


Dolores Batinić, Thomas Schmidt

RECONSTRUCTION OF SEPARABLE PARTICLE VERBS IN A CORPUS OF SPOKEN GERMAN

Mitglied der


Leibniz
Leibniz-Gemeinschaft

SEPARABLE VERBS

- German particle verbs:
 - may occur as one word or as a multiword expression:
 - Das **sieht** gut **aus**
 - ..., dass es gut **aussieht**
 - Challenge in CL:
 - current lemmatizers do not reconstruct the lemmas of separated verbs
 - Lemmatisation in case of separated verb

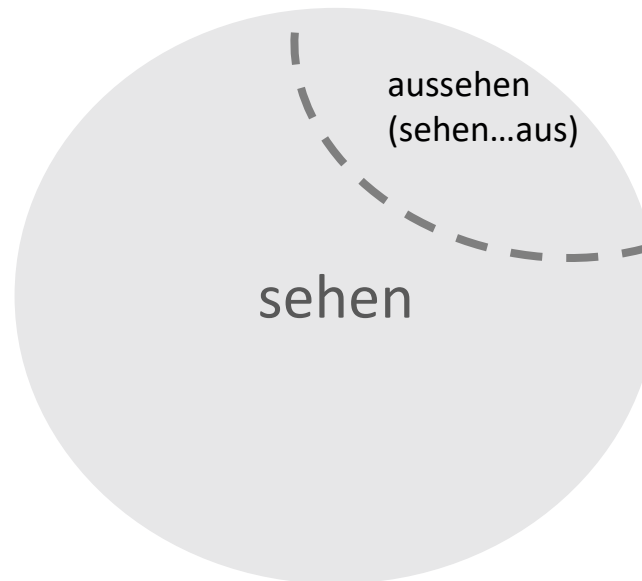
<i>segment</i>	das	sieht	gut	aus
<i>lemmas</i>	das	sehen	gut	aus

- Lemmatisation in case of a separable verb written together

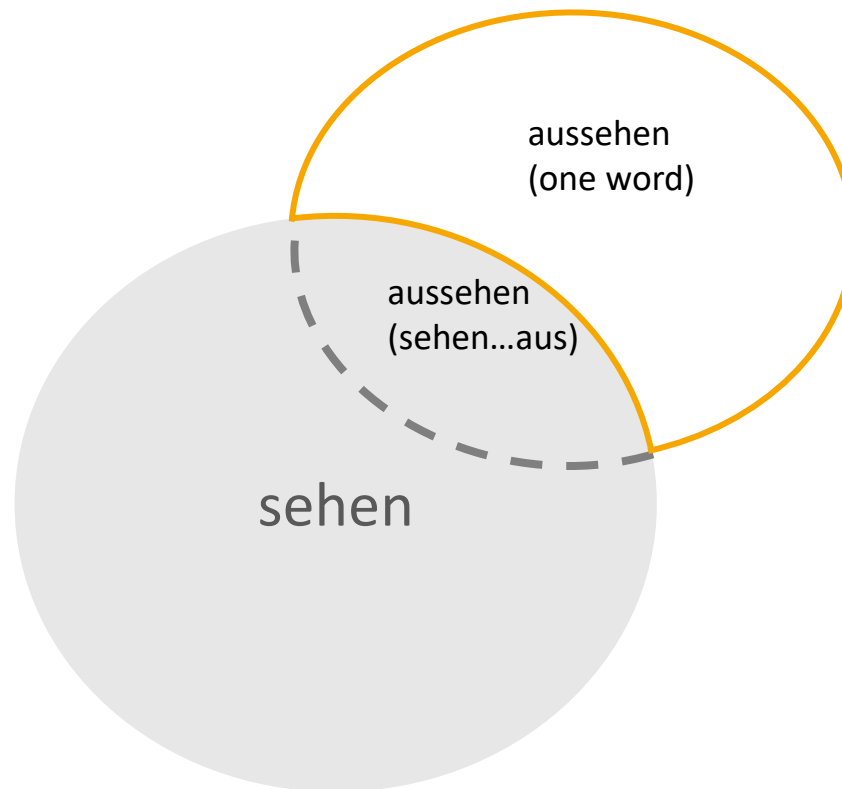
<i>segment</i>	dass	es	gut	aussieht
<i>lemmas</i>	dass	es	gut	aussehen



CURRENT LEMMATISATION





CURRENT LEMMATISATION



CORPUS EXAMPLES

Corpus examples for query *aussehen* in www.dwds.de

aussehen|  

Korpus: Referenz- und Zeitungskorpora (▾)

Start: 1473 **Ende:** 2016

Textklassen: Belletristik Wissenschaft Gebrauchsliteratur Zeitung

Anzeige: KWIC voll maximal

Sortierung: Datum absteigend ▾


Anzahl Treffer pro Seite: 50 ▾

1-50 von 63778 Treffern (65703 insgesamt) [Treffer exportieren](#)

1:	z	2016	donald_trump	...seits ließ sie das naiv und für höhere Ämter ungeeignet	aussehen	.
2:	z	2016	theater_euro	...eigen mit insgesamt acht Stücken nun, wie diese Vielfalt	ausseht	.
3:	z	2016	donald_trump	...d offengelassen, wie seine künftige Regierungsmannschaft	aussehen	wird – obwohl er sich in den vergangenen Tagen mit zahlr...
4:	z	2016	donald_trump		ausseht	, verriet er aber nicht.
5:	z	2016	tennis_kerber	... mir dennoch wichtig, ich selbst zu bleiben und natürlich	auszusehen	.
6:	z	2016	unfaelle_fue		ausgesehen	, als ob das Fahrzeug mit Wucht gegen einen Baum gepral...
7:	z	2016	elbvertiefung	Wie die beiderseitigen Wirtschaftsbeziehungen in Zukunft	aussehen	könnten, diskutieren 500 Teilnehmer der bedeutendsten c...
8:	z	2016	parteien_gab	Einerseits, weil das so	aussehen	würde, als ob die Sozialdemokraten hinter der Kanzlerin h...
9:	z	2016	fussball_auch	...rwart Felix Wiedwald, der bei der Aktion etwas unglücklich	aussah	.
10:	z	2016	fussball_auch	...rwart Felix Wiedwald, der bei der Aktion etwas unglücklich	aussah	.
11:	z	2016	obama_trump	Man müsse erst abwarten, wie Trumps Politik	aussehe	.
12:	z	2016	facebook_zuc	...y hatte aber kürzlich skizziert, wie eine technische Lösung	aussehen	könnte:
13:	z	2016	elbvertiefung	...n Zahl von Emoticons arbeiten, um zu beschreiben, wie es	ausseht	, wenn eine SMS von Udo eingeht.
14:	z	2016	fussball_leipz	...etern Nationaltourtour Leno, der beim Ausgleich nicht gut	aussah	.
15:	z	2016	fussball_ribry	...er ich kann jetzt noch nicht sagen, wie die Lösung genau	ausseht	."
16:	z	2016	international	"So wie es jetzt	ausseht	, hat es noch nicht genug Fortschritte gegeben", sagte Me...
17:	z	2016	auto_abgas_e	Zwar hatte es vor einem Jahr schon einmal so	ausgesehen	, als könnte sich die NOx-Affäre zur CO2-Affäre ausweiten.
18:	z	2016	auto_systeme	Zwar hatte es vor einem Jahr schon mal so	ausgesehen	, als könnte sich die NOx-Affäre zur CO2-Affäre ausweiten.
19:	z	2016	barack_obam	...rt nicht viel darüber, wie die nächsten konkreten Schritte	aussehen	könnten.
20:	z	2016	fluechtlingskr	...manchen konservativen Politiker jedoch die ideale Kirche	auszusehen	.
21:	z	2016	fussball_stoe	...lt voraus, auch wenn es in der Tabelle momentan nicht so	ausseht	", erklärte Stöger vor dem Duell des Elften gegen den Sec...
22:	z	2016	e_bike_hawk	... ich finde, dass man auf den meisten E-Bikes wie ein Depp	ausseht	– was an den Rädern liegt, nicht am Fahrer.
23:	z	2016	leute_eddie_r	Weil der Koffer ähnlich	ausseht	wie der des Filmhelden, ist der Schauspieler nun in der Zw...
24:	z	2016	bundespraesi	Und, was nach Coup hätte	aussehen	können, blockierte die CSU:
25:	z	2016	wissenschaft	Wie die Folgen eines Einschlags	aussehen	könnten, ließ 2013 allein schon die Explosion eines etwa 2...
26:	z	2016	donald_trump	Wie politische Auseinandersetzung mit Trump	aussehen	kann, zeigt Elizabeth Warren, demokratische Senatorin in ...
27:	z	2016	wahlen_tause	...t, wer Trump außenpolitisch berate und wie dessen Team	aussehe	.

CORPUS EXAMPLES

Corpus examples for query *aussehen* in www.dgd.ids-mannheim.de



KORPUSAUSWAHL

[Kein virtuelles Korpus geladen](#)
[Virtuelles Korpus laden](#)

- BW Berliner Wendekorpus
- FOLK Forschungs- u. Lehrkorpus für gesprochenes Deutsch
- PF Deutsche Umgangssprachen: Pfeffer-Korpus
- ZW Zwirner-Korpus
- AD Australiendeutsch
- DR Deutsche Mundarten: DDR
- DS Dialogstrukturen
- FR Grundstrukturen: Freiburger Korpus
- HL Deutsche Hochlautung
- IS Emigrantendeutsch in Israel

POSITION
TOKEN
KONTEXT
METADATEN
ANZEIGE

Transkribiert: z.B. "kannscht"

Lemmat:

Normalisiert: z.B. "kannst"

POS: z.B. "VMFIN"

Reguläre Ausdrücke

Recherche - Tokens

KWIC wird angezeigt. 00:00:01.0

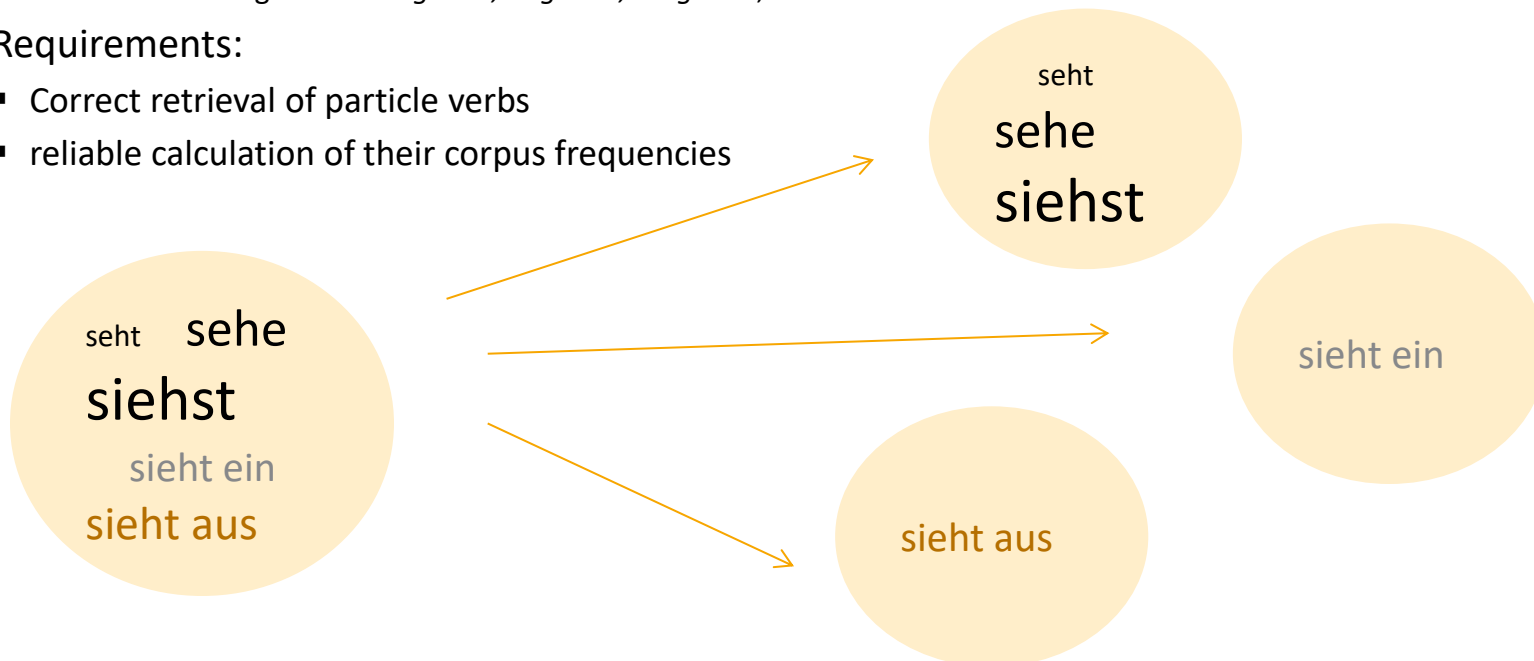
Ergebnisse 1 bis 20 von 243 (243 / 0 aus-/abgewählt) Seite 1 von 13

	Ereignis	Sprecher	Kontext	Treffer
<input checked="" type="checkbox"/>	FOLK_00001	LB	dann verifizieren ob des am skop genauso	aus sieht
<input checked="" type="checkbox"/>	FOLK_00001	LB	wie sieht die zünds wie kann die	aus sehen also entweder isse ganz abgefackelt oder
<input checked="" type="checkbox"/>	FOLK_00004	GS	kollege gescheüber weil er ja vielleicht besser	aus seh will
<input checked="" type="checkbox"/>	FOLK_00005	LB	prüft ja hier üvoltmeter so wie s	aus sieht
<input checked="" type="checkbox"/>	FOLK_00006	LB	mal anschauen wie s mi m zweite	aus sieht widerstand der induktivgeberspule
<input checked="" type="checkbox"/>	FOLK_00006	LB	uff jeden fall irgendwann nachdem isch komisch	aus sehe hab hab isch des ding dann weggeworfe des ging gott
<input checked="" type="checkbox"/>	FOLK_00007	GS	des soll könnte	aus sehen wie folgendes
<input checked="" type="checkbox"/>	FOLK_00007	JK	finden wenn er weiß überhaupt wie er	aus sieht
<input checked="" type="checkbox"/>	FOLK_00007	GS	dann muss ma wisse wie er	aus sieht richtig das heißt er braucht kognitive kennnisse
<input checked="" type="checkbox"/>	FOLK_00007	GS	mal er muss ja wisse wo er	aus sieht wo er suche muss ja
<input checked="" type="checkbox"/>	FOLK_00008	LB	wie würd s signal	aus sehen
<input checked="" type="checkbox"/>	FOLK_00008	LB	des blesch nur e kle bissl anersch	aus sieht oder e bissl aneri vorspannung hat verändertet sich sofort das
<input checked="" type="checkbox"/>	FOLK_00009	LB	net ob s im dacia jetzt annersch	aus sieht
<input checked="" type="checkbox"/>	FOLK_00009	LB	warum des bild jetzt anders	aus sieht
<input checked="" type="checkbox"/>	FOLK_00009	MK	hm wenn s bild normal	aus sieht
<input checked="" type="checkbox"/>	FOLK_00009	LB	isch die aussage wenn s bild normal	aus sieht
<input checked="" type="checkbox"/>	FOLK_00009	LB	isch defekt weil es darf nisch so	aus sehen
<input checked="" type="checkbox"/>	FOLK_00011	VK	ich weiß nich wie horst	aus sieht
<input checked="" type="checkbox"/>	FOLK_00013	CJ	unheimlich oben auf seinem schiff richtig unheimlich	aus sah
<input checked="" type="checkbox"/>	FOLK_00015	FR	ist weil ich weiß äh wie die	auszusehen hat

Ergebnisse 1 bis 20 von 243 (243 / 0 aus-/abgewählt) Seite 1 von 13

MOTIVATION

- LeGeDe project (Lexik des gesprochenen Deutsch), IDS Mannheim
 - Building a lexicographic resource for spoken German
 - source: FOLK corpus, 1.95 M tokens, tagger accuracy: 95%
 - current study on verbs which happen to be highly productive in terms of verb particle combinations:
 - Perception verbs: sehen -> *aussehen, ansehen, vorsehen*, etc.
 - Motion verbs: gehen -> *ausgehen, angehen, vorgehen*, etc.
- Requirements:
 - Correct retrieval of particle verbs
 - reliable calculation of their corpus frequencies



RECONSTRUCTING SEPARABLE VERBS

- Principle
 - attach the verb particle to the lemma of the nearest preceding finite verb in the same sentence
 - (Volk et al. 2016):

*das sieht gut **aus***

id	w1	w2	w3	w4
lemma	das	sehen	gut	aus
reconstruction	das	aussehen	gut	[ref. to w2]
STTS tag	PDS	VVFIN	ADJD	PTKVZ

- Also reconstructing:
 - Modal and auxiliary verbs + PTKVZ > finite verbs (es. vorhaben)
 - multiple tags for one verb (mach das an und dann aus)

MATERIAL

- FOLK
 - Research and Teaching corpus of Spoken German (Schmidt 2016)
 - Spontaneous interactions
 - Accessed via DGD (Schmidt 2014): structured corpus search with concordances combined with video and audio files
 - 4 annotation levels (transcription, lemmatisation, normalisation, pos-tags)
 - Metainformation about speakers and recording available
 - **no sentence boundaries:**
 - contributions : corpus segmentation based on pauses > 0.2s

CONTRIBUTIONS

- In many cases separated verb occurs in one and the same contribution:

FOLK_E_00021

0163 PL voll der streber

0164 CH [**guck**] dir hier [mal den profi **an**]

CONTRIBUTIONS

- ...but it can also cross contribution boundaries:

FOLK_E_00013

0761 CJ nun

07 (0.25)

0763 CJ **sah** er

0764 (0.21)

0765 CJ schon viel freundlicher **aus** (.) klar

CONTRIBUTIONS

- ...or it can be interrupted by another speaker:

FOLK_E_00006

1149 LB **guckt** eusch mal

1150 XM (is richtig)

1151 (0.41)

1152 LB die form des signals **an**

VERB RECONSTRUCTION

- Testing the reconstruction on three levels
 - Contribution as boundary
 - the contribution boundaries are limits within which the reconstruction is performed
 - Turn as boundary
 - the reconstruction is performed on the sequence of contributions belonging to one speaker
 - No boundaries
 - the reconstruction can skip preceding contributions of another speaker

VERB RECONSTRUCTION

- List of separable verbs from www.verblists.de
 - Extended with spoken German variants:
 - darauf -> drauf
 - daran -> dran
 - heran -> ran
 - herein -> rein, etc.
 - Total added: 630 types to the list of 7658 separable verbs

RESULTS

- Evaluation on gold standard (all tags correct)
 - FOLK GOLD Standard (Westphal & Schmidt 2016), 99247 tokens
 - Precision of 99% for all approaches
 - verb particle crossed the contribution boundary in 1 example out of 100
 - 11% of all finite verbs detected as separated separable verbs (597 out of 5240)
 - setting max. distance of 23 words between reconstruction candidates

FOLK_E_00180

0827	AAC2	äh (.) achtnhalb jahre im verein gespielt (.) und jetzt spiele ich nur (.) ähm aus spaß mit meinen freunden aus derstufe °h noch ei (.) aus_m
-------------	-------------	---

0828	(0.3)
-------------	-------

0829	AAC2	nach_er schule °hh in so ner mittwochsliga mit (.) in so ner (.) indoorhalle
-------------	-------------	---

RESULTS

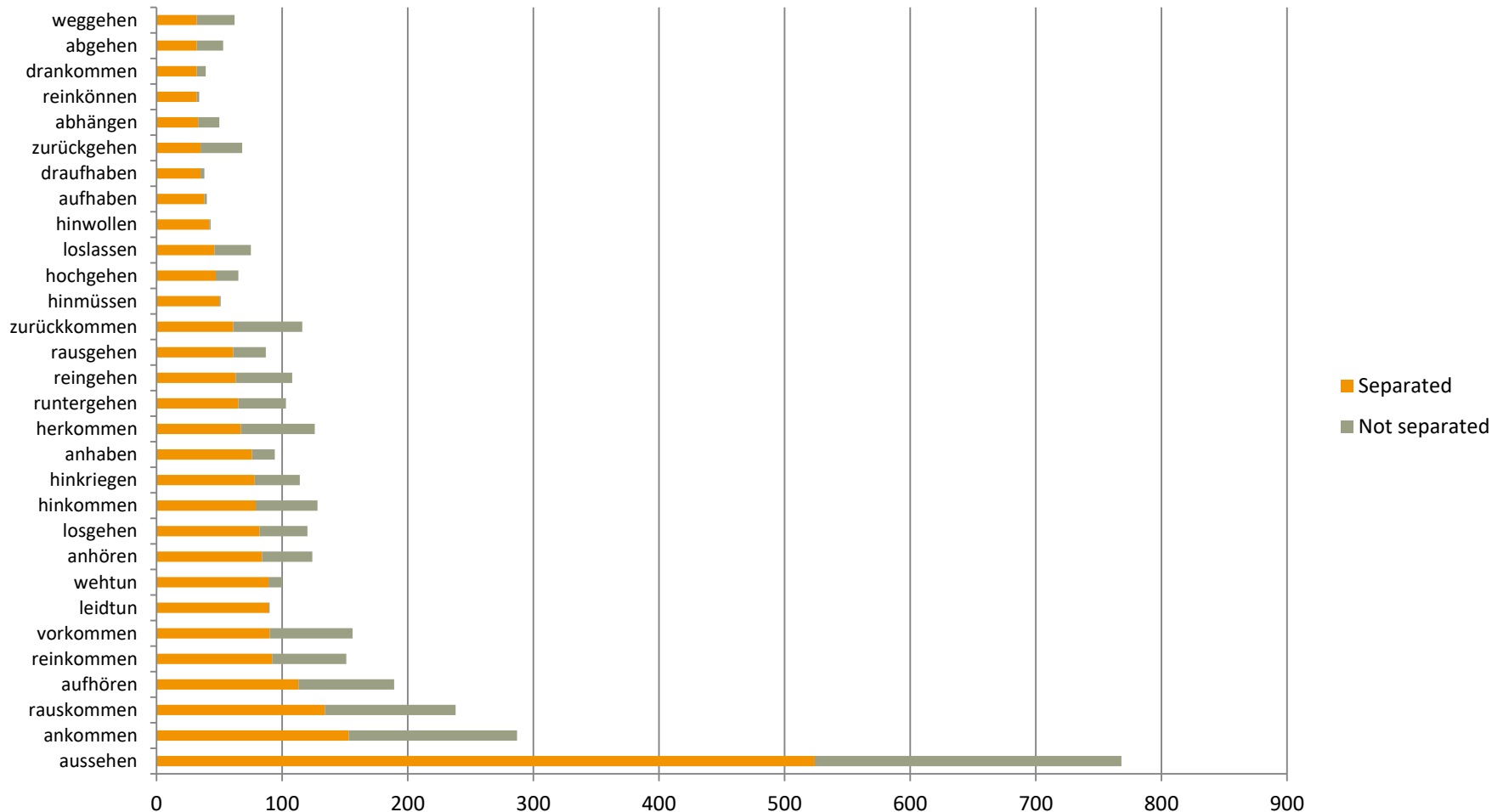
▪ Evaluation on FOLK

- 7% of all finite verbs were reconstructed
 - 1059 different verbs (types) with the contribution-oriented approach
 - 1140 with the turn-oriented approach
 - 1156 with the skipping approach
- Precision slightly higher when reconstructing within contributions: 0.91 vs. 0.88
- Recall slightly higher when crossing contribution boundaries: 0.98 vs. 0.95
- All three approaches are acceptable (accuracy: 0.9)

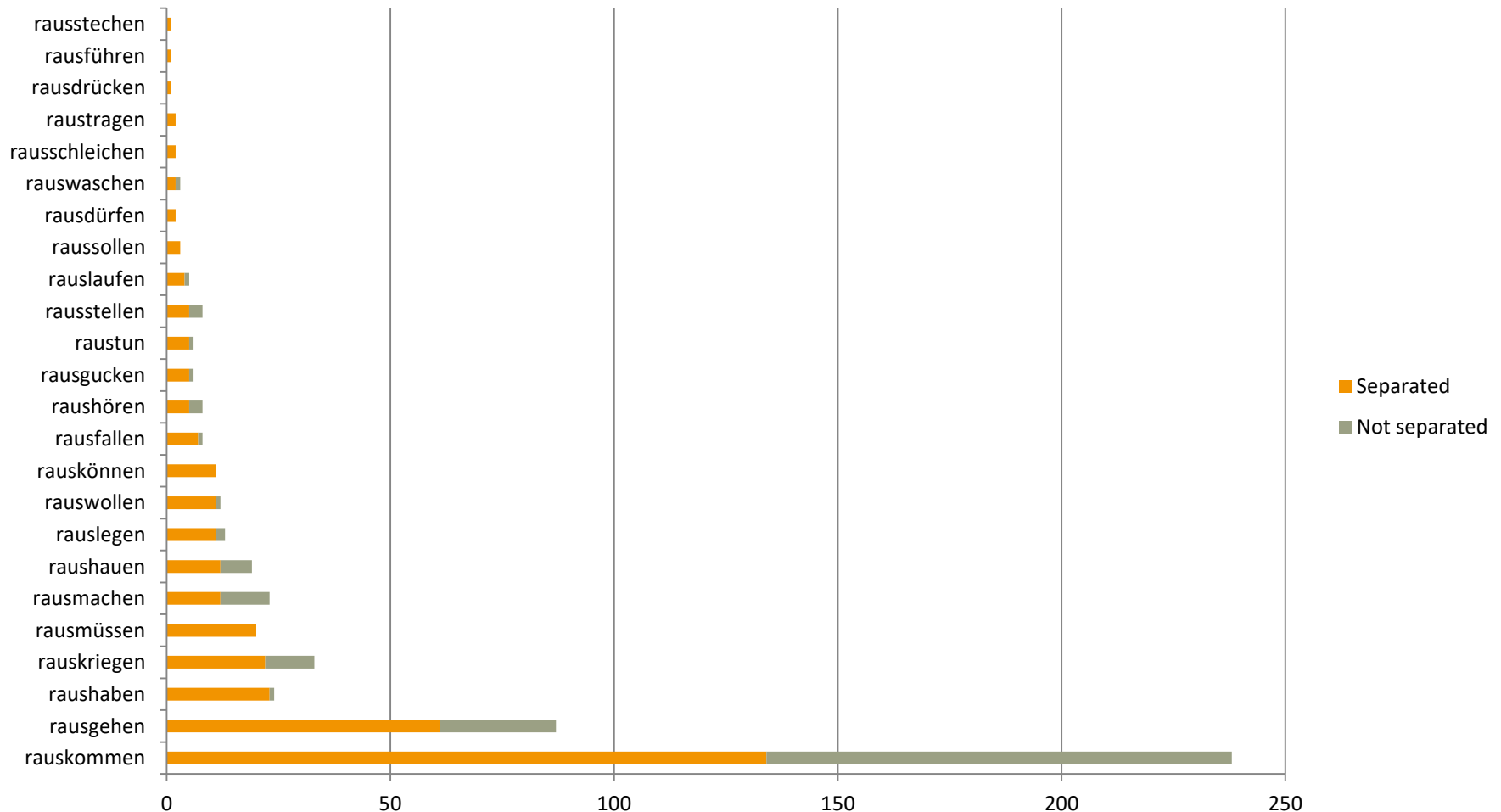
- For higher precision:
 - Set lower distance between base verb and the particle (average distance: 3.9 words)
 - Postprocessing in case of modal and auxiliary verbs + PTKVZ -> finite verbs (es. vorhaben)
 - Postprocessing in case of multiple particles for one verb (machen Sie einmal mit der Faust auf und zu)

- Challenges:
 - Nested clauses
 - Particles preceding the verb stem (mit nach Thailand nehmen)
 - Modal & auxiliary verbs

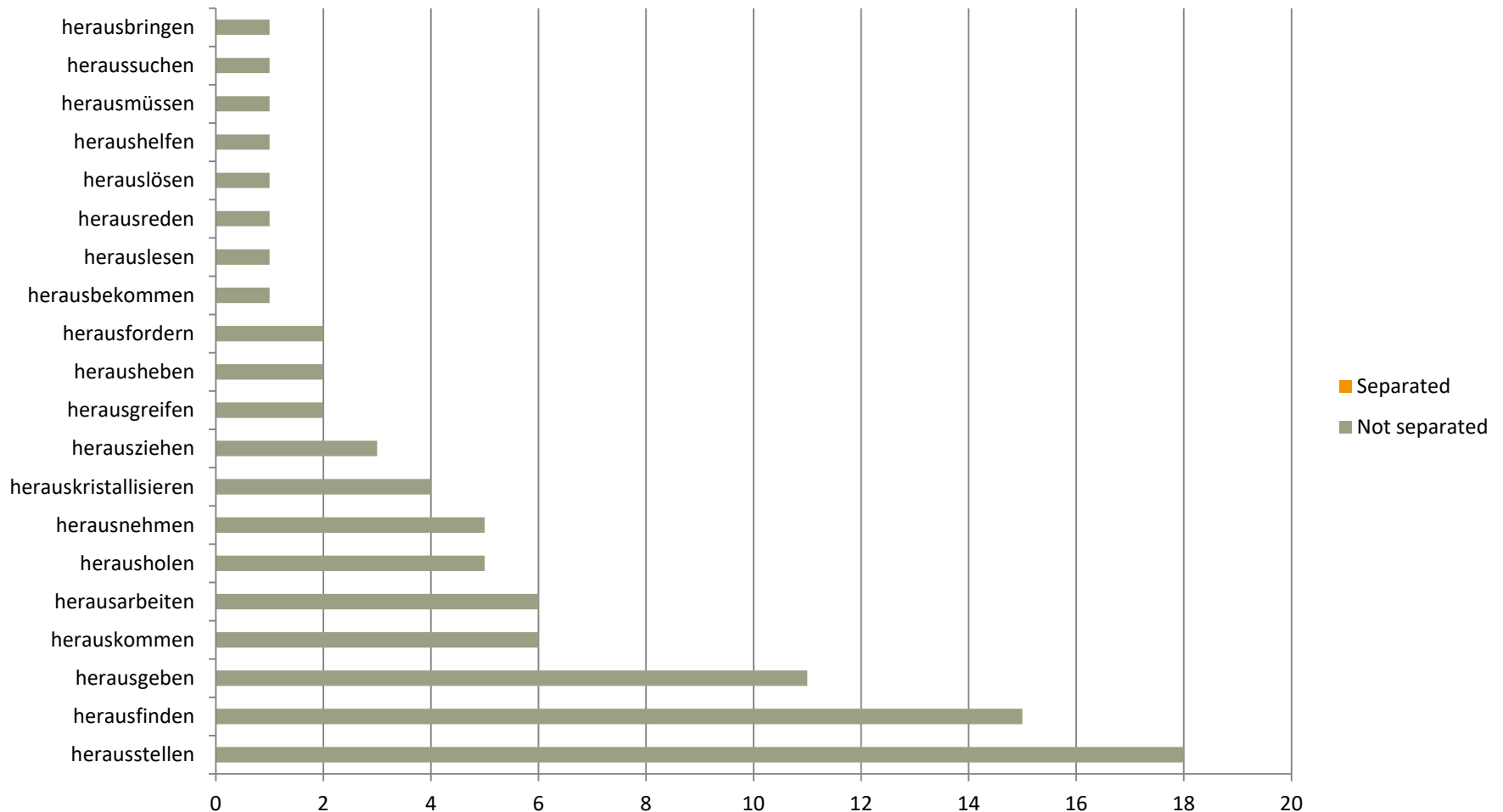
VERBS OCCURRING MORE OFTEN SEPARATED THAN TOGETHER (TOP 30)



COMBINATIONS WITH „RAUS“



COMBINATIONS WITH „HERAUS“



AMBIGUOUS CASES

- Repetitions of the same particle

FOLK_E_00106

0568	BUC1	und (.) °hh jetz geh mal von der linken (.) oberen ecke
0569		(1.13)
0570	BUC2	ja
0571		(0.49)
0572	BUC1	äh (.) so einen zentimeter raus praktisch so schräg raus
0573		(0.43)
0574	BUC1	in richtung [(.)] in die die ecke zeigt

AMBIGUOUS CASES

- Truncations

FOLK_E_00046

0463 **VW** [so die frau (.) **lebt** sozusagen]

0464 **VW** [oder beide **leben** ihre emotionale] seite halt **aus** die sie im alltag und in ihrer famil[ie halt nicht ausleben könnendes is so cool geschrieben und davon gibt_s halt] n zweiten teil und den [hat halt des buch hab ich ausgeliehen von meiner sozusagen schwiegermutter in spe]

AMBIGUOUS CASES

- Self-corrections

FOLK_E_00076

0218	DJ	°h währenddessen bricht der vulkan weiter auf
0219		(0.59)
0220	DJ	aus

AMBIGUOUS CASES

- Coordinated particles

FOLK_E_00055

0051	US	[...] der läuft auf der straße so hin und her also der [is wirklich] sturzbetrunken ich
------	----	---

OUTLOOK

- Extending the additional annotation layer with similar types of reconstructions
 - pronominal adverbs (Kaiser & Schmidt 2016)
 - ex. dafür vs. da...für

FOLK_E_00261

0520 OB (.) ich hab kohle **dafür** gekricht

FOLK_E_00044

0675 CT ja auf ihre (.) also das **da** zahlen wir nix **für**

THANK YOU