German Verb Particle Constructions in CCG

MALIN / MoL 2021

Pablo Lopez Alonso

Stony Brook University
al.pablo20@gmail.com

Dec 13, 2021

The Big Picture

- Syntax is the mediator between **meaning** and **sentences**.
- How do we model that mediaton?
- Can our model **simply**and **elegantly** account for typological differences?
- Or does it break down?

The project

- **CCG** is one of several language formalisms that try to do that.
  its claim to fame being its ability to deal with conjunction elegantly.
- Test CCG against a language phemnomenon to see if it can consistently deal with it.
- Point out the problems found, alternatives, and direction to pursue.

Outline

1. Background
   - What are Verb Particle Constructions?
   - German Word Order and VPCs in German
2. Introduction
   - CCG as a framework for language modelling
   - English VPCs in CCG
   - Current approaches
3. German (VPCs) in CCG
   - Deriving German Sentences in CCG
   - Deriving German VPCs in CCG
4. VPC Coordination
   - CCG
   - Minimalism
5. Conclusion & work to do

Background

- Verb Particle Constructions (VPCs) are a phenomenon unique to the **Germanic** languages. (Dehé 2002; Harbert 2006)
- Compound verbs that consist of a **verb** and a preposition-like **particle**
- In English, also known as Phrasal Verbs

Two possible orders

- **Continuous**

  (1) a. The police **tracked down** the thief.
  b. Anna **looked up** the book.

- **Discontinuous**

  (2) a. The police **tracked** the thief **down**.
  b. Anna **looked** the book **up**.

Background

- Verb Particle Constructions (VPCs) are a phenomenon unique to the **Germanic** languages. (Dehé 2002; Harbert 2006)
- Compound verbs that consist of a **verb** and a preposition-like **particle**
- In English, also known as Phrasal Verbs

Two possible orders

- **Continuous**

  (1) a. The police **tracked down** the thief.
  b. Anna **looked up** the book.

- **Discontinuous**

  (2) a. The police **tracked** the thief **down**.
  b. Anna **looked** the book **up**.

Background cont.: German Word Order

- German also has **VPCs**!
- But German is a **V2** language in matrix clauses

(3) a. *Peter* ***gibt*** *Maria das Buch.*
P gave M the book

b. *ein Buch* ***gibt*** *Peter Maria.*
a book gave Peter Maria

c. *dann* ***gibt*** *Peter Maria ein Buch.*
then gave P M a book
'Peter gave Maria the book.'

- Relative and embedded clauses are verb final

(4) *dass Peter Maria das Buch* ***gibt.***
that P M a book gave
'that Peter gave Maria the book.'

- Rich inflectional morphology

Background cont.: German Word Order

- German also has **VPCs**!
- But German is a **V2** language in matrix clauses

(3) a. *Peter* ***gibt*** *Maria das Buch.*
P gave M the book

b. *ein Buch* ***gibt*** *Peter Maria.*
a book gave Peter Maria

c. *dann* ***gibt*** *Peter Maria ein Buch.*
then gave P M a book
'Peter gave Maria the book.'

- Relative and embedded clauses are verb final

(4) *dass Peter Maria das Buch* ***gibt.***
that P M a book gave
'that Peter gave Maria the book.'

- Rich inflectional morphology

Background cont.: German Word Order

- German also has **VPCs**!
- But German is a **V2** language in matrix clauses

(3) a. *Peter* ***gibt*** *Maria das Buch.*
P gave M the book

b. *ein Buch* ***gibt*** *Peter Maria.*
a book gave Peter Maria

c. *dann* ***gibt*** *Peter Maria ein Buch.*
then gave P M a book
'Peter gave Maria the book.'

- Relative and embedded clauses are verb final

(4) *dass Peter Maria das Buch* ***gibt***.
that P M a book gave
'that Peter gave Maria the book.'

- Rich inflectional morphology

Background cont.: German Word Order

- German also has **VPCs**!
- But German is a **V2** language in matrix clauses

(3) a. *Peter* ***gibt*** *Maria das Buch.*
P gave M the book

b. *ein Buch* ***gibt*** *Peter Maria.*
a book gave Peter Maria

c. *dann* ***gibt*** *Peter Maria ein Buch.*
then gave P M a book
'Peter gave Maria the book.'

- Relative and embedded clauses are verb final

(4) *dass Peter Maria das Buch* ***gibt***.
that P M a book gave
'that Peter gave Maria the book.'

- Rich inflectional morphology

VPCs in German

- German VPC in a main clause

(5) *Ich* ***rufe*** *Angela* ***an.***
I call Angela PRT
'I am calling Angela (up)'

- German VPC in a main clause with a fonted adjunct

(6) *dann* ***rufe*** *ich Angela* ***an.***
then call I Angela PRT
'then I call Angela (up)'

- German VPC in an embedded clause

(7) *dass ich Angela* ***anrufe.***
that I Angela PRT.call
'that I am calling Angela (up)'

CCG

- Combinatory Categorial Grammar (Steedman 2017) is a **highly lexicalized** grammar formalism
- aims to keep **syntax** and **semantics** as closely linked as possible.
- Every word is associated with a **syntactic type** made out of **directional slashes**

  Category X/Y $\mapsto$ X after taking a Y to its right

CCG Cont.

- keeps theory as close to psychological and computational mechanisms

sentences $\mapsto$ **meaning** $\mapsto$ **sentences**

so that "the child or computer can learn the grammar for any human language" Steedman (2017)

Is CCG **THE** language model?

Well, we gotta test it typologically...

CCG Cont.

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

The Project

- The **goal** of my project is to capture German VPCs in CCG.
- Can CCG **satisfactorily account** for German VPCs?

Roadmap

- First, we look at English VPCs in CCG.
- Constable and Curran's (2009) VPC proposal for English
- Hockenmaier's (2006) German CCG
- My proposal for dealing with German VPCs in CCG
- Talk about inconsistencies and problems with coordination
- Propose an alternative
- Conclusion and work to do

The Project

- The **goal** of my project is to capture German VPCs in CCG.
- Can CCG **satisfactorily account** for German VPCs?

Roadmap

- First, we look at English VPCs in CCG.
- Constable and Curran's (2009) VPC proposal for English
- Hockenmaier's (2006) German CCG
- My proposal for dealing with German VPCs in CCG
- Talk about inconsistencies and problems with coordination
- Propose an alternative
- Conclusion and work to do

VPCs in CCG

- CCGBank - the main corpus for CCG-related work
- Despite their prevalence in English, CCGbank currently mishandles them (Constable and Curran 2009)
- CCGbank treats particles as **adverbial modifiers**.

Three Problems

1. Particles are core, adverbs are optional

   (8) a. The police **tracked down** the thief.
   b. ?? The police **tracked** the thief.

2. Iterability

   (9) a. Anna **looked up** the book.
   b. * Anna looked **out down up** the book.

3. Word order (with respect to other adjuncts)

   (10) a. ?? The police tracked **yesterday down** the thief.
   b. ?* Anna looked **then up** the book.

VPCs in CCG

- CCGBank - the main corpus for CCG-related work
- Despite their prevalence in English, CCGbank currently mishandles them (Constable and Curran 2009)
- CCGbank treats particles as **adverbial modifiers**.

Three Problems

1. Particles are core, adverbs are optional

   (8) a. The police **tracked down** the thief.
   b. ?? The police **tracked** the thief.

2. Iterability

   (9) a. Anna **looked up** the book.
   b. * Anna looked **out down up** the book.

3. Word order (with respect to other adjuncts)

   (10) a. ?? The police tracked **yesterday down** the thief.
   b. ?* Anna looked **then up** the book.

Constable and Curran (2009)

- None of the current categories (N, NP, S, PP) can be used without opening the door to undesired transformations.
- Create a new category for particles **RP**

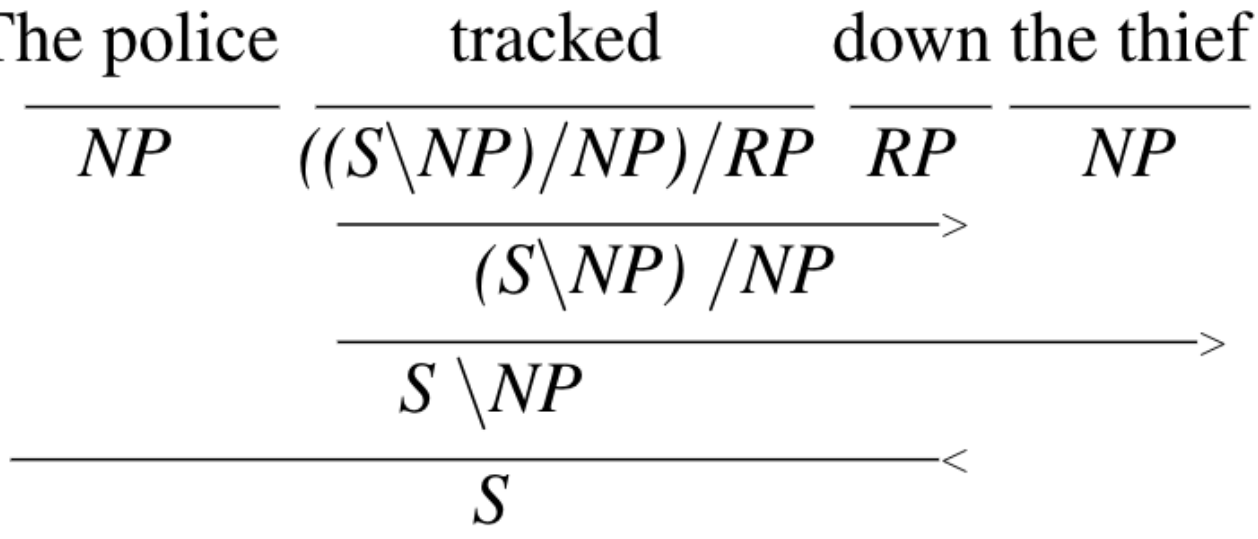

Figure 1: C&C (2009) continuous order derivation

Constable and Curran (2009)

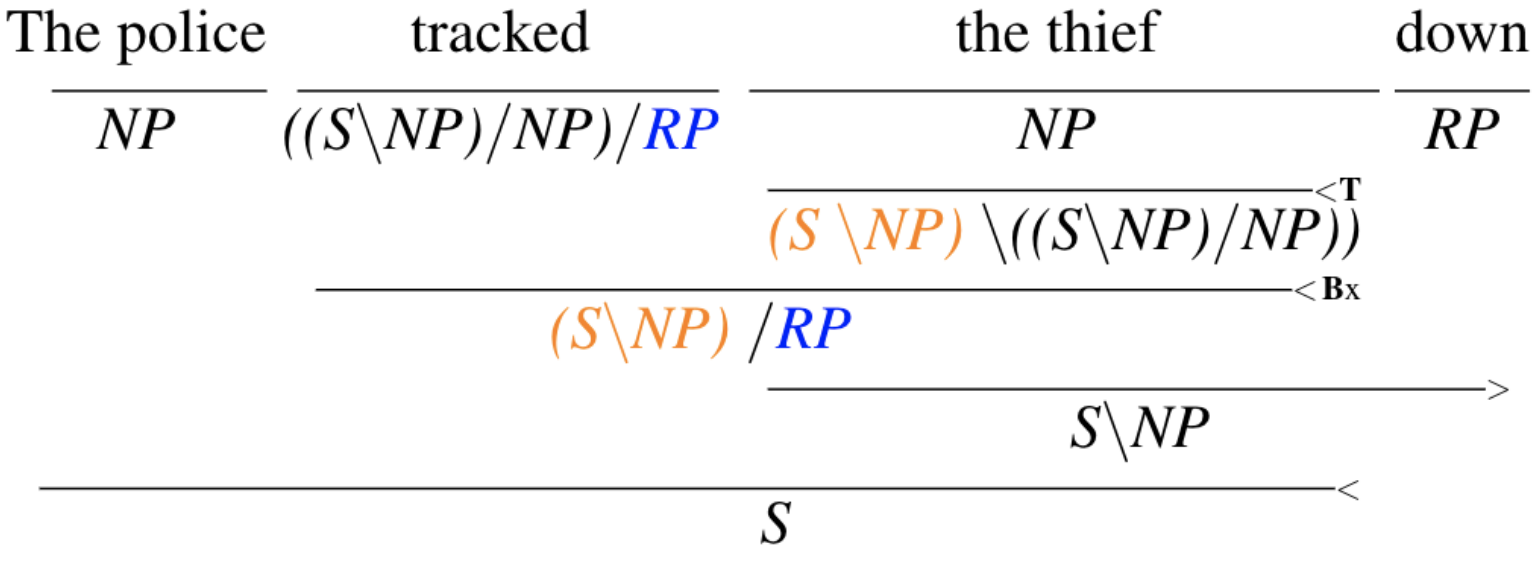

Figure 2: Constable and Curran(2009) discontinuous order derivation

German CCG

- What about German?
- Hockenmaier (2006) created a CCGbank for German using the German Tiger Corpus. unclear whether it includes VPCs, but it should; they are very prevalent in German.
- Accounts for word order and morphology

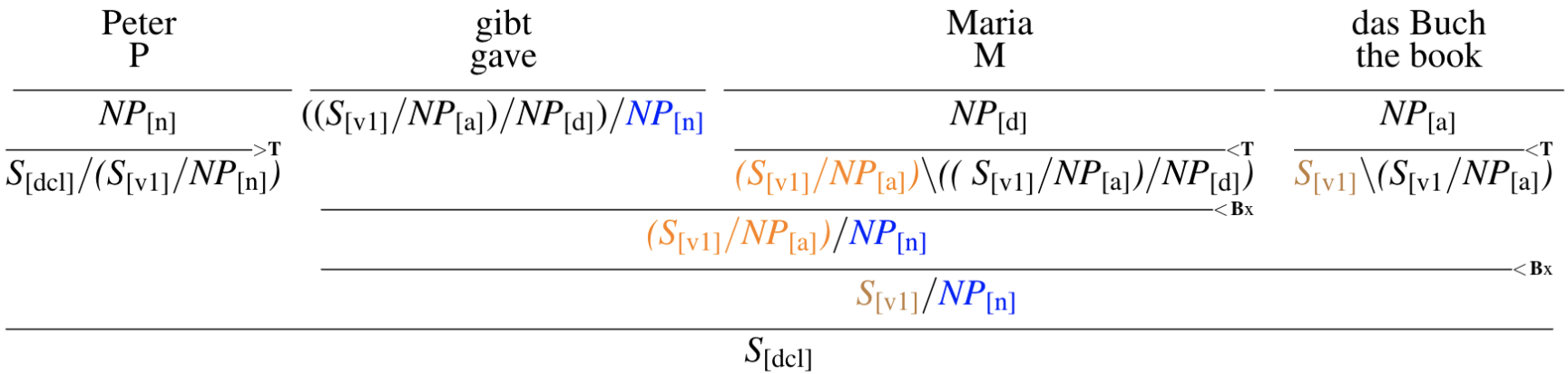

Figure 3: "Peter gave Maria the book" (Hockenmaier 2006)

German CCG cont.

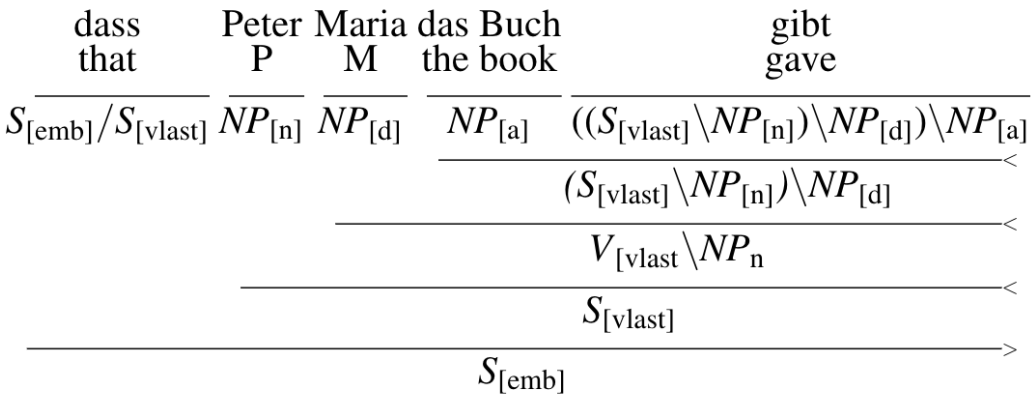

Figure 4: subordinate German VPC sentence (Hockenmaier 2006)

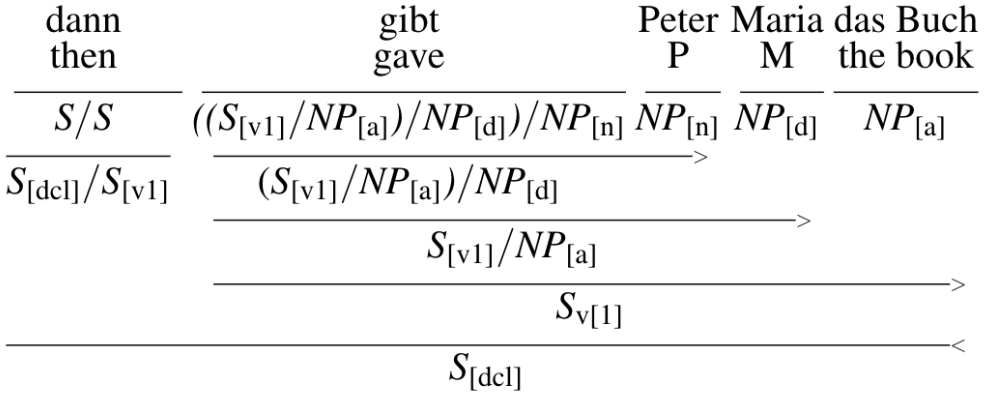

Figure 5: German VPC constuction, fronted adjunct (Hockenmaier 2006)

Interim Summary

- VPCs are **misnhandled** in CCGbank.
- Constable and Curran (2009) created a new label (**RP**) to deal with them.
- Hockenmaier (2006) translated a German Corpus to a German CCGbank adding features and capturing word order.

Where we are headed

- Capturing German VPCs in CCG
- Talk about problems with the approach
- Directions for further work

German VPCs in CCG

- So how do we derive **German VPCs** in CCG?
  I start out with the **simplest** sentence I could think of...

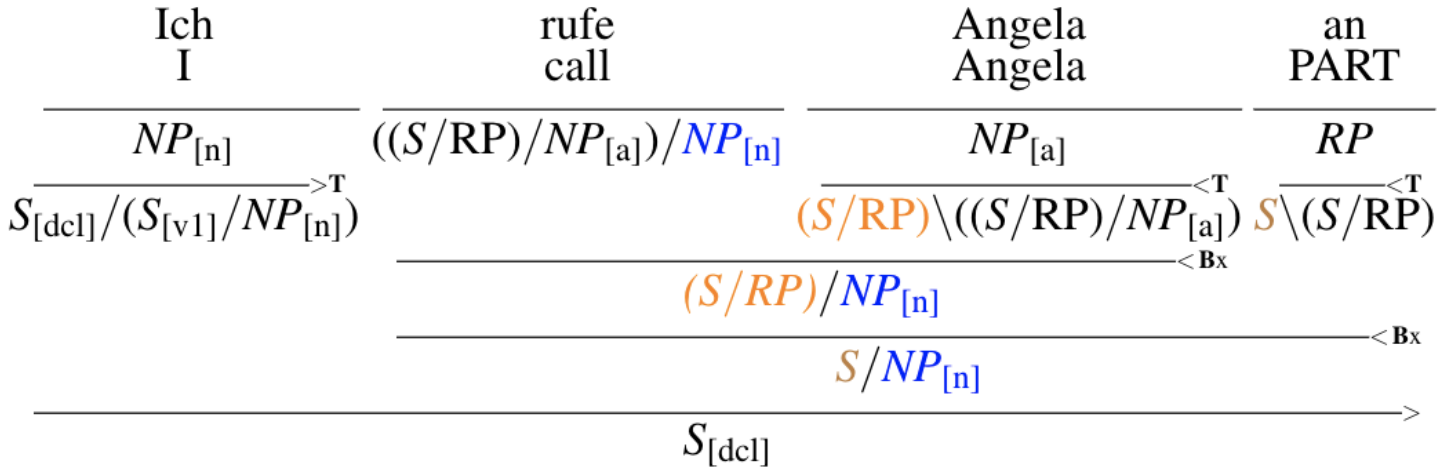

Figure 6: "I am calling Angela (up)" à la Hockenmaier (2006)

- It **works**!

More German VPCs in CCG

- How about adjuncts?

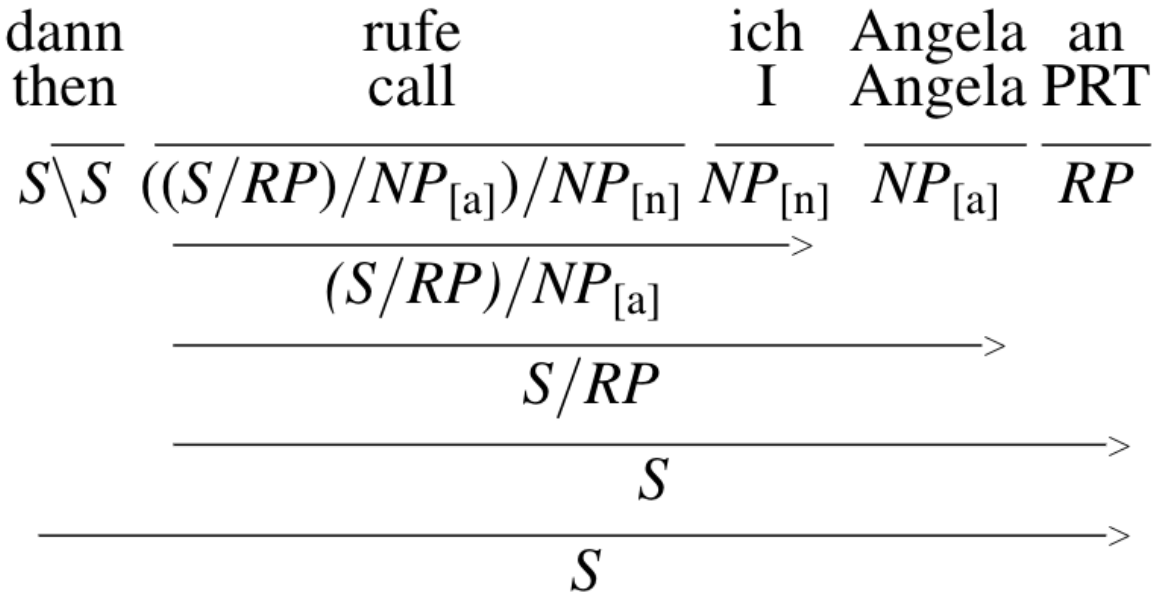

Figure 7: 'Then I call Angela (up)' à la Hockenmaier (2006)

- It still **works**!

  and we just had to use functional application!

Even More German VPCs in CCG

- But things get complicated with embedded clauses. . .

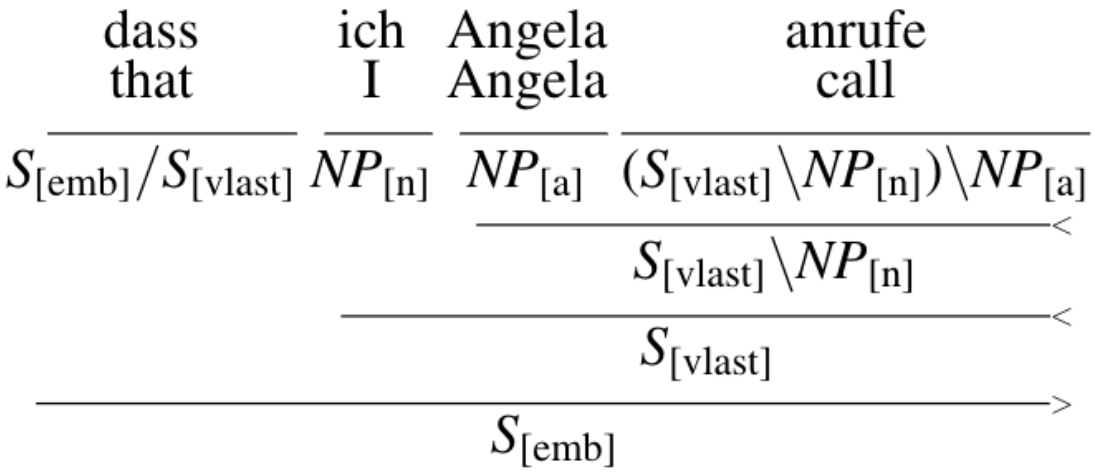

Figure 8: "that I am calling Angela (up)"

- It still **works**!
  Kind of. . .
- The lexical entry for the verb is not **consistent**.
  Is *anrufen* 1 or 2 words?

More German VPCs in CCG

- Can we be consistent with the lexical entry for *anrufe*?

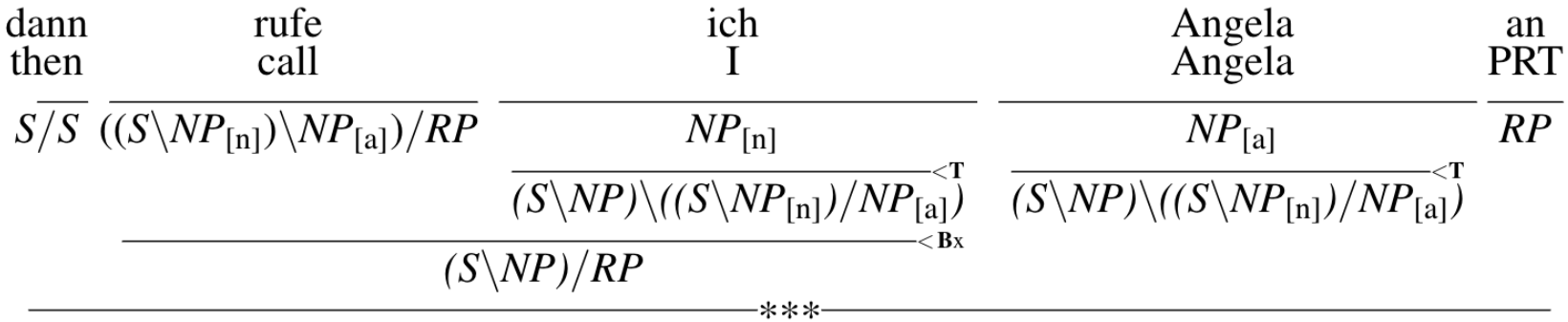

Figure 9: 'then I call Angela (up)'

- The derivation for the fronted adjunct clause doesn't work anymore.
  Can we make it work with one same type for the verb?

Conjunction

- VPC coordination presents another **challenge** for CCG.
- I look at **three** types of coordination
  - **particle sharing** - two different verb stems are associated to the same particle
  - **verb sharing conjunction** - verb stem associated with two different particles
  - **dual valency conjunction** - the verb acts as both intransitive and transitive across conjuncts

Particle Sharing

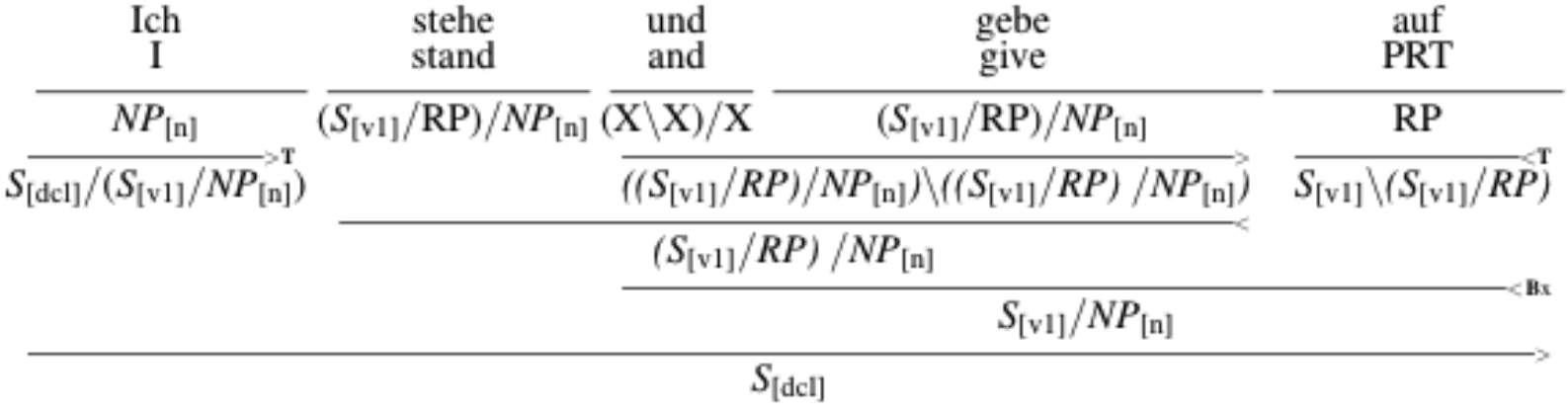

Figure 10: CCG derivation for particle sharing conjunction

- Verbs combine with one another first, then with the particle, and last with the subject.
- One elegant feature of this solution is that the verb remains the same as that for regular main clauses.

Verb Sharing

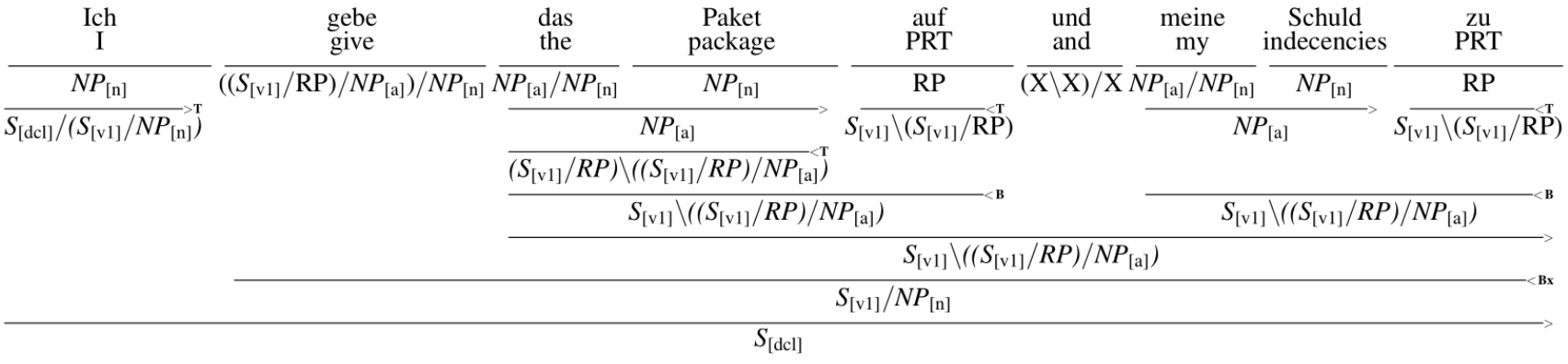

Figure 11: CCG derivation for verb sharing conjunction

- First, combine the two object conjuncts, and then compose them with the verb.
- Last, by functional application we derive the full sentence
- Once again, we keep the verb category *consistent* with the main clause verb.

Dual Valency

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

Figure 12: Failed CCG derivation for dual valency conjunction

- This is another **problematic** point for VPCs in CCG
- The verb needs to have two different categories to be able to derive this sentence - it must be **transitive** *and* **intransitive** at the same time

Take home message

- CCG bank **mishandles** English VPCs
- Constable and Curran (2009) propose a solution with a new category **RP**
- Hockenmaier (2006) creates a German CCG that handles word order correctly
- I propose a mix of both approaches as a **first step** to handle VPCs in German
- CCG **cannot** handle German VPCs in a cohesive and consistent manner!

Take home message cont.: Embedded clauses

- We need to allow VPCs to **separate** in **main clauses**,
  but to **stay together** in **embedded** and **relative clauses**.
  without allowing anything undesired in between.

Constraints

- We want to allow past tense formation.
  an**ge**rufen
  'called'
- Disallow other adjuncts between particle and verb.
  *an**gestern**rufen
  PRT.yesterday.call

- it doesn't seem we can do that (yet!)

Take home message cont.: Conjunction

- Particle sharing and verb sharing are **easily** handled
- Dual valency conjunction. . . not so much

Constraints

- Verbs like *aufgeben*, *zugeben*, and *geben* need to be able to coordinate with their respective complements. . . or lack thereof as part of a conjunction

(11) *Ich* ***gebe*** *auf* *und* *meine Schuld* *zu*
I give PRT and my indecencies PRT
'I admit my indecencies and give up my candidacy.'

(12) *Ich* ***gebe*** *dieses Rennen auf* *und* *dir* *den ersten Platz.*
I give this race PRT and you the first place
'I give up this race and (give) you the first place.'

Coordination in Minimalism

- The minimalist program has given **considerable** attention to VPCs
- Their **status** (phrasal vs. lexical) remains **unclear**
- However, Minimalism handles VPCs (along with their movement and coordination) more readily than CCG

VPCs in Minimalism

- I follow Wurmbrand (2000) in deriving VPCs in Minimalism
- For cases of particle sharing the two VPCs coordinate and the particle then moves to a higher position, and the verb moves to cannonical V2

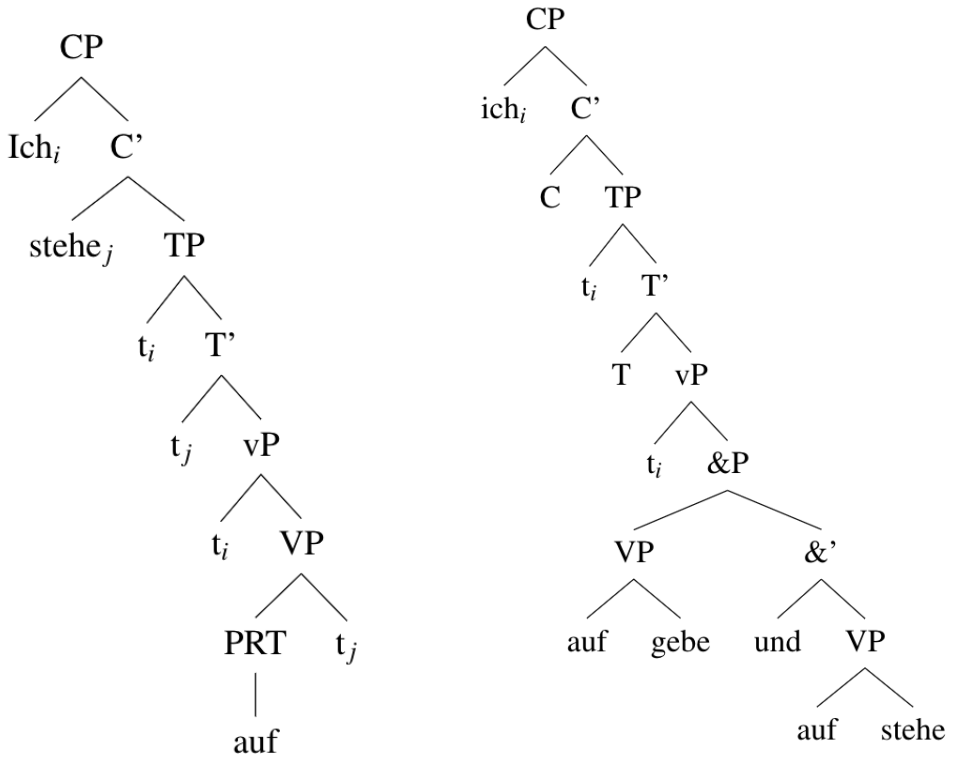

Figure 13: Minimalist derivation for *Ich stehe auf* 'I stand up' (left) and verb stem coordination (right)

Dual Valency in Minimalism

- Minimalism has no issue in dealing with dual valency coordination

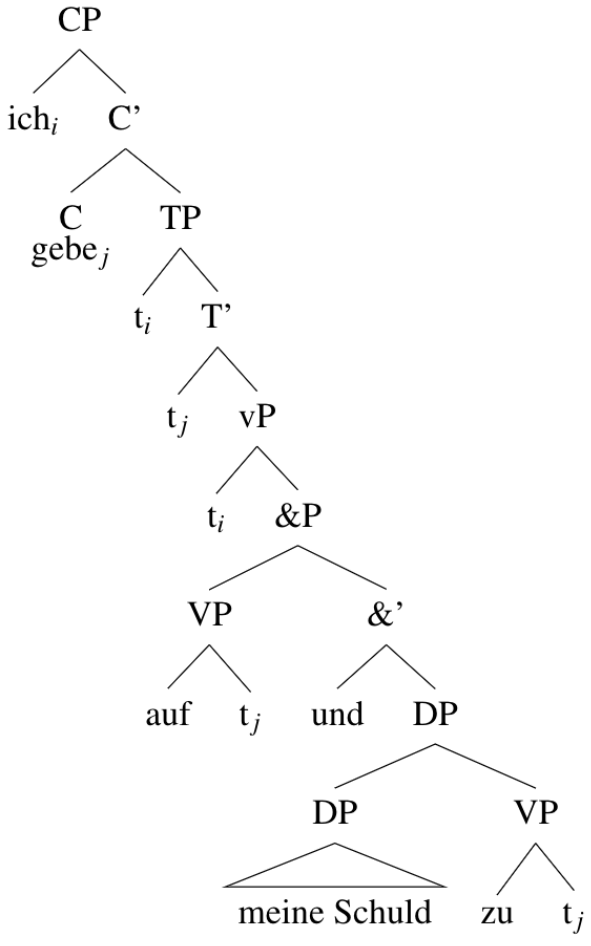

Figure 14: Dual valency coordination in minimalism

Conclusion and future work

- VPCs are **mishandled** in CCG
- Using Constable and Curran (2009) and Hockenmaier (2006) I porpose a **first step** to model German VPCs in CCG
- German VPCs present a greater **challenge** for CCG (esp. embedded clauses and coordination)
- Minimalism seems **better equipped** to deal with German VPCs

Future work

- Is there a way to model German VPCs in CCG? (how do we do it for other Germanic languages?)
- Is the minimalist approach adequate for German VPCs? Does it overgenerate?

Conclusion and future work

- VPCs are **mishandled** in CCG
- Using Constable and Curran (2009) and Hockenmaier (2006) I porpose a **first step** to model German VPCs in CCG
- German VPCs present a greater **challenge** for CCG (esp. embedded clauses and coordination)
- Minimalism seems **better equipped** to deal with German VPCs

Future work

- Is there a way to model German VPCs in CCG? (how do we do it for other Germanic languages?)
- Is the minimalist approach adequate for German VPCs? Does it overgenerate?

References

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

Dehé, N. (2002). *Particle Verbs in English: Syntax, information structure and intonation*. Linguistik Aktuell/Linguistics Today. John Benjamins Publishing Company.

Harbert, W. (2006). *The Germanic Languages*. Cambridge Language Surveys. Cambridge University Press.

Hockenmaier, J. (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.

Steedman, M. (2017). *Combinatory Categorial Grammar*. Unpublished manuscript.

Wurmbrand, S. (2000). The structure (s) of particle verbs. *Ms*

Appendix I: Verb inconsistencies

- The lexical entry for ‘anrufen‘ is **inconsistent** in **main and adjunct clauses vs. embedded ones**

$$\frac{\text{rufe}\\ \text{call}}{((S/RP)/NP_{[a]})/NP_{[n]}} \qquad \frac{\text{anrufe}\\ \text{call}}{(S_{[vlast]}\backslash NP_{[n]})\backslash NP_{[a]}}$$

Figure 15: CCG Types for *rufe. . . an* and *anrufe*