German Verb Particle Constructions in CCG

Pablo Lopez Alonso
Department of Linguistics
Stony Brook University
Stony Brook, NY
al.pablo20@gmail.com

Abstract

As a theory of grammar, CCG (Steedman, 2000) is said to keep theory closely linked to psychological and computational mechanisms, to model how child or computer can learn any language (Steedman, 2017). In this paper, I test a Combinatory Categorial Grammar (CCG) against German Verb Particle Constructions (VPCs). Following previous work on English VPCs (Constable and Curran, 2009) and a German CCGbank (Hockenmaier, 2006), I analyze three types of German VPC constructions in CCG – main clauses, embedded clauses, and coordination clauses. Problems with modeling these sentences using CCG are discussed, and an alternative to deriving coordination sentences is presented in the Minimalist framework.

1 Introduction

An effective grammar formalism must represent the computational and psychological reality of natural language. To do so, it must be able to account elegantly and easily for natural language phenomena. Its mechanisms should correctly accommodate empirical data and make the correct predictions and rule out incorrect ones.

The central goal of this paper is to test Combinatory Categorial Grammar (CCG) (Steedman, 2000) against a set of compound verbs known as Verb Particle Constructions (VPCs), which present two possible orders:

(1) **Continuous Order**

a. The police **tracked down** the thief.

b. Anna **looked up** the book.

(2) **Discontinuous Order**

a. The police **tracked** the thief **down**.

b. Anna **looked** the book **up**.

VPCs have been extensively studied, but their syntactic status remains controversial (Dehé, 2015), (Haiden, 2017). They form a single semantic unit composed by two lexemes that show paradoxical behavior – they behave as both a word and a phrase. The analysis in this paper consists of two parts: first, we look at CCG's ability to deal with these verbs in German in simple main and embedded clauses. Then, we look at how effective this formalism is at dealing with a more complex structure, namely coordination. Crucially, we want CCG to deal with three types of VPC coordination – coordination of two different verb stems, coordination of one verb stem with two different particles, and dual valency coordination. This latter one involves a verb that is both a regular standalone verb and a verb that is part of a VPC.

The approach I follow here, borrows directly from Constable and Curran's (Constable and Curran, 2009) analysis of VPCs in CCGBank for English, as well as from Hockenmaier's (Hockenmaier, 2006) CCG Bank for German. I build on both approaches by adding Constable and Curran's new label for particles in VPCs to Hockenmaier's feature-rich German CCG to derive the correct German word order.

My proposal indicates that the approach used for English by Constable and Curran can account for some, but not all, German VPC sentences in CCG. Crucially, embedded sentences pose a challenge as VPCs in German have a different configuration in these constructions than they do in English. Furthermore, Across the Board Movement (ATB) in Minimalist Syntax is better able to capture dual valency VPC coordination constructions.

The paper proceeds as follows. Section 2 presents the preliminaries of German Syntax. Examples of VPCs in main and embedded clauses are presented, as well as those of coordination with VPCs with dual valency. Section 3 presents a brief

Proceedings of the 17th Meeting on the Mathematics of Language, pages 14–22, Montpellier (online), December 13, 2021.

introduction to CCG, as well as Hockenmaier's CCG rules for a German CCGBank. Section 4 delves into the analysis of VPC constructions in CCG. First, I summarize Constable and Curran's approach for VPCs in English, then I present my own analysis combining their analysis with Hockenmaier's. This section contains the bulk of the paper, and presents the problematic cases of VPCs for the proposed German CCG. In the last section, I discuss how Minimalist Syntax is better equipped to deal with cases of dual valency coordination.

2 German Syntax

2.1 Word order

German has three different word orders depending on the sentence type. Main clauses are verb second (V2) (3), embedded and relative clauses are verb final (4), and interrogatives and imperatives are verb-initial (5) (examples from (Hockenmaier, 2006)).

(3) a. *Peter gibt Maria das Buch*
   'Peter gives Mary the book.'
  b. *ein Buch gibt Peter Maria.*
  c. *dann gibt Peter Maria das Buch.*

(4) a. *dass Peter Maria das Buch gibt.*
  b. *das Buch, das Peter Maria gibt.*

(5) a. *Gibt Peter Maria das Buch?*
  b. *Gib Maria das Buch!*

For the purposes of this paper, we look only at VPCs in the context of V2 and verb-final sentences.

2.2 VPCs in German

VPCs in German must follow the same word order outlined in section 2.1:

(6) a. Ich stehe auf.
   I stand PRT
   'I stand up.'
  b. Ich höre dir zu.
   I hear you.DAT PRT
   'I listen to you.'

In a main clause (6), the verb appears in its usual V2 position and the particle remains stranded in the final position. However, in an embedded clause (7), the verb and particle form one lexical unit and both appear in the final position.

(7) a. dass ich aufstehe.
   that I PRT.stand
   'that I stand up.'
  b. dass ich dir zuhöre.
   that I you.DAT PRT.hear
   'that I listen to you.'

Embedded sentences, in particular, pose a challenge for dealing with VPCs in CCG as they display idiosyncratic behavior. I take this issue up in section 4.2.

3 CCG

3.1 Combinatory Categorial Grammar

Combinatory Categorial Grammar (Steedman, 2000) is a highly lexicalized grammar formalism. In CCG, every word is associated with a syntactic type made up of atomic categories and directional slashes. A category of the type X/Y is a functor that takes a type Y to the right and returns type X after application. CCG uses function application to combine different constituents; however, other types of functions are also available.

Application:

$$\sigma/\tau \quad \tau \quad \mapsto \quad \sigma$$
$$\tau \quad \sigma\backslash\tau \quad \mapsto \quad \sigma$$

Composition (B):

$$\sigma/\tau \quad \tau/\rho \quad \mapsto \quad \sigma/\rho$$
$$\tau\backslash\rho \quad \sigma\backslash\tau \quad \mapsto \quad \sigma\backslash\rho$$

Crossing Composition (B$_x$):

$$\sigma/\tau \quad \tau\backslash\rho \quad \mapsto \quad \sigma\backslash\rho$$
$$\tau/\rho \quad \sigma\backslash\tau \quad \mapsto \quad \sigma/\rho$$

Type-Raising (T):

$$\tau \quad \mapsto \quad \sigma/(\sigma\backslash\tau)$$
$$\tau \quad \mapsto \quad \sigma\backslash(\sigma/\tau)$$

As a theory of grammar, CCG keeps the syntax and semantics closely linked together. It also purports to be consistent with linguistic facts, keeping the theory as close as possible to computational and psychological mechanisms, allowing any language to be learned by both child and computer (Steedman, 2017).

CCG deals elegantly with certain linguistic phenomena like conjunction (Steedman and Baldridge, 2006). However, it has been noted that CCG Bank (Hockenmaier and Steedman, 2007), the main corpus for CCG-related work, which uses the formalism's combinatory rules, has several flaws. For example, it struggles to deal with complement/adjunct distinctions, compound nouns, and

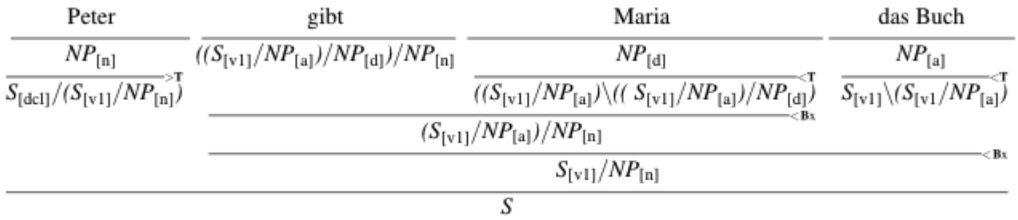

Figure 1: Standard main clause in Hockenmaier's German CCG

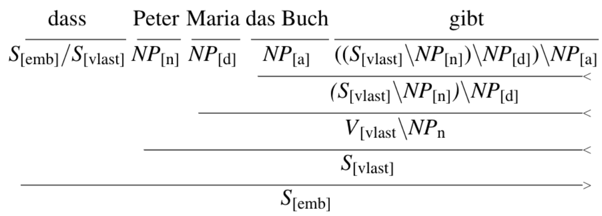

Figure 2: Embedded clause in Hockenmaier's German CCG

phrasal verbs or Verb Particle Constructions (Constable and Curran, 2009). I discuss this issue further in section 4.

3.2 A CCG for German

A first step to analyzing German VPCs in CCG is to capture the German word order. Hockenmaier (Hockenmaier, 2006) creates a CCG Bank for German, in which she advocates for augmenting CCG Bank with features to derive the correct word order (see Fig.1 and Fig.2 above). Features such as $S_{[v1]}$ and $S_{[vlast]}$ are introduced to capture the correct word order for V2 in main clauses and verb-final embedded clauses, for example. Other features such as $[n]$, $[a]$, and $[d]$ are introduced for nouns inflected for case.

Admittedly, contrary to Steedman (Steedman, 2000), Hockenmaier assumes that German is verb-initial (whence the $[v1]$ feature). This notion, however, runs counter to the literature on German verb headedness (Harbert, 2006), (Haider, 2010).

3.3 CCG Treatment of VPCs

CCGBank (Hockenmaier and Steedman, 2007), as mentioned in section 3.1, is the primary corpus for CCG-related work. It has been noted that it varies in its management of particles, but it tends to treat them as adverbial modifiers (Constable and Curran, 2009). This notion is problematic since particles are a core part of the VPC construction. Similarly, current German CCG tools do not seem to have a clear way to handle particles or VPCs (German CCG Bank, (Hockenmaier, 2006)) or they only deal with them in their sentence-final position (CCGWeb, (Evang et al., 2019)).

At least one attempt has been made to remedy this problem for English (Constable and Curran, 2009). Constable and Curran add a new atomic category *RP* to the existing set (*N, NP, S, PP*). This approach adds the particle directly into the verb's subcategorization and ensures the particle is a required part of the construction instead of an adverbial modifier. Admittedly, this model tends to over-generate, and it cannot rule out ungrammatical VPCs with a pronominal object in the split configuration (**she took away it*). Moreover, Constable and Curran observed a decrease in performance when parsing using this new mechanism. Nonetheless, the model is more consistent with linguistic facts than previous treatments of VPCs in CCG. As such, I take it up and use it to augment Hockenmaier's German CCG Bank rules in the next section.

My analysis is presented in contrast to Steedman's (personal communication, 2019) proposal that verbs such as *geben* 'give' and its VPC counterpart *zugeben* 'to admit' are "accidental homophones." Steedman assumes that these verbs are two distinct lexical units – a regular verb and a light-verb that combines with a particle. This proposal weakens the semantic connection between the two verbal constructions that intuitively we know exists (cf. *stehen* 'stand' v. *aufstehen* 'stand up').

4 German VPCs in CCG

In this section I analyze some VPC sentences in German CCG. First, I look at both intransitive and

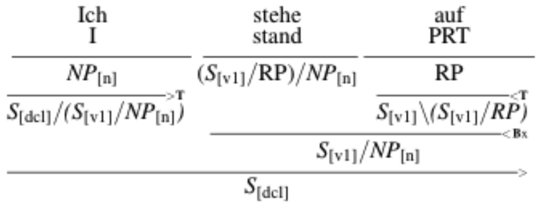

Figure 3: Main clause VPC sentence in German CCG

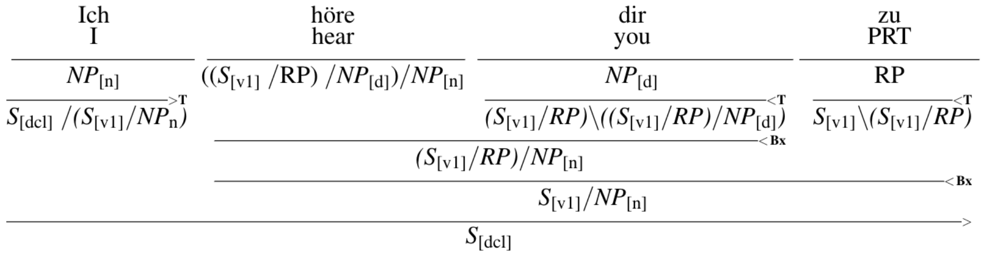

Figure 4: Transitive VPC construction in German CCG (dative object)

transitive VPCs in a main clause. Then, I move on to embedded sentences, and finally I look at examples of VPC coordination and verbs with dual valency.

4.1 Main Clauses

The combined approach I proposed in section 3.3, deals with intransitive sentences easily (Fig. 3). Following Hockenmaier's rules for German CCG Bank, the subject is type raised to be able to combine with the VPC. Similarly, here the Particle must type raise to combine with the verb. Unlike Constable and Curran (Constable and Curran, 2009), I have chosen to make the verb select for the particle first, and only then select for its other arguments. This decision was based on two facts: first, it allows us to more closely follow Hockenmaier's rules for German CCG derivations, and second, it captures the notion that the verb and particle form one unit despite their distance in the sentence.

This approach easily handles main clauses with transitive VPCs as well. Two examples are presented here, one with a dative object (Fig.4 above) and one with an accusative object (Fig.5 below).

Both examples follow the same template. First, the verb combines with its object (*accusative or dative*) through crossing composition, then with the particle through the same process. Lastly, the subject selects for the rest of the sentence, giving a declarative sentence as a result.

4.2 Embedded Clauses

Dealing with embedded VPC clauses in CCG presents the biggest challenge. As mentioned in section 2.2, VPCs in German show as a single lexical unit in the final position in embedded clauses. Since we know in a main clause the verb behaves as two separate lexical units, this leaves us with two options in CCG.

First, we could assume that the verb is one word only (Fig.6). In this approach we switch the functionality of the functor and make the verb select for the subject to its left. This is in line with what Hockenmaier (Hockenmaier, 2006) does for embedded clauses as well.

The second option is to split the verb into the particle and verb stem, much like we have done for the main clause (Fig. 7).

None of the previous approaches are without fault. The first approach requires that we stipulate that the main clause VPC and the embedded clause VPC have two different lexical categories. This is problematic as it increases the category ambiguity of the words by introducing a new category for each instance of the verb. In principle, CCG opposes this type of variation as it prefers to handle those differences by using combinatory rules.

The second approach allows us to keep the category of the verb consistent across main and embedded clauses by keeping the verb-particle split. However, in this instance, there is no mechanism that would prevent us from allowing adjuncts to come between the verb and the particle. It overgenerates

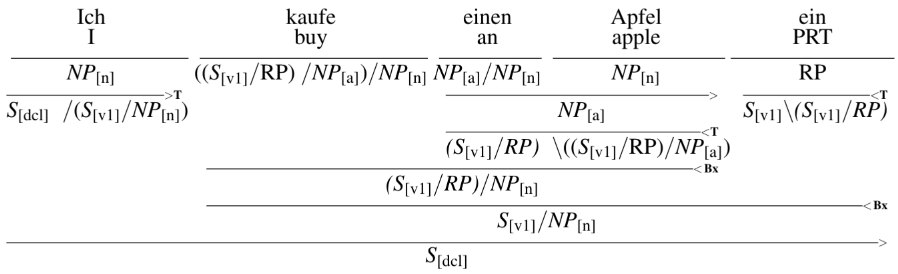

Figure 5: Transitive VPC construction in German CCG (accusative object)

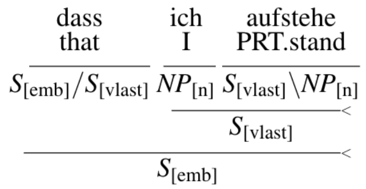

Figure 6: Embedded clause with VPC as a single lexical unit

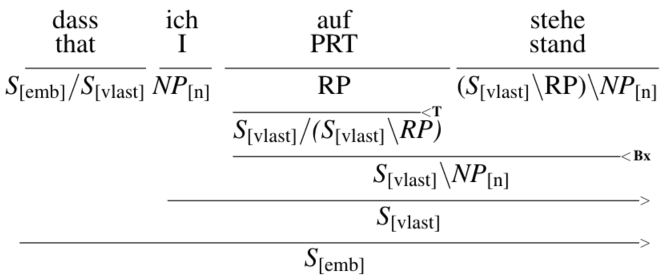

Figure 7: Embedded clause with VPC as a phrase

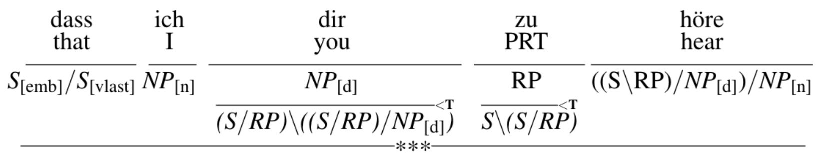

Figure 8: Failed CCG derivation for an embedded transitive VPC

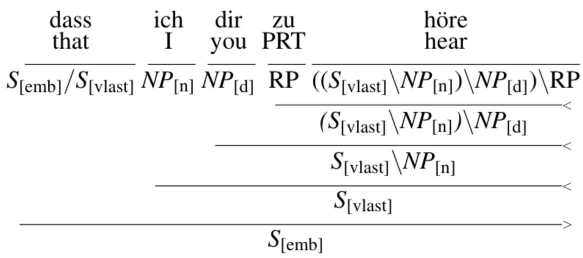

Figure 9: Fully derived VPC CCG derivation; verb category has changed

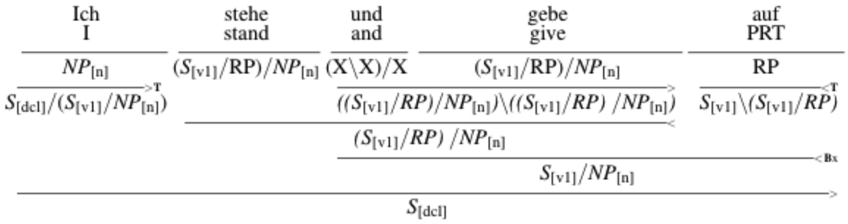

Figure 10: CCG derivation for particle sharing conjunction

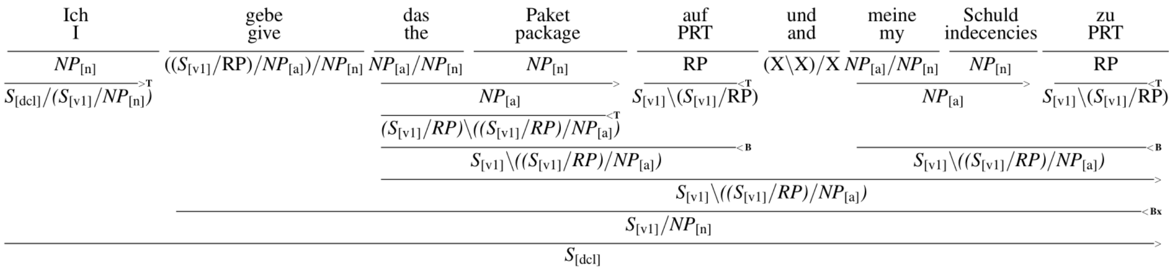

Figure 11: CCG derivation for verb sharing conjunction

for ungrammatical words like **aufgesternstehen* 'PRT-*yesterday*-stand'. At the moment, I do not have a solution for this problem, but if solved it would make CCG able to handle VPCs in embedded sentences.

Transitive VPCs in embedded clauses present even more of a challenge (Fig.8). In addition to the issue of whether the verb should be split from the particle or not, the verb has to select for an object.

It is impossible to derive this sentence and keep the verb category consistent with that of the main clause. Here again, changing the category assigned to the verb could give us a fully derived sentence. However, as mentioned above this is undesirable as it increases category ambiguity, thus assuming that the main clause verb and the embedded clause verb are two different lexical items. I present one possible derivation here for illustration purposes (Fig.9). Note, however, that this is an undesirable CCG derivation within our framework.

4.3 Coordination

VPC coordination presents even more of a challenge for CCG. Three types of sentences were analyzed for coordination – particle sharing conjunction, verb sharing conjunction, and dual valency conjunction. The first type of coordination involves two different verb stems in each conjunct associated with the same particle. The second type of coordination involves a verb stem that is associated with two different particles. The last type is a clause where the verb acts as both an intransitive and transitive across the conjuncts.

Particle sharing instances are easily handled (Fig.10). The verbs combine with one another first, then with the particle, and last with the subject. One elegant feature of this type of derivation is that the verb types remain the same as those for the regular main clause sentences, thus keeping the design simplified for the formalism.

Instances of verb sharing are handled much like Steedman and Baldridge (Steedman and Baldridge, 2006) handle some instances of Across The Board (ATB) movement (Fig. 11). First, we combine the two object conjuncts, and then we compose them with the verb. Last, by application we derive the full declarative sentence.

Once again, this example allows us to keep the category of the verb $((S_{[v1]}/RP)/NP_{[a]})/NP_{[n]}$ consistent with that of the main clause verb keeping the formalism simple.

Although CCG handles verb and particle sharing easily, it is unable to handle dual valency coordination at all (Fig.12). Crucially, the verb stem needs to select for an object for one conjunct, but for no object for the other. That is to say, the verb needs to be both transitive and intransitive. There is no formal mechanism to assign the verb two different categories.

As a transitive VPC, the verb must have the category $((S_{[v1]}/RP)/NP_{[a]})/NP_{[n]}$. However, as an intransitive verb, it needs the category $S_{[v1]}/RP/NP_{[n]}$. Since CCG deals with these surface variations only through combinatory rules,

Ich	gebe	auf	und	meine	Schuld	zu
I	give	PRT	and	my	indecencies	PRT
$NP_{[n]}$	$((S_{[v1]}/RP)/NP_{[a]})/NP_{[n]}$	RP	$(X\backslash X)/X$	$NP_{[a]}/NP_{[n]}$	$NP_{[n]}$	RP
			***			

Figure 12: Failed CCG derivation for dual valency conjunction

there is no way to assign both valencies to the verb, and the derivation is impossible.

As a grammar formalism, CCG is unable to handle VPCs in a cohesive and consistent manner. In this section I have highlighted only two problems that are readily apparent – embedded clauses and coordination. Embedded clauses present two problems. In their intransitive configuration the verb has to be categorized as a lexical unit with the particle or as a phrasal one. This problem is not unique to CCG and remains unresolved in the literature (Haiden, 2017; Dehé, 2015). However, in contrast to other analyses, both approaches are too powerful. The lexical approach is too restrictive and presents no mechanism for verb inflection. The phrasal approach is too permissive in that it allows for ungrammatical derivations of the verb. The second problem of German VPCs in CCG is that of coordination. Intransitive and transitive coordination fare decently under the formalism, however, dual valency coordination is impossible. The coordination problem receives a simpler analysis under the Minimalist Program (Chomsky, 1995). In the next section I discuss those analyses as an alternative to CCG.

5 VPC Coordination in Minimalism

Within the Minimalist Program (Chomsky, 1995), VPCs have received considerable attention (cf. (Haiden, 2017) for a general review and (Dehé, 2002) for VPCs in English). The syntactic status of VPCs as either complex heads or small clauses remains unresolved. Wurmbrand (Wurmbrand, 2000) provides an analysis that splits the complex head/small clause debate along semantic lines and which I follow here. The coordination problem, however, remains, to my knowledge, unadressed. As such, in this section, I make use of Wurmbrand's approach to show how Minimalism is better equipped to deal with the coordination cases mentioned in section 4.3.

Wurmbrand's analysis splits VPCs into transparent and idiomatic VPCs. Transparent VPCs receive a small clause treatment, while idiomatic ones receive a complex head treatment. This analysis

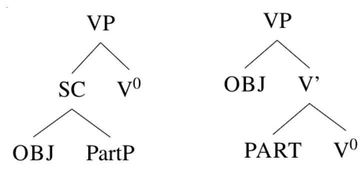

Figure 13: Transparent VPC (left) vs. (Semi-)idiomatic VPC structures (right)

easily accounts for different properties of VPCs in Germanic languages such as topicalization (Fig.13)

The transparent/idiomatic distinction is not always clear-cut, however, and as such is largely ignored here. Note that for our purposes this distinction does not affect the analysis proposed here. Under Wurmbrand's analysis a main clause VPC looks as follows.

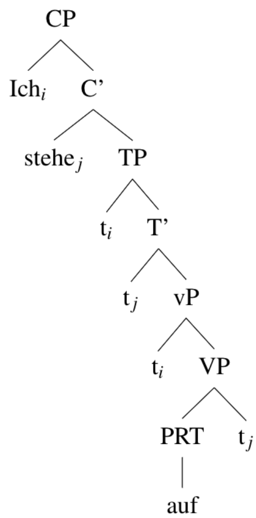

Figure 14: Minimalist derivation for *Ich stehe auf* 'I stand up'

I follow Haider (Haider, 2010) in assuming that no functional heads in German are head-final, only the verb is head-final. Canonically, the verb raises to the T head before reaching its final position in the C head for its V2 position in main clauses. The subject must then raise to Spec C in order to derive the correct word order. Crucially, the particle remains in its base position giving rise to particle stranding in German VPCs.

Coordination cases are easily accommodated by this analysis (Fig.15). For intransitive cases of particle sharing, the derivation starts with the VPC conjunction *aufgebe und aufstehe*. Through ATB movement the particle moves to a position above the coordination and the verbs must move out of the coordination to reach their V2 position as well. Just how this mechanism would fully work is out of the scope of this paper, but it would follow some type of sideward movement like that proposed in (Torr and Stabler, 2016).

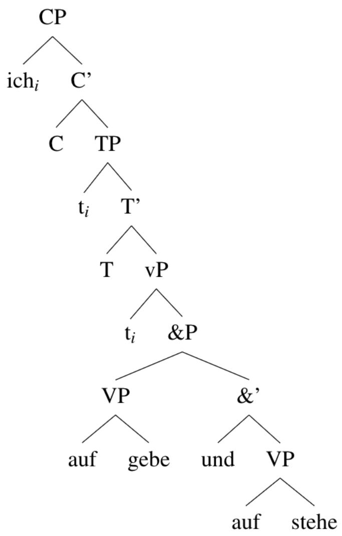

Figure 15: Verb stem coordination in Minimalism

Crucially, Minimalism has no difficulty in dealing with cases of dual valency mentioned in section 4.3.

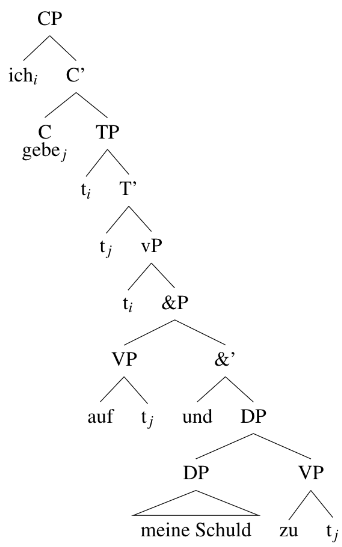

Figure 16: Dual valency coordination

The derivation starts with the conjunction of the two VPCs *aufgeben* and *meine Schuld zugeben*. Through ATB the verb follows its canonical movement to the T head and then to the C head to land in its V2 position. This operation results in the correct word order and the fully derived sentence. ATB in Minimalism avoids the problem CCG has of assigning two different categories to the same lexical item and thus creating ambiguity in the formalism. It remains to be seen if the minimalist approach would overgenerate, however, I leave that work for future research.

6 Conclusion

The account proposed in this paper draws from Constable and Curran (Constable and Curran, 2009) and Hockenmaier (Hockenmaier, 2006) to derive an analysis for German VPCs in CCG. The analysis presented here is able to accommodate VPCs in their main clause configuration, but fails at deriving the embedded clause counterparts. CCG is unable to capture distinctions such as the complex head/small clause analysis that has been proposed for VPCs in the literature. We are forced to choose one of the two analyses and the formalism becomes too powerful. Either it overgenerates for ungrammatical sentences under the small clause account or extremely restricts the verb under the complex head account, disallowing verbal inflection. In instances of embedded clauses with transitive VPCs, CCG seems completely unable to derive these clauses as the combinatory rules do not allow any possible combination of the constituents. Steedman's (p.c.) account for VPCs in CCG is equally unsatisfactory as it hinges on the idea that the verb in the VPC is a light-verb equivalent of a regular verb. At best, this analysis weakens a semantic connection between the verbs and at worst it completely denies it. Additionally, as a language formalism, CCG is unable to deal with examples of coordination such as dual valency.

While the question of the syntactic status of VPCs remains open, it is clear that as a theory of grammar, CCG offers no new insights to this matter. If we are to embrace this formalism as a theory of grammar that enables any child or computer to learn any language, further work needs to occur to refine the formalism in a manner that can better account for syntactic constructions such as the ones examined here.

References

Noam Chomsky. 1995. *The Minimalist Program*. MIT Press, Cambridge, Massachussets.

James Constable and James Curran. 2009. Integrating verb-particle constructions into CCG parsing. In *Proceedings of the Australasian Language Technology Association Workshop 2009*, pages 114–118, Sydney, Australia.

Nicole Dehé. 2002. *Particle Verbs in English: Syntax, Information Structure, and Intonation (Linguistik Aktuell/Linguistics Today 59)*.

Nicole Dehé. 2015. Particle verbs in germanic. In Peter O. Müller, editor, *Word-Formation : an International Handbook of the Languages of Europe ; vol. 1*, number 40,1 in Handbücher zur Sprach- und Kommunikationswissenschaft, pages 611–626. De Gruyter Mouton, Berlin; Boston.

Kilian Evang, Lasha Abzianidze, and Johan Bos. 2019. CCGweb: a new annotation tool and a first quadrilingual CCG treebank. In *Proceedings of the 13th Linguistic Annotation Workshop*, pages 37–42, Florence, Italy. Association for Computational Linguistics.

Martin Haiden. 2017. *Verb Particle Constructions*. The Wiley Blackwell Companion to Syntax.

Hubert Haider. 2010. *The Syntax of German*. Cambridge Syntax Guides. Cambridge University Press.

Wayne Harbert. 2006. *The Germanic Languages*. Cambridge Language Surveys. Cambridge University Press.

Julia Hockenmaier. 2006. Creating a CCGbank and a wide-coverage CCG lexicon for German. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 505–512, Sydney, Australia. Association for Computational Linguistics.

Julia Hockenmaier and Mark Steedman. 2007. CCGbank: A corpus of CCG derivations and dependency structures extracted from the Penn Treebank. *Computational Linguistics*, 33(3):355–396.

M. Steedman and Jason Baldridge. 2006. Combinatory categorial grammar. *Encyclopedia of Language Linguistics*, 2.

Mark Steedman. 2000. *The Syntactic Process*. MIT Press, Cambridge, MA, USA.

Mark Steedman. 2017. Combinatory categorial grammar: An introduction.

John Torr and Edward P. Stabler. 2016. Coordination in Minimalist Grammars: Excorporation and across the board (head) movement. In *Proceedings of the 12th International Workshop on Tree Adjoining Grammars and Related Formalisms (TAG+12)*, pages 1–17, Düsseldorf, Germany.

Susi Wurmbrand. 2000. The structure (s) of particle verbs. *Ms., McGill University*.