

Between complex predicates and regular phrases: German collocational clusters

Philippa Cook

Freie Universität Berlin

Proceedings of the 21st International Conference on
Head-Driven Phrase Structure Grammar

University at Buffalo

Stefan Müller (Editor)

2014

CSLI Publications

pages 48–62

<http://csli-publications.stanford.edu/HPSG/2014>

Cook, Philippa. 2014. Between complex predicates and regular phrases: German collocational clusters. In Müller, Stefan (Ed.), *Proceedings of the 21st International Conference on Head-Driven Phrase Structure Grammar, University at Buffalo*, 48–62. Stanford, CA: CSLI Publications.



Abstract

I argue for a new type of non-standard constituent in German; a *modifier-collocational-cluster*. This type of cluster combines (i) a modifier and (ii) a PP from a light-verb construction (or a Funktionsverbgefüge (FVG) as they are known in German) or a bare noun. Such strings are found in German in initial (prefield) position in certain cases of apparent multiple fronting. We are dealing with a syntax-semantics mismatch here since the modifier does not semantically modify the element with which it can first syntactically combine. I show that the modifier is a collocate of both its co-prefield element but also of the verb. I propose a schema which lexically licenses the building of such clusters and I show how we can encode information about what I refer to as collocational selection in the lexical entries of the type of lexemes involved in these multi-word strings. The analysis can be seen as lexical but does not require lexical storage of phrasal elements.

1 Introduction

I propose a new analysis of certain multi-word strings in German such as (i) *heftig in die Kritik geraten* 'to be heavily criticised', (ii) *weltweit für Aufregung sorgen* 'to cause worldwide concern' or (iii) *richtig Geld verdienen* 'to make real money', postulating units I will call (*modifier-*)*collocational chunks*. The strings in (i) and (ii) involve a (semi-compositional) support verb construction, cf. Krenn & Erbach (1994), Steinitz (1989), *in die Kritik geraten* (literally: into the criticism fall) or *für Aufregung sorgen* (literally: for excitement provide) with modification by an adverbally used adjective, *heftig* 'harsh(ly)' or *weltweit* 'worldwide' respectively.¹ In (iii) a verb *verdienen* 'earn' selects a bare noun, and there is again modification by an adverbally used adjective *richtig*.² I argue that these strings are lexically encoded as multi-word expressions but we will see that this does not mean they have to be stored as phrasal entries. They are, I believe, situated on a continuum between genuine complex predicates at one extreme of the spectrum and canonically composed syntactic phrases at the other extreme. I take these lexical strings to

[†]Thanks to Felix Bildhauer, Stefan Müller, John Payne and Elodie Winckel as well as the audience and anonymous reviewers of HPSG 2014 (especially Farrell Ackerman, Anne Abeillé, Berthold Crysmann, Ray Jackendoff and Bob Levine) for comments and suggestions. Thanks also to the audience of the *MehrWortverbindungen* conference at the University of Basel in October 2014. This Research was funded by a DFG grant to project A6 of the SFB 632.

¹The part of speech *adjective* can be used in German as a pre-nominal modifier or predicatively but **also** in the function of an adverb with no morphological difference. A word-form such as *richtig* is therefore ambiguous in isolation. This ambiguity is undoubtedly a central contributing factor in the licensing of what I am calling *modifier-collocation-clusters*.

²In the data I will be discussing, *richtig* functions as an intensifier rather than as the manner adverb 'correctly'. Since *richtig* has a dual status (manner adverb or intensifier), the string is in principle ambiguous. There are similar, but less compositional, strings for which the manner reading is much less salient than the intensifier reading, e.g. *richtig Gas geben* 'to increase effort/to really go for it', lit: really give gasoline, viz. example (1d) below. Note, *geben* is 2-place here and clearly semantically bleached.

be exemplific of several larger classes of data patterning similarly, although with small differences across subclasses (and not all involving modifiers). For reasons of space I cannot document the full array of data here but see the comments in § 5 below.

The paper is structured as follows: In section 2, data is introduced which suggests we may have to accept a non-standard type of constituent in German, licensed only in the presence of certain combinations of lexical material. In section 3, the collocational relationships, cf. Firth (1957), Sinclair (1991, 1996), Evert (2008), spanning all three subcomponents of the string are discussed and in Section 4 it is proposed that sub-parts of such strings (namely the modifier and the PP/bare noun) may combine in German via a special schema for building collocational chunks, rather than building traditionally known syntactic constituents. The proposed schema is inspired by Function Composition known from Combinatorial Categorical Grammar (CCG). Although each of the three elements in the string is individually a syntactic atom of a multi-word string, the combination as a whole should be viewed as one complex lexeme, the building of which is licensed lexically.

2 Apparent cases of multiple fronting

German main clause declaratives are subject to the verb-second constraint; i.e. precisely one constituent may occur in the initial position preceding the finite verb (in a position referred to as the *prefield*. (1a)-(1d) instantiate (a certain type of) so-called *apparent multiple fronting* construction in which the clause-initial position before the finite verb contains a string that does not fit the traditional definition of constituent. Here, we have a modifier and a bare noun or PP. Semantically the modifier in initial position modifies the whole PP/N + V string. Syntactically, though, the modifier (surprisingly) combines with the PP or the N. Not only do we have a non-isomorphism of syntax and semantics (a syntax-semantics mismatch), but also a curious constituent structure.

- (1) a. [**Weltweit**] [**für Aufregung**] sorgt eine Werbekompagne von
 worldwide for upset provides an advertising-campaign from
 Benetton³
 Benetton
 ‘A Benetton advertising campaign is causing international concern’
- b. [**Heftig**] [**in die Kritik**] geriet der Kostenrechnungsbericht des
 heavy into the criticism fell the finance report the
 Jugendamtes für 2002⁴
 youth service for 2002
 ‘The youth service’s 2002 financial report got slated’

³<http://woodz.schwarzwaelder-bote.de/alltag/lifestyle/8422-benetton-zieht-kuss-foto-von-papst-zurueck.html>, checked 14.10.2014

⁴COSMAS, RHZ03/SEP.09166 Rhein-Zeitung, 12.09.2003

- c. **[Richtig] [Geld]** wird nur im Briefgeschäft verdient⁵
 right money is only in letter.business earned
 ‘You can only make real money with letters’
- d. **[Richtig] [Gas]** wird in der Großraum Disco ”Cocos Club” ab den
 right Gas will in the large-scale Disco Cocos Club from the
 Sa. 16.02.2008 gegeben⁶
 Sat. 16.02.2008 given
 ‘It’s going to be all-go in the large-scale Disco “Cocos Club” as of
 Saturday 16th February 2008’

This phenomenon has been documented by Müller (2003, 2005) who proposes an analysis in which the initial position houses a VP-constituent with an empty head (a structure that is used anyway in many approaches to German). I provide an alternative analysis for *apparent multiple fronting* data specifically of the type in (1) drawing on the concept of collocation.⁷ Support for the claim that strings such as *heftig in die Kritik* are **collocational clusters** (a string akin to some kind of *chunk/prefab*) can be gleaned from the observation that the material in the purported cluster prefers to permute (scramble) together rather than individually, viz.

- (2) a. weil **heftig in Kritik** der Bericht geriet
 because heavy in criticism the report fell
 ‘because the report got slated’
- b. ? weil **heftig** der Bericht **in Kritik** geriet
 because heavy the report in criticism fell
 intended: ‘because the report got slated’
- c. ? weil **in Kritik** der Bericht **heftig** geriet
 because in criticism the report heavy fell
 intended: ‘because the report got slated’

3 The collocational nature of the lexemes in the multiple fronting data

It has been noticed that the material in the initial string in constructions known as apparent *multiple fronting* intuitively forms a tight unit and that often at least one prefield element forms some kind of unit with the verb. We will see below that in the cases under discussion here, both prefield elements form a bond with the verb. Taking the strings from (1) above we can use a collocation association measure to ascertain whether or not this intuitive *unithood* can be verified. I employ the *Wortprofil 3.0* tool offered by the *Digitales Wörterbuch der Deutschen Sprache*

⁵taz 28./29.10.2000, p. 5, taken from Müller (2005)

⁶<http://www.my-nrw.de/nachtflug.php?kat=91&id=9806>, checked 14.10.2014

⁷Müller’s analysis covers a much broader range of data than discussed here. It is conceivable that Müller’s analysis could be retained and, for the type of data discussed here, be enhanced to include some kind of collocational analysis.

(DWDS) corpus, cf. Didakowski & Geyken (2013), which uses the LogDice measure of Rychlý (2008). Such lexicographically-oriented approaches to collocation use a notion of headword or node and examine collocates in a relation of dependence to one another such that e.g. in a modifier-noun collocation, the noun is headword and in a verb-object collocation, the verb is headword. The association measures for the lexemes mentioned, in the stated dependency relation, are given here.

	association using LogD	frequency
<i>heftig</i> as modifier of <i>Kritik</i>	11.12	9882
<i>Kritik</i> as object of <i>geraten</i>	9.27	2453
<i>heftig</i> as modifier of <i>geraten</i>	5.8	174

	association using LogD	frequency
<i>weltweit</i> as modifier of <i>Aufregung</i>	3.51	16
<i>Aufregung</i> as object of <i>sorgen</i>	9.13	3774
<i>international</i> as modifier of <i>sorgen</i>	4.41	107

	association using LogD	frequency
<i>richtig</i> as modifier of <i>Geld</i>	5.07	241
<i>Geld</i> as object of <i>verdienen</i>	11.51	22226
<i>richtig</i> as modifier of <i>verdienen</i>	6.09	332

A comprehensive study of the collocational behaviour of these tuples would warrant a separate paper but, for now, the measures suffice to illustrate that the intuitively perceived bond between the components of the string is statistically verified. In a further study, the collocation of complex strings (e.g. *Geld verdienen*, *in die Kritik geraten*) with the modifier will also be measured.⁸

4 Function Composition for collocational selection

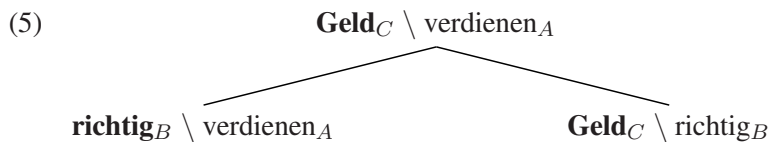
4.1 The spirit of Function Composition as a basis for the analysis

A solution to the syntax-semantics mismatch mentioned at the start of § 2 is the use of Function Composition (FC) instead of Functional Application to combine elements in syntax, cf. Jacobson (1990). Function Composition (FC) combines two functors to yield a new functor as sketched here:

⁸Annelies Häcki Buhofer suggested to me that modification of *geraten* by *heftig* does not seem semantically likely (in contrast to cases such as e.g. *richtig verdienen*) and Kathrin Steyer suggested that the modifier should only be considered as a collocate of the whole FVG, e.g. *in Kritik geraten* 'get criticised'. This needs to be more closely examined although I note for now that in the DWDS Corpus *heftig* clearly also collocates (as a modifier) with other forms related to *geraten* such as *aneinandergeraten* 'clash with one another/come into contact with one another'. That the modifier's scope extends across the whole FVG or the whole N+V string follows from my analysis although I only actually encode the modifier as a modifier of a verb.

- (3) Forward Function Composition: $A/B * B/C = A/C$
 (4) Backward Function Composition: $B \setminus A * C \setminus B = C \setminus A$

Forward FC allows A/B to combine with B/C yielding A/C ; a category requiring a C in order to be saturated. The *need* for a C at the initial level is postponed to the next level. Backward FC similarly postpones saturation (this time of A) to the next level. Within HPSG, Argument Inheritance draws on this type of combinatory rule, cf. Hinrichs & Nakazawa (1994) and much subsequent work on the licensing of verbal clusters. The spirit of Backward FC can be transferred to collocational cluster formation if we assume Backward FC can combine *richtig* + *Geld* (in bold-face below), postponing the "requirement" for *verdienen* 'earn'. By "requirement" for *verdienen*, I am referring to the modification domain of *richtig*; the modifier is actually (informally speaking) looking for the verb to modify but combines syntactically with a different element first. FC yields a *special* instantiation of *Geld* which can syntactically combine with the modifier and yet still requires the verb, as sketched here:



Below, I will show how the spirit of this type of syntactic combination could be captured in HPSG through a combination of lexical entries and a schema that licenses the type of cluster I am arguing for. Since HPSG makes no division between lexicon and syntax in the sense that lexical entries of words and rules of syntactic combination (schemata) are stored together, cf. e.g. Müller (2013, p. 8), Jackendoff (2010, p. 19f), this means the analysis is lexical and we can think of these multi-word expressions as being lexically stored. It is conceivable that particularly frequently co-occurring material is **also** stored as a (ready-built) chunk or prefab as well, cf. the notion of *conventionalized collocation* and *prefabs* discussed by (Bybee, 2006, p. 713-4, 727). Cases in which elements of the lexical string are non-contiguously realized (e.g. in multiple fronting, partial topicalization etc.) probably then involve a schema, as we propose below, since individual atoms of the string are aligned non-adjacent to other atoms.

4.2 Lexical Entries

I will take the string *richtig Geld verdienen* 'to make heaps' as an example throughout. The lexical entry for *richtig* in its function as an intensifier is given here:

$$(6) \left[\begin{array}{l} \text{word} \\ \text{PHON} \quad \langle \text{richtig} \rangle \\ \text{SS|LOC} \quad \left[\begin{array}{l} \text{CAT} \quad \left[\begin{array}{l} \text{HEAD} \quad \left[\begin{array}{l} \text{MOD} \quad V[\text{LID } \boxed{4}] \\ \text{LID} \quad \text{richtig-intensifier} \end{array} \right] \\ \text{SUBCAT} \quad \langle \rangle \end{array} \right] \end{array} \right] \\ \text{CONT} \quad [\text{intensify } \boxed{4}] \\ \text{COLL|LID} \quad \boxed{4} \text{verdienen-idiomatic} \end{array} \right]$$

I make use of the LID (lexical identifier) feature appropriate for the sort *head* to identify specific instantiations of words (Richter & Sailer (1999); Soehn (2004); Sag (2012); Spencer (2005)). Thus this word has the value *richtig_{intensifier}* for the feature LID in its lexical entry. The COLL feature (which I take to be appropriate for the sort *word* and *cluster*) encodes in the lexical entry of a word (or cluster) that it collocates with (the LID value of) a particular word (cf. Sailer (2003), Richter & Sailer (1999)). I refer to this as *collocational selection*. Thus we see here that *richtig_{intensifier}* collocates with the verb *verdienen* 'earn' (in its idiomatic instantiation). The intensifier is lexically encoded as a verb modifier (viz. the head feature MOD) and it also collocates with the verb it modifies (viz. the label 4 above). One could generalize the lexical entry so that the intensifier *richtig* always modifies the verb it collocationally selects if that turns out to be empirically correct.

I now give the lexical entry for *Geld* in the (semi-light-)verb phrase use:

$$(7) \left[\begin{array}{l} \text{word} \\ \text{PHON} \quad \langle \text{Geld} \rangle \\ \text{SS|LOC} \quad \left[\begin{array}{l} \text{CAT} \quad \left[\begin{array}{l} \text{HEAD} \quad [\text{LID } \text{Geld-idiomatic}] \\ \text{SUBCAT} \quad \langle \rangle \\ \text{SPR} \quad \langle \rangle \end{array} \right] \\ \text{CONT} \quad [\text{INDEX } \text{non-referential}] \\ \text{COLL|LID} \quad \text{richtig-intensifier} \end{array} \right] \end{array} \right]$$

The idiomatic bare noun is lexically encoded as a collocate of the intensifier *richtig*. I am also assuming the noun is lexically specified as non-referential (this is certainly the case for nouns such as *Gas* in *Gas geben*) and cannot take a specifier (i.e. must be saturated). A separate lexical entry in which the value of COLL|LID is *verdienen_{idio}* handles occurrences of the verb phrase *Geld verdienen* without *richtig*.⁹

The lexical entry for for the (semi-)light verb *verdienen* 'earn' is given next, below. The lexical entry would also be structured in the same way for a less-compositional (clearly) light verb such as *geben* 'give' (in e.g. *richtig Gas geben*):

⁹In the case of the bare-noun strings, we find frequent data such as the following which I think support the claims about collocation of *richtig* with *Geld* and *richtig* with *verdienen*:

- (i) *er hat richtig Geld* 'he is really rich' and
- (ii) *er verdient richtig* 'he earns loads'

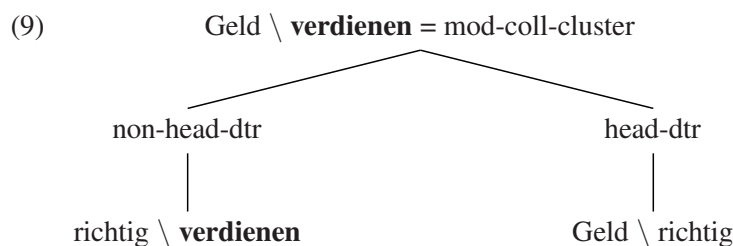
That the existence of such frequent strings (in particular the first one) facilitates apparent multiple fronting constructions seems highly plausible but remains to be further studied.

$$(8) \left[\begin{array}{l} \text{word} \\ \text{PHON} \quad \langle \text{verdienen} \rangle \\ \text{SS|LOC} \quad \left[\text{CAT} \left[\begin{array}{l} \text{HEAD} \quad [\text{LID } \textit{verdienen-idiotic}] \\ \text{SUBCAT} \quad \langle \text{NP-nom } \boxed{5} \rangle \end{array} \right] \right] \\ \text{CONT} \quad \left[\text{RELS} \quad \left\langle \left[\begin{array}{l} \textit{earn heaps} \\ \text{AGENT } \boxed{5} \end{array} \right] \right\rangle \right] \\ \text{COLL|LID} \quad \textit{Geld-idiotic} \end{array} \right]$$

This constitutes a new HPSG treatment of light verb phrases (or *Funktionsverbgefüge*, FVG) in German.¹⁰ The light verb *collocationally selects* (not subcategorizes) the (athematic) object but selects the subject NP in the normal way via SUBCAT. That this verb cannot undergo personal passive in the idiomatic use follows from the fact that there is no regular thematic argument other than the subject. Of course, impersonal passive (a subjectless construction, always requiring 3rd person singular verbal morphology, in German) is possible as we see in examples (1c-d) above). I am thus analyzing object-verb collocations involving bare nouns (e.g. *Geld verdienen* 'earn money') on a par with light verb phrases. I believe that this analysis can be extended to account for integrated objects (in the sense of Jacobs (1993, 1999) but cannot go into details here.

4.3 The *modifier-collocational-cluster* schema

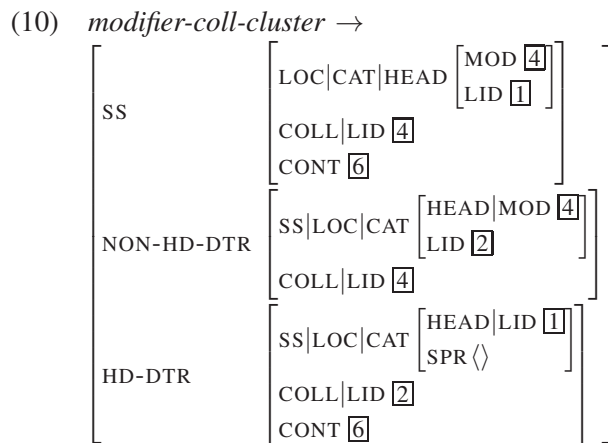
In this section, I introduce the schema which licenses the *modifier-collocational-clusters* such as e.g. *richtig Geld* [lit. real money]. The composition of the *modifier-collocational-cluster*, e.g. *richtig Geld*, with the idiomatic verb *verdienen* will then be shown next. First, recall the spirit of Backward FC. I indicate here the structures that I will be assuming in the schema. In particular, I assume here that the modifier is the non-head daughter and the bare noun is the head daughter of the cluster. This essentially translates the notion of headword or node from the lexicographically-oriented approach to collocation I mentioned earlier.



The modifier *richtig* has to 'wait' until it finds the verb it modifies. What I analyze as a *modifier-collocation-cluster* is therefore not a type of *head-adjunct-structure*

¹⁰Recent work on the processing of light verb phrases indicates increased processing load at the verb in light verb phrases. This effect can be interpreted as providing evidence that light verb phrases are not stored as complex (phrasal) entries but, rather, require some kind of syntactic combination or, perhaps, some operation involving argument-structure merging, cf. Wittenberg & Piñango (2011), Wittenberg et al. (2014). I believe my analysis is in keeping with these findings.

(since semantic modification does not occur here). This makes sense because the tuple is, I believe, in fact more like one complex lexeme. In fact, *richtig* in its intensifier function is not a normal modifier but somewhere between modifier and argument (= a collocational modifier). The *modifier-collocational-cluster* schema given below captures this:



The HD-DTR, e.g. *Geld*, collocationally selects the NON-HD-DTR, e.g. *richtig*. The NON-HD-DTR collocationally selects the verb it modifies. At the cluster level, the mother inherits the COLL|LID and MOD values from the NON-HD-DTR (= the postponement mentioned above). At the cluster level, the mother also inherits the CONT value of the HD-DTR; in keeping with the Semantics Principle. The cluster (mother) inherits the LID value from HD-DTR. In this way, the cluster can be seen as a special version of the lexeme *Geld-idiomatic*. The sub-tree for *richtig Geld* licensed by the *modifier-collocation-cluster* schema is given in Figure 1 on the following page.

The remaining question now is how the verb, in our case *verdienen* 'earn', combines with the collocational cluster *richtig Geld* 'real money'. In fact, the *modifier-collocation-cluster richtig Geld*, headed by *Geld*, collocationally selects the (idiomatic) verb *verdienen* but it also selects it via MOD. The *mod-coll-cluster* and the verb can combine via the normal *head-adjunct-schema*. The idiomatic semantics of the verb (encoded at the HD-DTR) percolate to the mother node and the (postponed) semantic modification of the verb can apply. The sub-tree for the combination of the cluster *richtig Geld* and *verdienen* is given in Figure 2 on the next page.

5 Extensions and further work

The analysis I have sketched here can, hopefully, be extended to handle a bigger range of data which behave similarly to those discussed here. In particular, there are certain lexical strings which offer an open slot which can be instantiated not just by lexically specified (collocating) material but which is, rather, open for any

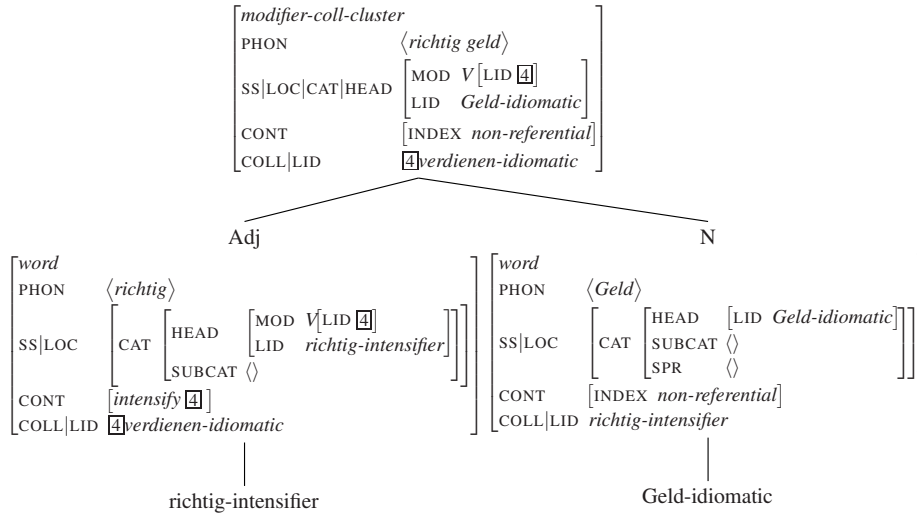


Figure 1: Sub-tree for the *modifier-collocational-cluster* *richtig Geld*

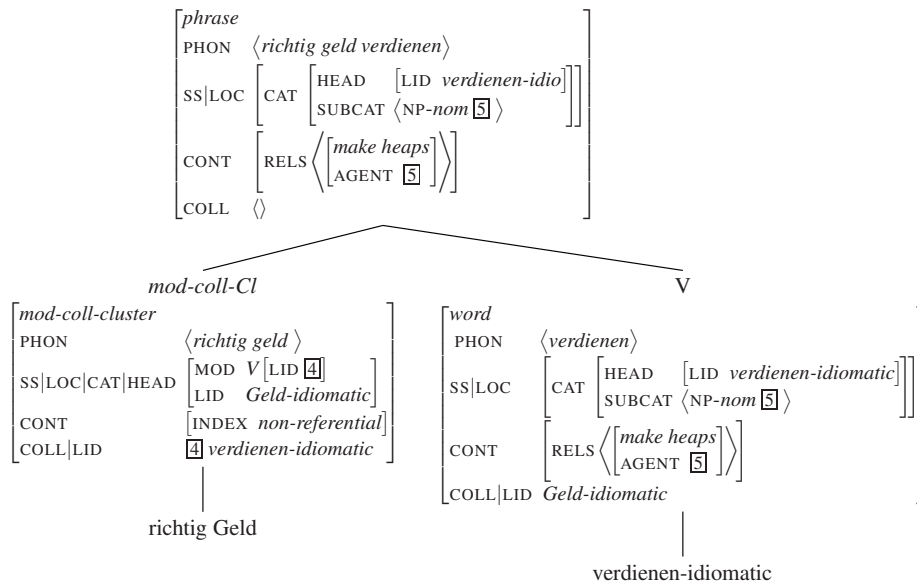


Figure 2: Sub-tree for the combination of the *modifier-collocational-cluster* *richtig Geld* with the verb *verdienen*

material of a particular class. For instance, the strings in the table below have all been attested with multiple fronting but range from fixed idioms, through collocating strings such as those discussed here to strings with slots for directional prepositional phrases, for instance. The strings vary in degrees of schematicity and form a continuum from full idioms to near-compositional phrases.

<i>Licht</i>	<i>ins Dunkel</i>	<i>bringen</i>	'bring light into the dark = shed light onto sth.'
<i>richtig</i>	<i>Gas</i>	<i>geben</i>	'really give Gas = increase effort'
<i>hart</i>	<i>ins Gericht</i>	<i>gehen</i>	'go hard into court = roast s.o.'
<i>ihm</i>	<i>zur Seite</i>	<i>stehen</i>	'stand by him'
<i>am billigsten</i>	<i>in XP</i>	<i>kommen</i>	'get to X the cheapest (way)'
<i>trocken</i>	<i>durch XP</i>	<i>kommen</i>	'come dry through X'
<i>postiv/negativ</i>	<i>auf XP</i>	<i>wirken</i>	'react positively/negatively to X'

It remains to be fully worked out how the range of data can be accommodated in the type of analysis proposed here.

A different consequence of the analysis proposed here concerns the possibility of topicalization of the collocational clusters for which I am arguing. It now seems plausible that this could be handled analogously to fronting of coherent verbal clusters, as in (11b), and could potentially offer an alternative analysis for (some) multiple fronting constructions:

- (11) a. [**richtig Gas**] gibt er immer
 right gas gives he always
- b. [**zu schlafen versucht**] hat er
 to sleep try has he
 'he tried to sleep'

Just as a string *zu schlafen versucht*_{verbal-cluster} can be realized in initial position, so could potentially a string *richtig Gas*_{modifier-collocational-cluster}. In fact, it is interesting to note that the availability of cluster formation discussed here could well be closely related to the availability of cluster-formation more generally in a given language (i.e. languages allowing verbal clusters may well be languages that allow other kinds of clusters too).

A further fascinating area is the extension of the current analysis to also cover (free) datives in the prefield, as in examples such as (12) where we have a free dative together with a PP belonging to a light verb phrase in the prefield:¹¹

- (12) [**Ihm**] [**zur Seite**] steht als stellvertretender Vorstandschef Gerd
 he-DAT to.the side stands as acting ceo Gerd

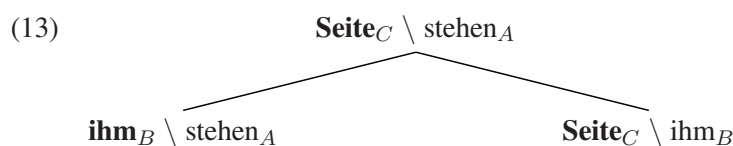
¹¹Examples of apparent *multiple fronting* involving a dative and an accusative object in the prefield are extremely rare. An analysis of argument-clusters as proposed by Mouret (2006) for co-ordination structures might be relevant but it would certainly over-generate as it stands since the material that can occur in the multiple fronting data is lexically very restricted.

Tenzer¹²

Tenzer

‘Gerd Tenzer is helping him out as acting CEO’

One can treat the dative as a benefactive modifier, addable to the argument-structure of any verb in German (e.g. by lexical rule). The dative is, however, also concomitantly possessor of the noun *Seite* ‘side’; i.e. it is also a modifier of the type which I assume to be introducible into the argument-structure of any noun. I informally sketch here how the FC-style analysis could be extended to cover such data:



6 Conclusion

I have argued here for a new type of cluster in German; a *modifier-collocation-cluster*. Clearly, we must extend the part-of-speech hierarchy accordingly to accommodate such elements. I believe introducing this type of cluster is a justified step, though. The analysis presented here has significance for our ideas about constituency and how it interacts with usage/frequency information, cf. Bybee & Cacoullos (2009); Beckner & Bybee (2009); Bod (1998), and also for the issue of the modifier-argument distinction. We know there is a close relation between frequently co-occurring elements and standard constituents but we must also capture units beyond those standardly acknowledged up to now, I firmly believe. Collocationally selected modifiers are situated inbetween arguments and true modifiers. The availability of what I have treated as collocationally selected items seems to generalize to form a pattern, to provide a slot fillable by material of a certain grammatical class (cf. Dowty (2003)). An extension of the current analysis to handle this kind of phenomenon is an exciting prospect. The analysis has, moreover, certain advantages for HPSG and specifically for the analysis of German. It interfaces usage data and a usage-based view of ‘constituency’ with the HPSG formalism. Further, it begins to capture the analogy between verb clusters (cluster – chunk) and the (non-standard) constituents for which I have argued in German. With some additional modification, it also offers the basis for a syntactic solution for handling Integration of nouns and PPs as discussed in (Jacobs, 1993, 1999).

References

Beckner, Clay & Joan Bybee. 2009. A usage-based account of constituency and reanalysis. *Language Learning* 59. 27–46.

¹²taz, 18.07.2002, p. 7, taken from Müller (2005)

- Bod, Rens. 1998. *Beyond grammar. an experience-based theory of language*. Stanford: CSLI Publications.
- Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language* 82(4). 711–733.
- Bybee, Joan L. & Rena Torres Cacoulos. 2009. The role of prefabs in grammaticization: How the particular and the general interact in language change. In Hamid Ouali Roberta L. Corrigan, Edith A. Moravcsik & Kathleen Wheatley (eds.), *Formulaic language: Volume 1. distribution and historical change*, 187 – 217. Amsterdam: John Benjamins.
- Didakowski, Jörg & Alexander Geyken. 2013. From DWDS corpora to a German word profile – methodological problems and solutions. In *Network strategies, access structures and automatic extraction of lexicographical information. 2nd work report of the academic network "internet lexicography" OPAL - Online publizierte Arbeiten zur Linguistik X/2012*, 43 – 52. Mannheim: Institut für deutsche Sprache.
- Dowty, Davis. 2003. The dual analysis of adjuncts and complements in categorial grammar. In Ewald Lang, Claudia Maienborn & Cathrine Fabricius-Hansen (eds.), *Modifying Adjuncts*, 1 – 22. de Gruyter.
- Evert, Stefan. 2008. Corpora and collocations. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics. an international handbook. article 58*, 1212 – 1248. Berlin: Mouton de Gruyter.
- Firth, J. R. 1957. A synopsis of linguistic theory 1930–55. In *Studies in linguistic analysis*, 1 – 32. Oxford: The Philological Society.
- Hinrichs, Erhard W. & Tsuneko Nakazawa. 1994. Linearizing AUXs in German verbal complexes. In Nerbonne et al. (1994) 11–38.
- Jackendoff, Ray. 2010. *Meaning and the Lexicon. The Parallel Architecture 1975–2010*. Oxford: Oxford University Press.
- Jacobs, Joachim. 1993. Integration. In Marga Reis (ed.), *Wortstellung und Informationsstruktur*, 63–116. Tübingen: Niemeyer.
- Jacobs, Joachim. 1999. Informational Autonomy. In Peter Bosch & Robert van der Sandt (eds.), *Focus. Linguistic, cognitive and computational perspectives*, 56–81. Cambridge University Press.
- Jacobson, Pauline. 1990. Raising as function composition. *Linguistics and Philosophy* 13. 423–475.
- Krenn, Brigitte & Gregor Erbach. 1994. Idioms and support verb constructions. In Nerbonne et al. (1994) 365–396.

- Mouret, François. 2006. A phrase structure approach to argument cluster coordination. In *Proceedings of the 13th international conference on Head-Driven Phrase Structure Grammar, Bulgarian Academy of Sciences, Bulgaria*, 247–267. Stanford: CSLI Publications. <http://csli-publications.stanford.edu/HPSG/2009/>.
- Müller, Stefan. 2003. Mehrfache Vorfeldbesetzung. *Deutsche Sprache* 31(1). 29–62.
- Müller, Stefan. 2005. Zur Analyse scheinbarer V3-sätze. In Franz-Josef d’Avis (ed.), *Deutsche Syntax: Empirie und Theorie. Symposium Göteborg 13–15 Mai 2004* (Göteborgger Germanistische Forschungen 46), 173–194. Acta Universitatis Gothoburgensis.
- Müller, Stefan. 2013. Unifying everything. *Language* 89(4). 920–950. <http://hpsg.fu-berlin.de/~stefan/Pub/unifying-everything.html>.
- Nerbonne, John, Klaus Netter & Carl J. Pollard (eds.). 1994. *German in Head-Driven Phrase Structure Grammar* (CSLI Lecture Notes 46). Stanford: CSLI Publications.
- Richter, Frank & Manfred Sailer. 1999. LF conditions on expressions of ty2: An HPSG analysis of negative concord in Polish. In Robert D. Borsley & Adam Przepiórkowski (eds.), *Slavic in Head-Driven Phrase Structure Grammar*, 247 – 282. Stanford: CSLI.
- Rychlý, Pavel. 2008. A lexicographer-friendly association score. In *Raslan 2008*, 6–9. Brno: Masarykova Univerzita.
- Sag, Ivan. 2012. Sign-based construction grammar: An informal synopsis. In Hans Boas & Ivan Sag (eds.), *Sign-based construction grammar*, 69 – 202. Stanford: CSLI.
- Sailer, Manfred. 2003. *Combinatorial semantics and idiomatic expressions in Head-Driven Phrase Structure Grammar*. Working Papers of the SFB 340, Vol. 161: Eberhard-Karls-Universität Tübingen Dissertation (2000).
- Sinclair, John. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sinclair, John. 1996. Beginning the study of lexis. In C. E. Bazell, J. C. Catford, M. A. K. Halliday & R. H. Robins (eds.), *In memory of J. R. Firth*, 410 – 430. London: Longman.
- Soehn, Jan-Philipp. 2004. License to coll: How to bind bound words and readings to their contexts. In Stefan Müller (ed.), *Proceedings of the 11th international conference on Head-Driven Phrase Structure Grammar, Center for Computational Linguistics, Katholieke Universiteit Leuven*, 261 – 273. Stanford: CSLI.

- Spencer, Andrew. 2005. Generalized paradigm function morphology – a synopsis. In Alexandra Galani & Beck Sinar (eds.), *York papers in linguistics*, vol. 2 Papers from the York-Essex Morphology Meeting, 2003, 93 – 106. University of York, UK.
- Steinitz, Renate. 1989. Überlegungen zum Prädikativ. In Wolfgang Motsch (ed.), *Wortstruktur und Satzstruktur*, 211 – 234. Berlin: Zentralinstitut für Sprachwissenschaft.
- Wittenberg, Eva, Martin Paczynski, Heike Wiese, Ray Jackendoff & Gina Kuperberg. 2014. The difference between "giving a rose" and "giving a kiss": Sustained neural activity to the light verb construction. *Journal of Memory and Language* 73. 31–42.
- Wittenberg, Eva & Maria Piñango. 2011. Processing light verb constructions. *Mental Lexicon* 6(3). 393–413.