Isolating the syntactic factor in non-standard punctuation : a thought experiment with pilot study on the Vorfeldkomma in written German

Date of secondary publication: 30.08.2023

Version of Record (Published Version), Bookpart

Persistent identifier: urn:nbn:de:bvb:473-irb-902616

Primary publication

Reiner, Tabea (2022): Isolating the syntactic factor in non-standard punctuation : a thought experiment with pilot study on the Vorfeldkomma in written German. In: Vergleichende Interpunktion = Comparative Punctuation. Rössler, Paul; Besl, Peter; Saller, Anna (Ed.). Berlin : De Gruyter S. 91-108. DOI: 10.1515/9783110756319-005

Tabea Reiner

Isolating the Syntactic Factor in Non-Standard Punctuation

A Thought Experiment with Pilot Study on the *Vorfeldkomma* in Written German

Abstract: The present contribution demonstrates that *Vorfeldkommas*, i. e. non-standard commas after prefields, are not only relevant for language teaching but also from a syntax theoretic point of view. Didactic research has over the years identified a wealth of potential factors for these commas, e. g., prefield length or syntactic function. More recently, these variables were also tested for the *Vorfeldkomma's* counterpart, i. e. correctly lacking commas in the same place. However, it is hardly ever asked whether syntactic position as such attracts non-standard commas and how this might be accounted for theoretically. The paper sets out to fill this gap and concludes that indeed syntax acts separately to a certain extent.

1 The *Vorfeldkomma* – what is it?

1.1 Definition and examples

Literally, *Vorfeldkomma* translates as 'prefield comma' where the prefield is, roughly, the position in front of the finite verb in declarative sentences, according to the so called topological model (for a short introduction to this framework in English, cf. St. Müller in prep.: ch. 1). This position hosts exactly one constituent, which may be a noun phrase, an adjective/adverb phrase, or a clause. In this sense, the end of the prefield is a syntactically distinguished spot in German. Although syntax appears to play a major role in norms for comma placement (when viewed as systems, cf. Primus 1993, Primus 2007, Bredel 2008), the end of the prefield does not require a comma *per se* (Rat für Deutsche Rechtschreibung 2018, henceforth *Amtliche Regelung*). Put differently: in written Standard German the prefield ends in a comma if (and only if) there is another reason to use one, e. g. marking a clause boundary.

Against this background, the *Vorfeldkomma* is commonly defined as a superfluous, hence non-standard, comma after the prefield. Following are two arbitrarily picked examples with glossing for the prefield and finite verb.

https://doi.org/10.1515/9783110756319-005

(1) Nach einer über 2000-jährigen Phase einer vergleichsweise
after a more.than 2000.years.long phase of.a comparatively
langsamen Entwicklung, sind
slow development are

die letzten 50 Jahre der Linguistik von einem außerordentlichen Wissenszuwachs gekennzeichnet. (H. Müller: 15)

'After more than 2000 years of comparatively slow development, the last 50 years of linguistics have been marked by an extraordinary increase in knowledge.'

(2) Durch die Exzellenzinitiative, den Hochschulpakt und den
through the Excellence Initiative the Higher Education Pact and the
Pakt für Forschung und Innovation, hat
Pact for Research and Innovation has

sich die Zahl der Qualifikationsstellen innerhalb weniger Jahre von 108.000 auf 158.000, [...], erhöht. (Resolution des 65. DHV-Tages in Mainz)

'Through the Excellence Initiative, the Higher Education Pact and the Pact for Research and Innovation, the number of qualification positions has increased from 108,000 to 158,000 [...] within a few years.'

1.2 Aims and scope of previous research

The *Vorfeldkomma* has been a topic in the didactics of L1-orthography for more than 40 years now, with frequency counts suggesting that it is an error rather than a mistake (Berg et al. 2020: 86). The systematic nature of the phenomenon raised the following questions:

1. Why do writers, including but not limited to students at school, use and accept the *Vorfeldkomma*?
2. How can we improve teaching in this respect?

I will focus on question number 1. With respect to this question, a whole range of possible factors and combinations of them has been brought to the fore, (non-exhaustively) summarized in table 1.

In table 1, "indirectly" means that the pertaining variable derives from one or more of the other factors in the same column. "Possibly" indicates that the respective factor is cited from previous studies and considered plausible. Moreover, Sappok's (2011) two factors – information structure and length – are to be understood as connected by OR: it is either the one or the other or both.

Tab. 1: *Vorfeldkomma* – suggested explanatory variables

	Barthel/Löffler (1976: 108–109)	Naumann (1995: 215, 224)[a]	Müller (2007: 210–220)	Sappok (2011: 331–332)	Primus/Kirchhoff (2014: 217–218)	Bredel/Hlebec (2015: 37)	Lindbüchl (2015: 82)[b]	Boettcher (2016: 357–358)	Hochstadt/ Olsen (2016)	Rössler (2017)	Berg et al. (2019; 2020)
prosodic boundary after prefield	x	x	indirectly		x	x	?	x	possibly	indirectly	
information-structural boundary after prefield			x (new topic)	x				indirectly (thematic and argumentative weight)		possibly	
certain length of prefield	x	indirectly	–	x	indirectly	x		indirectly		x	x
first element of prefield = preposition						indirectly				indirectly	x
high complexity of prefield									x		
syntactic function of prefield constituent = adverbial	x				indirectly		x			indirectly	x
content of prefield constituent = separate prediction						x			possibly	x	–
language contact							?	x	possibly		
misguided didactics						x				indirectly	

[a] Speaking of "Erstposition" 'first position'. [b] Speaking of "satzinitialer Position" 'sentence initial position'.

What happens rarely in the literature on *Vorfeldkommas* is that factors that were previously suggested are explicitly challenged. However, two exceptions from this pattern concern "length" and "separate predication". As to length, Müller (2007: 213, 215), though on the whole arguing very tentatively, states that his study could not confirm prefield length as a trigger for non-standard commas after prefields. All the more surprisingly, this factor resurfaces in the two most recent studies from table 1, i. e. Rössler (2017) and Berg et al. (2019; 2020). This is remarkable not only because these studies are recent but also because especially the latter one appears to be unprecedented in empirical coverage. Please note, however, that the critical number of words is discerned as ~7 by Rössler (2017) whereas for Berg et al. (2019; 2020) it is ≥ 7 (for more details on length as a controversial factor, cf. section 2.1).

As to predication, this is Rössler's (2017) other central factor while Berg et al. (2020: 105) could not replicate his finding with respect to deverbal nouns followed by a genitive noun phrase encoding the former object. Then again, in trying to operationalize a factor that is admittedly hard to operationalize, they might have overspecified their query. Overspecification may arise from two sources.

First, there are certainly predications conveyed by nominalizations coming without any genitive modifier, especially when the genitive's alternative [*von* + dative] is available. (3) is an example found via WebCorp Live[1].

(3) Nach dem Auffinden von Palladino und ihrem Kavat Rook bittet
after the finding of Palladino and her Kavat Rook asks

sie den Tenno, sie nach Eisenwacht zu bringen, und er muss sie gegen Schatten des Red Veil verteidigen, bevor ihr abreisen könnt. (https://warframe.fandom.com/de/wiki/Ketten_des_Harrow, 14 July 2020)

'After finding Palladino and her Kavat Rook, she asks the Tenno to take her to Eisenwacht and he must defend her against shadows of the Red Veil before you can leave.'

Thus, diagnosing a separate predication in a nominal does not need the presence of a genitive modifier that conserves (part of) the former valency information. Moreover, with respect to a certain class of verbs, this type of genitive modifier is not even available: verbs that require as their object a prepositional phrase rather than a nominal phrase keep this prepositional phrase when undergoing nominal-

1 http://www.webcorp.org.uk/live/index.jsp; I continue glossing only the prefield and finite verb.

ization (needless to say, the prepositional phrase is optional then). (4) is another example found via WebCorp Live.

(4) Trotz des Wartens auf den neuen Passat bleibt
despite the waiting for the new Passat remains

VW Flottenmarke Nummer eins. (https://www.autoflotte.de/artikelarchiv/artikel/wachsen-und-warten-1577724.html, 14 July 2020)

'Despite waiting for the new Passat, VW remains the number one fleet brand.'

Second, overspecification may arise from the requirement that the genitive noun phrase be definite. Berg et al. (2020: 105) claim that this is the most frequent type without giving justification. To be sure, their claim is not unplausible, for example searching the *Deutsches Referenzkorpus* (subcorpus *W–öffentlich*) for the string *des Problems* ('of the problem') yields 45,998 hits but searching the same corpus for the string *eines Problems* ('of a problem') yields merely only 3,222 hits. Still, by ignoring indefinite genitive phrases Berg et al. exclude a set of relevant instances from the outset. (5) is, again, an example found via WebCorp Live.

(5) Nach dem Beantworten einer Bestätigungsmail wird
after the replying to.a confirmation e-mail becomes

die Unterschrift gültig, […]. (https://bdsgruppebonn.wordpress.com/2017/05/02/unterstuetzt-den-hungerstreik-der-palaestinensischen-gefangenen/, 14 July 2020)

'After replying to a confirmation e-mail, the signature becomes valid.'

Summing up on predication, I have to remain agnostic about its role as a trigger of *Vorfeldkommas* for the moment. On the one hand, the considerations above cast serious doubt on Berg et al.'s (2020) refutation of this factor; on the other hand, it is by no means a given that a more inclusive search would eventually lead to a different result. Moreover, the two studies are based on different types of corpora (a manually compiled corpus of term papers in the case of Rössler 2017 vs. a large corpus from the web in the case of Berg et al. 2019; 2020). For purposes of the present contribution, I do include predication as a potential factor; however, future research should probably target this factor in particular.

What happens virtually never in the *Vorfeldkomma*-literature is that syntax as such is recognized as a potential factor, i. e.: prefields are usually not considered to trigger non-standard commas at their right edge just because they are prefields. Certainly, Bredel & Hlebec (2015: 37) as well as Bredel (2015: 209–212, not in table 1) and, subsequently, Hochstadt & Olsen (2016: 171–172) do concede a certain role to

syntax in general and to the prefield boundary in particular – but mainly as a basis for the real factors to operate on. This seems natural since the position of *Vorfeldkommas* is part of the phenomenon to be explained, hence it cannot figure as an explanatory variable like any other. However, section 2 of the present paper will show that a distinctly syntactic perspective on these commas is possible and even necessary.

2 Is there anything new to say about the *Vorfeldkomma*? A thought experiment on the syntactic side of the phenomenon

2.1 A new perspective on previous research

This section aims at viewing previous research from a different angle than in table 1: besides listing the explanatory variables, I intend to show how the response variable is composed. In fact, it consists of three individual variables, which can be separated – and began to be separated only recently.

Table 2 presents the "classical" kind of research on the *Vorfeldkomma*: it is about the phenomenon as a whole, i. e. about commas that writers have non-standardly placed directly after the prefield.

Tab. 2: *Vorfeldkomma* – as a holistic response variable

	[+after prefield] & [+placed] & [–standard]
prosodic boundary after prefield	?
information-structural boundary after prefield	?
certain length of prefield	?
first element of prefield = preposition	?
high complexity of prefield	?
syntactic function of prefield constituent = adverbial	?
content of prefield constituent = separate predication	?
language contact	?
misguided teaching	?

Tab. 3: *Vorfeldkomma* – as a decomposed response variable in Berg et al. (2019; 2020); explanatory variables not considered by the authors are in white

	[+after prefield] & [+placed] & [–standard]	**[+after prefield] & [–placed] & [+standard]**
prosodic boundary after prefield	?	?
information-structural boundary after prefield	?	?
length of prefield ≥ 7 words	?	?
first element of prefield = preposition	?	?
high complexity of prefield	?	?
syntactic function of prefield constituent = adverbial	?	?
content of prefield constituent = separate predication	?	?
language contact	?	?
misguided teaching	?	?

Apart from occasional remarks, the classical approach prevailed until Berg et al.'s (2019; 2020) study. These researchers are the first ones to compare the number of *Vorfeldkommas* with the number of commas that are not there. More precisely, they expand table 2 to table 3.

Interestingly, Berg et al. modify two of the individual response variables at once: [+placed] becomes [–placed] and [–standard] becomes [+standard]. This can be seen from their discussion of examples (6) and (7) (Berg et al. 2020: 94, numbers (20a) and (20b) there).

(6) Nach langer Wartezeit auf entsprechende finanzielle Mittel, sind
after long waiting time for appropriate financial means are

wir nun stolz unsere (relativ) neuen Rechner präsentieren zu dürfen.

'After a long time of waiting for appropriate financial means we are now proud to present our (relatively) new computers.'

(7) Die junge Generation_ diskutierte
the young generation discussed

über ihre Meinungen, Ansichten und Vorschläge.

'The young generation discussed their opinions, views and proposals.'

Each one represents one of the two structures they are looking for, the first one involving a classical *Vorfeldkomma*, i. e. a comma with the feature values [+after prefield] & [+placed] & [–standard], the second one showing a normatively required decision against placing a comma at this position, i. e. a "comma" with the feature values [+after prefield] & [–placed] & [+standard].

Why do they opt for precisely this modification rather than any other like, e. g., [+after prefield] & [+placed] & [+standard], which would be represented by prefields consisting of an adverbial clause or ending in an attributive clause? I hold that there are two reasons, one implicit and one explicit.

The implicit one is that the comparison in table 3 makes sense from a didactic perspective: we want to compare the *wrong* (i. e. not conforming to the current standard) decision *for* placing commas after prefields to the *correct* (i. e. conforming to the current standard) decision *against* placing commas in the same spot. Thus, 100 % equals all comma decisions after non-comma requiring prefields and we can determine the percentage of the wrong decisions.

The explicit reason is that by including prefields without commas they can check whether Rössler's (2017) version of the factor length is a side effect of very long prefields being rare in general (Berg et al. 2020: 89), i. e. whether *Vorfeldkommas* after more than 7 words are rare just because prefields of more than 7 words are rare. They conclude that this is indeed the case: the longer the prefield gets the more likely *Vorfeldkommas* become, without an upper boundary (Berg et al. 2020: 99). However, their pertinent diagram (Berg et al. 2020: 98) rather supports the view that there is one peak at 7 words and another one at 11 words, after which the percentage of *Vorfeldkommas* drops, albeit staying on a high level. To be sure, for 14+ words the diagram shows an increase again; however, this may very well be an artefact of the fact that prefields of 14, 15, 16, 17,… words are accumulated here. Thus, we do not know on which level the percentages are for prefields of 14 words, prefields of 15 words and so on when recorded individually. In sum, I see two peaks rather than a linear increase. So I suspect that the influence of prefield length on placing *Vorfeldkommas* is more complicated than both, Rössler's (2017) finding that prefield length in words has to be around 7 as well as Berg et al.'s (2020) conclusion that it has to be at least 7. It remains a task for future research to investigate whether the second peak can be replicated. For the rest of the present paper, I will simply treat the two current versions of the factor length side by side. To sum up, Berg et al. (2019, 2020) start decomposing the *Vorfeldkomma* as a response variable into its constituent parts. However, they have reasons to restrict themselves to examining merely one alternative composition of these parts.

2.2 The psychological reality of prefield boundaries

Considering that the *Vorfeldkomma* as a response variable is composed of three features with two values each, there are in total $2^3 = 8$ ways of combining them. These combinations represent phenomena that differ from the original phenomenon to varying degrees – and not all of them appear to be equally interesting. For instance, the combination [–after prefield] & [–placed] & [+standard] covers every right boundary of a word (apart from the right edge of the prefield) that must not and at the same time is not followed by a comma, i. e. almost all right word boundaries.

What I intend to do in the present contribution, however, is a bit more radical than merely recomposing the response variable. I will pull out the syntactic feature, i. e. [+/–after prefield], from the phenomenon to be explained and push it into the explanatory variables. That is, I ask the question that is virtually never asked in *Vorfeldkomma* research: whether prefields trigger non-standard commas at their right edge just because they are prefields. From a didactic point of view, this might or might not be an interesting question to ask. From a syntax theoretic point of view, however, it is a quite fundamental one to ask: if a syntactic

Tab. 4: *Vorfeldkomma* – dragging the syntactic variable form the composite response variable into the explanatory variables

	[+placed] & [–standard]	**[–placed] & [–standard]**
[+after prefield]	?	?
[–after prefield]	?	?
[+after ~7 words from start of sentence]	?	?
[–after ~7 words from start of sentence]	?	?
[+after ≥ 7 words from start of sentence]	?	?
[–after ≥ 7 words from start of sentence]	?	?
[+constituent preceding comma position starts with preposition]	?	?
[–constituent preceding comma position starts with preposition]	?	?
[+constituent preceding comma position is an adverbial]	?	?
[–constituent preceding comma position is an adverbial]	?	?
[+comma position is between predications]	?	?
[–comma position is between predications]	?	?

boundary as such triggers non-standard commas then this is an indication that the boundary is psychologically real in some sense.

In more detail, table 4 visualizes the comparisons I will draw. Please note that isolating the syntactic variable from the composite response variable and treating it as an explanatory variable requires the other explanatory variables to be redesigned, too. I focus on (redesigned versions of) those factors that were suggested in the two most recent studies, i. e. in Rössler (2017) and Berg et al. (2019; 2020).

For systematic reasons as well as for Berg et al.'s (2020: 89) reasoning on long prefields, table 4 should be expanded by two further columns: [+placed] & [+standard] and [–placed] & [+standard]. However, I leave this expansion to future research and focus on non-standard commas here.

3 Probing the syntactic factor in a pilot study

3.1 Data

3.1.1 Materials

In order to explore how the gaps in table 4 may be filled, I compiled a corpus from term papers, very much in the spirit of Rössler's (2017) study. However, since mine is no more than a pilot study, probing a new analytic direction, my corpus is very small, consisting of merely five papers. These were written in a BA course on language acquisition in the winter term of 2016/2017 and fully corrected then. This setting comes with two assets. First, at the time of correcting, I did not know that I was creating data for a future study, hence I was not biased in any direction, e. g., towards finding more *Vorfeldkommas* than any other type of non-standard comma. Second, I know the writers' linguistic background so that I was able to pick those who grew up with German as their sole L1.

The essential part of these data is assessing punctuation, so some details on the standards I used are worth mentioning. For one thing, I basically followed the *Amtliche Regelung* but ignored two of its supplementary rules, i. e. § 75 E1 and E2. These rules relax certain principles introduced before (requiring certain commas and banning others) under the condition that communication is, at the very least, not impeded. In most cases, I find it hard to decide whether this condition holds, so ignoring the supplementary rules altogether appeared to be the only option. The same holds for § 73. For another thing, I spotted a few blunders in my earlier corrections. In these cases I revised the correction and indicated my revision in the data sheets (cf. section 3.2.1).

3.1.2 Processing

Originally, the term papers were available as .docx-files with mark-ups. For tokenization, I converted the bodies of the papers into tab separated text files and, finally, into Excel sheets with no more than one graphematic word per cell, e. g. ‹prosodischen›, ‹Merkmalen,›. Each row corresponds to one graphematic sentence[2], each column corresponds to one linear position (up to position number 52), and each sheet corresponds to one paper (apart from the last sheet). Moreover, I kept headlines but ignored them with respect to annotation (cf. section 3.2.1). Direct quotations however, were kept and annotated as long as they were syntactically and – in the original file – typographically integrated. This introduces some alien material but ensures that the position count stays realistic.

3.2 Method

3.2.1 Annotation: inventory

In total, I used 13 labels, 9 of which will be directly relevant to the results and discussion (cf. sections 3.3 and 3.4). The whole set is given in table 5 and table 6 with the important cells marked grey.

Tab. 5: Annotations in cells

tag in cell	specification
<7OK–>	no comma at position 6, 7, or 8; standard
<7OK+>	comma at position 6, 7, or 8; standard
<VOK–>	no comma after prefield; standard
<VOK+>	comma after prefield; standard
<del>	comma, non-standard (read: deleted)
<deln>	comma, non-standard (read: deleted); new
<ins>	no comma, non-standard (read: inserted)
<insn>	no comma, non-standard (read: inserted); new

2 This notion might be refined by Neef's (2019; this volume) notion of *Schreibäußerung*.

Tab. 6: Annotations in columns

feature in columns after position numbers	feature values	specification
len [read: length]	1–	for <del(n)>/<ins(n)>: position number for <VOK–/+>: prefield length
prefield	+/–	for <del(n)>/<ins(n)>: comma position is after prefield
initial prep	+/–	for <del(n)>/<ins(n)>: constituent preceding comma position starts with preposition
adverbial	+/–	for <del(n)>/<ins(n)>: constituent preceding comma position is an adverbial
double pred	+/–	for <del(n)>/<ins(n)>: comma position is between predications

3.2.2 Annotation: procedure and premises

The fact that I chose graphematic sentences as reference point has two consequences. For one thing, there may be more than one prefield per sentence = row. In these cases, I added an extra row in order to accommodate the second prefield tag (<VOK–/+> or <del(n)/ins(n)>). For another thing, if this tag was <del(n)> and came with [+prefield] then, in effect, I was dealing with a *Vorfeldkomma* whose position number is higher than the prefield is long. However, this only happened once in the data set, cf. (8).

(8) Kindgerichtete Sprache bezeichnet kein starres Phänomen, denn
child-directed language means no rigid phenomenon for
neben der in 2.1.1 und 2.1.3 bemerkten Dynamik<del>,<del> weist
besides the in 2.1.1 and 2.1.3 noted dynamics exhibits
sie auch kulturbedingte Unterschiede auf. [cb110a][3]
'Child-directed language is not a rigid phenomenon, because in addition to the dynamics noted in 2.1.1 and 2.1.3, it also exhibits cultural differences.'

In (8), *denn* is a coordinating conjunction in the so called pre-prefield while the prefield proper extends from *neben* to *Dynamik*. It is 8 words long (counting the

3 Letters refer to authors, numbers refer to row IDs.

section numbers as one word each) but the comma is at position 15 in the graphematic sentence.

Theoretically, the same situation may arise as a byproduct of long pre-prefields. In practice, however, I found only one type: [$_{\text{pre-prefield}}$ V1-conditional] [$_{\text{prefield}}$ *so*] V_{finite}. Since its tokens involved exclusively standard comma decisions and I focus on non-standard commas here (cf. table 4), these examples did not have an impact on the results.

A more pressing analytical question is how to identify separate predications. As heuristics I used participles and nominalizations (cf. Rössler 2017); however, the decisive criterion was my intuition that, in the end, there is a separate propositional nucleus. Admittedly, this introduces a certain degree of subjectivity.

In contrast, the topological assumptions on which I based the analysis were clear-cut and I list them here to conclude the present section:

- There are no invisible prefields (e. g., in conjunction reduction).
- There is no prefield if the finite verb is invisible.
- There is no prefield in relative clauses (Dürscheid 2012: 93).

3.3 Results

An overview of the results is given by table 7.

As has been remarked repeatedly in the literature at least since Pießnack & Schübel (2005: 58), lacking commas are much more common than superfluous commas in written Standard German. With respect to table 7, this means that we have to expect the numbers in the right most column to be invariably higher than the numbers to the left. However, there are two exceptions to this rule in table 7: the rows [+after prefield] and [+… adverbial]. Additionally, [+… preposition] comes close to patterning with these two.

Thus, the topological factor as well as one or two of Berg et al.'s (2019; 2020) factors behave differently from the others. As to the latter ones, though, this might be an artefact of the requirement to re-design these factors for purposes of the present study (cf. section 2.2): in cases where there was no (complete) constituent preceding the comma position on neither level (syntactic sentence or clause), the labels were simply not applicable – which affected the count of inserted commas more as four times as often as the count of deleted commas (21 times vs. 5 times)[4]. Thus, there seems to be an independent reason why the usual trend of

4 This is also why not all pairs of rows in table 7 add up to 31 deleted commas and 59 inserted commas.

Tab. 7: Results of pilot study

	[+placed] & [–standard], i. e. <del(n)>	[–placed] & [–standard], i. e. <ins(n)>
[+after prefield]	12	9
[–after prefield]	19	50
[+after ~7 words from start of sentence] [a]	10	14
[–after ~7 words from start of sentence]	21	45
[+after ≥ 7 words from start of sentence]	22	44
[–after ≥ 7 words from start of sentence]	9	15
[+constituent preceding comma position starts with preposition]	10	11
[–constituent preceding comma position starts with preposition]	16	27
[+constituent preceding comma position is an adverbial]	12	7
[–constituent preceding comma position is an adverbial]	14	31
[+comma position is between predications]	10	14
[–comma position is between predications]	21	45

[a] I operationalized "~7" as 6, 7, or 8.

comma insertions outnumbering comma deletions is inverted under the conditions [+…adverbial] and [+…preposition]. In contrast, for the condition [+after prefield], there is no such reason. Additionally, it is safe to assume that the numbers in the first two rows of table 7 are not in random distribution ($\chi^2(1) = 5.0075$, $p = 0.02524$), notwithstanding that effect size is small ($\varphi = 0.264$).

Summarizing the results of the pilot study, the topological factor in isolation turns out to reverse the usual trend in comma faults from lacking commas towards superfluous commas. In other words: prefields as such appear to trigger non-standard commas. However, this finding is based on a very small sample and hence preliminary. It remains to be seen whether the results can be replicated in large-scale studies.

3.4 Discussion

The results presented in the previous section provide initial indications that syntax, more precisely: topology, has a role to play in the placement of *Vorfeldkommas*. That is, in part, the phenomenon might explain itself: prefields attract commas just because they are prefields. However, it is fair to ask why prefields in particular should have this effect. As Berg et al. (2020: 89) succinctly put it with reference to Hochstadt & Olsen (2016: 171): there are more syntactically distinguished spots in a German sentence than the right prefield boundary and no equivalent clustering of non-standard commas is reported for any of them (cf. also Bredel 2015: 209–210). Answering this question depends in an obvious way on which syntactic theory one favours more generally. In any case, the topological model is easily compatible with the CP-IP-VP tree structure assumed in Government & Binding theory (Wöllstein 2014: 156–159). More concretely, the specifier of CP corresponds to the prefield and C° corresponds to the position of the finite verb in declaratives. The right prefield boundary, then, is located between these two nodes. This boundary is a special one indeed: above it, nothing can be base generated, i. e. the underlying clause ends here. If there is anything psychologically real to the theory at all, I would not be surprised to find that it is this boundary in particular that is mentally represented and, among other things, guides language users in their comma decisions.

4 Conclusion

The present contribution indicated that the topological position of *Vorfeldkommas* plays a more independent role than previously assumed. However, as hinted at above, large scale studies are needed in order to replicate the results. This means, among other things, building real data bases, not just an Excel workbook. In the process, tokenization might be done in such a way that, eventually, every conceivable comparison can be drawn – even with non-commas anywhere in the sentence. For example, Python's Natural Language Toolkit seems to provide the possibility to keep spaces as tokens so that spaces after letters can be searched for.

Coming back to language teaching for a moment, the paper as a whole might create the impression that I am arguing for permitting or even prescribing a comma after every prefield. This impression is wrong. Although I would consider such a move syntactically motivated, it would not make sense in a syntactic punctuation system like Bredel's – since not every prefield needs at its right edge what she calls a subordination blocker (Bredel 2016: 27–28, cf. also Lindbüchl

2015). Gallmann (2019: 351) even states that the didactic explication of topology impedes *Vorfeldkommas*. So the way from topology to orthography is not straightforward and I do not intend to give any practical recommendations in this paper. Rather, I intend to highlight the fact that *Vorfeldkommas* are both, a didactic and a syntax-theoretical subject-matter, in particular to the extent that they can be given a theoretic account as sketched in the previous section. So not only should theoretical linguistics go to the class room (cf. Trotzke 2020) but also vice versa.

Acknowledgements

I thank the students who gave permission to use their term papers as data.

References

[Amtliche Regelung] (2018): *Deutsche Rechtschreibung. Regeln und Wörterverzeichnis. Aktualisierte Fassung des amtlichen Regelwerks entsprechend den Empfehlungen des Rats für deutsche Rechtschreibung 2016*. Mannheim.

Barthel, Hans & Erich Löffler (1976): Die Zeichensetzung in den Aufsätzen unserer Schüler. *Deutschunterricht* 29 (2/3), 100–112.

Berg, Kristian, Ursula Bredel, Nanna Fuhrhop & Niklas Schreiber (2019): Das Vorfeldkomma – Fehler von heute, Regel von morgen? *Der Deutschunterricht* 71 (4), 45–56.

Berg, Kristian, Ursula Bredel, Nanna Fuhrhop & Niklas Schreiber (2020): Was determiniert das Vorfeldkomma? Untersuchungen zur Verteilung einer nicht-standardisierten Kommatierung. *Linguistische Berichte* 261, 85–116.

Boettcher, Wolfgang (2016): Komma & Co unter dem Kooperationsprinzip. Interpunktionslernen im Kompetenzbereich ‚Schreiben und Sprachreflexion'. In Ralph Olsen, Christiane Hochstadt & Simona Colombo-Scheffold (eds.), *Ohne Punkt und Komma… Beiträge zu Theorie, Empirie und Didaktik der Interpunktion* (Transfer 10), 327–361. Berlin: RabenStück.

Bredel, Ursula (2015): Topologie und Orthographie. In Angelika Wöllstein (ed.), *Das topologische Modell für die Schule* (Thema Sprache – Wissenschaft für den Unterricht 19), 205–217. Baltmannsweiler: Schneider Hohengehren.

Bredel, Ursula (2016): Interpunktion. System und Erwerb. In Ralph Olsen, Christiane Hochstadt & Simona Colombo-Scheffold (eds.), *Ohne Punkt und Komma… Beiträge zu Theorie, Empirie und Didaktik der Interpunktion* (Transfer 10), 18–50. Berlin: RabenStück.

Bredel, Ursula & Hrvoje Hlebec (2015): Kommasetzung im Prozess. *Praxis Deutsch* 254, 36–43.

Dürscheid, Christa (2012): *Syntax. Grundlagen und Theorien*. 6. Aufl. Göttingen: Vandenhoeck & Ruprecht (UTB).

Gallmann, Peter (2019): Das topologische Modell. *Mitteilungen des Deutschen Germanistenverbandes* 66 (4), 344–352.

Hochstadt, Christiane & Ralph Olsen (2016): Zur Kommatierungskompetenz von Lehramtsstudierenden am Beispiel überflüssiger ‚Vorfeldkommas'. In Ralph Olsen, Christiane Hochstadt & Simona Colombo-Scheffold (eds.), *Ohne Punkt und Komma… Beiträge zu Theorie, Empirie und Didaktik der Interpunktion* (Transfer 10), 158–176. Berlin: Raben-Stück.

Kirchhoff, Frank & Beatrice Primus (2014): The architecture of punctuation systems. A historical case study of the comma in German. *Written Language & Literacy* 17 (2), 195–224.

Lindbüchl, Isabell (2015): Ein Komma für den Leser. Sprachverarbeitung und Interpunktion im Deutschen, Englischen und Französischen am Beispiel des Kommas. *Bavarian Working Papers in Linguistics* 4, 69–84.

Müller, Hans-Georg (2007): *Zum „Komma nach Gefühl". Implizite und explizite Kommakompetenz von Berliner Schülerinnen und Schülern im Vergleich* (Theorie und Vermittlung der Sprache 50). Frankfurt: Peter Lang.

Müller, Stefan (in prep.): *German clause structure. An analysis with special consideration of so-called multiple frontings* (Empirically Oriented Theoretical Morphology and Syntax). Berlin: Language Science Press. https://hpsg.hu-berlin.de/~stefan/PS/german-sentence.pdf.

Naumann, Carl Ludwig (1995): Interpunktions-„Fehler". Welchen Regeln folgen die SchreiberInnen bei der Kommasetzung? In Petra Ewald & Karl-Ernst Sommerfeldt (eds.), *Beiträge zur Schriftlinguistik. Festschrift zum 60. Geburtstag von Prof. Dr. phil. Habil. Dieter Nerius* (Sprache – System und Tätigkeit 15), 211–233. Frankfurt: Peter Lang.

Neef, Martin (2019): Eine beschränkungsbasierte Modellierung der Kommasetzung im Deutschen. Presentation at *Interpunktion International*, University of Regensburg.

Pießnack, Christian & Adelbert Schübel (2005): Untersuchungen zur orthographischen Kompetenz von Abiturientinnen und Abiturienten im Land Brandenburg. *LLF-Berichte* 20, 50–72. https://publishup.uni-potsdam.de/opus4-ubp/solrsearch/index/search/searchtype/collection/id/16357

Primus, Beatrice (1993): Sprachnorm und Sprachregularität: Das Komma im Deutschen. *Deutsche Sprache* 21, 244–263.

Primus, Beatrice (2007): The typological and historical variation of punctuation systems. Comma constraints. *Written Language & Literacy* 10 (2), 103–128.

Rössler, Paul (2017): Nicht kodifizierte Kommasetzungsprinzipien nach Vorfeld. In Kristin George, Miriam Langlotz, Urania Milevski & Katharina Siedschlag (eds.), *Interpunktion im Spannungsfeld zwischen Norm und stilistischer Freiheit. Sprachwissenschaftliche, sprachdidaktische und literaturwissenschaftliche Perspektiven* (Medien – Literaturen – Sprachen in Anglistik/Amerikanistik, Germanistik und Romanistik 24), 63–94. Frankfurt: Peter Lang.

Sappock, Christopher (2011): *Das deutsche Komma im Spiegel von Sprachdidaktik und Prosodieforschung. Forschungslage – „Parsing vs. Phrasing" – Experimente* (Studien zur Linguistik 17). Berlin: LIT.

Trotzke, Andreas (2020): Pädagogische Linguistik – jetzt! *Linguistische Berichte* 261, 3–24.

Wöllstein, Angelika (2014): Topologisches Satzmodell. In Jörg Hagemann & Sven Staffeldt (eds.), *Syntaxtheorien. Analysen im Vergleich* (Stauffenburg Einführungen 28), 143–164. Tübingen: Stauffenburg.

Data

Deutsches Referenzkorpus, partially searchable online: https://cosmas2.ids-mannheim.de/cosmas2-web/.

Müller, Horst M. (2009). *Arbeitsbuch Linguistik. Eine Einführung in die Sprachwissenschaft* (UTB 2169). 2nd en. Paderborn et al: Schöningh UTB.

Resolution des 65. DHV-Tages in Mainz. Reproduced in *Handbuch für den wissenschaftlichen Nachwuchs 2018*; 125.

[Data of the present study are available upon request.]