margit Breckle (2010a). ALeSKo - an annotated learner corpus. Poster presented at the poster session of the 32. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft (DGfS), Humboldt-Universität zu Berlin. <http://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/mitarbeiterinnen/amir/Posters/Zinsmeister_Breckle.pdf> (30.6.2010)
- Zinsmeister, Heike, Margit Breckle (2010b). Starting a sentence in L2 German – Discourse annotation of a learner corpus. *Proceedings of KONVENS 2010*. Saarland University.