On the Evolutionary Profile of *again* and Complex predicates: The Decomposition Corpus

Formal Diachronic Semantic—FoDS9

28 November, 2024, Bologna

Remus Gergel, Martin Kopf
Saarland University

{ remus.gergel | martin.kopf } @uni-saarland.de

Decomposing Decomposition in Time

Funded by
DFG Deutsche Forschungsgemeinschaft
German Research Foundation

Outline

1. Introducing the Decomposition Corpus (DCC) as part of a larger ongoing enterprise
    - Introducing *again* and its main readings
    - Properties of the DCC
2. Evolutionary profile of *again*
    - *Again*, complex predicates and visibility
3. Conclusion and Outlook

1. Background

Larger enterprise: A corpus with designated annotations for multiple semantic phenomena: the **Structured Diachronic Semantics corpus (SDS)**

- i.e., **semantic annotations** on top of **syntactically parsed corpora**

York-Toronto-Helsinki Corpus of Old English Prose – **YCOE** (Taylor et al. 2003) and

Penn Parsed Corpora of Historical English – **PPCHE** (Kroch et al. 2000, 2004, 2016)

Currently under discussion: The **decomposition layer** of the SDS

1. The Saarland DeComposition Corpus—DCC

Objective:

- Stand-off **annotation layer** for **decompositional items**

  e.g., *again, eft* (OE ‘again’), *re*-verbs, *almost*

1. The Saarland DeComposition Corpus—DCC

Motivation:

- Explore how the **structural** and **meaning** components **interface diachronically**.
- Moreover, a **historical perspective can elucidate synchronic debates**; e.g., negotiate between competing theoretical (e.g. structuralist vs. lexicalist w.r.t. decompositional item *again*) accounts (Gergel & Beck, 2015; Degano & Aloni, 2021).
- **Reveal readings** that have not been systematically analyzed so far; e.g. redirectional/'send-it-some-where-else' uses (Gergel, 2023).
- Advance understanding of presuppositions and **presuppositional change**. (Gergel, forthc.)

1. The Saarland DeComposition Corpus—DCC

- **Semantic annotations on top-of syntactically annotated corpus data**——YCOE & PPCHE

  (Taylor et al. (2003); Kroch et al. (2000, 2004, 2016))

- Decompositional presupposition triggers–iterative items:
  - 4,378 uses of adverb *again* (mostly in PPCHE)
  - 3,389 uses of adverb *eft* (OE ‘again’; mostly in YCOE)
  - 2,236 *re*-verbs (in PPCHE; mostly where *re-* was productive during PPCHE)
- Exhaustive annotation for the above data, including:
  - Reading (repetitive, counterdirectional/restitutive, …)
  - Target and antecedent predicate (main verb)

} more details below

1. *again* in the DCC – Overview data in PPCHE

4,378 uses of *again*:

• PPCME2	1150–1500 CE	(Kroch et al. 2000)	Middle English (ME)	945 *again*s	0.079%
• PPCEME	1500–1720 CE	(Kroch et al. 2004)	Early Modern English (EME)	1,532 *again*s	0.088%
• PPCMBE2	1700–1910 CE	(Kroch et al. 2016)	Late Modern English (LME)	1,901 *again*s	0.068%

1. *again* in the DCC – Readings

Repetitive *again* (‘rep’)

(1) [A]ll the plants then must be **examined** , […] and those which are planted in pots, should in the following year’s bloom be **again examined** (FALLOWFIELD-1791-2,34.349, *‘Gardening Calendar’*)

- **repetitive** reading presupposes an event of the same kind — ‘**rep**’

1. *again* in the DCC – Readings

Counterdirectional/Restitutive *again* (‘cd/res’)

(2) [Mr. Knightley] sat really lost in thought for the first few minutes; […] He hesitated, **got up** . […] and he **sat down again**; (AUSTEN-1815-2,169.633, *‘Emma’*)

- **restitutive/counterdirectional** reading presupposes an event in the opposite direction; the result state of the again-event is the restoration of a previous state — ‘**cd/res**’

1. *again* in the DCC – Readings

Counterdirectional/Restitutive *again* (‘cd/res’)

(2) [Mr. Knightley] sat really lost in thought for the first few minutes; […] He hesitated, **got up** . […] and he **sat down again**; (AUSTEN-1815-2,169.633, *‘Emma’*)

- **restitutive/counterdirectional** reading presupposes an event in the opposite direction; the result state of the again-event is the restoration of a previous state — ‘**cd/res**’

(on a structural analysis, *again* shares semantics with repetitive uses)

1. *again* in the DCC – Readings

Counterdirectional *again* (‘cd’) — (counterdirectional proper)

(3) queen Gwynyvere _had_ hym in greet favoure […] and so he **loved** the queen agayne aboven all other ladyes (CMMALORY-M4,180.2394)

- ‘**counterdirectional (proper)’** reading presupposes an event in the opposite direction; no feasible result state available — ‘***ctd***’

Note: Some uses of again have a discourse organizing function (‘dm’); other smaller readings are subsumed as ‘other’.

1. *again* in the DCC – Readings

Counterdirectional *again* (‘cd’) — (counterdirectional proper)

(3) queen Gwynyvere _had_ hym in greet favoure […] and so he **loved** the queen agayne aboven all other ladyes (CMMALORY-M4,180.2394)

- ‘**counterdirectional (proper)’** reading presupposes an event in the opposite direction; no feasible result state available — ‘***ctd***’

Note: Some uses of again have a discourse organizing function (‘dm’); other smaller readings are subsumed as ‘other’.

➡ **Disambiguation** depends on presupposition (PSP) satisfaction in the **context**.

1. *again* in the DCC – Context and disambiguation

(4) In two hours, I called Joseph to ***carry* him up again**; (Bronte-1848-2,2,287.311, Wuthering Heights)

(referents: *I*–Heathcliff, *him*–Linton)

Two options: Did Joseph carry Linton up before?

Or: Did Joseph carry Linton down before? —Or: Did Linton move himself?

1. *again* in the DCC – Context and disambiguation

I _**brought**_ **him down** one evening, the day before yesterday, (306)

and just set him in a chair, (307)

and never touched him afterwards. (308)

I sent Hareton out, (309)

and we had the room to ourselves. (310)

(4) In two hours, I called Joseph to ***carry*** **him up again**; (Bronte-1848-2,2,287.311, Wuthering Heights)

(referents: *I*–Heathcliff, *him*–Linton)

Antecedent _*brought*_ *him down* in (.306) satisfies the cd/res presupposition.

1. *again* in the DCC – Annotations and labels

(1) [A]ll the plants then must be _**examined**_, […] and those which are planted in pots, should in the following year's bloom be **again *examined*** (FALLOWFIELD-1791-2,34.349, '*Gardening Calendar*')

- **repetitive** reading presupposes an event of the same kind — tag as **'rep'**

i. classify uses of again	'rep'(etitive)
ii. mark the main verb of the again-predicate	*examined*
iii. mark the main verb of the antecedent	_examined_

→ 2 independent annotations + 3rd (sometimes 4th) annotator/team review (for 4,378 *again*s)

1. *again* in the DCC – Annotations and labels

(1) [A]ll the plants then must be **_examined_**, […] and those which are planted in pots, should in the following year's bloom be **again** ***examined*** (FALLOWFIELD-1791-2,34.349, '*Gardening Calendar*')

- **repetitive** reading presupposes an event of the same kind — tag as **rep**

i. classify uses of again — 'rep'(etitive)

ii. mark the main verb of the again-predicate — *examined*

iii. mark the main verb of the antecedent — _examined_

→ 2 independent annotations + 3rd (sometimes 4th) annotator/team review (for 4,378 *again*s)

1. *again* in the DCC – Annotations and labels

- No verb present → substitute categories

  e.g. ***wet***$_{\textbf{ADJ}}$ (b.) & ***rain***$_{\textbf{N}}$ (d.):

(5) a. August 28th. Monday. Avallon. (273—275)

b. […] Quite grey and _**wet**_ again **to-day**, but not so cold. (282)

c. August 29th. Tuesday. Avallon. (283—285)

d. ***Rain*** **again**, and all black above. (RUSKIN-1882-2,3,1019.286)

1. Overt satisfaction vs. inferential satisfaction

- So far: Overt satisfaction (e.g. *examined—examined again*; *get up—sit down again*)
- However: Frequently a PSP has to be inferred from the context

(6) Linton has not been feeling well lately and has spent the days up in his room.
But today he _**had**_ **dinner** in the **downstairs** dining room. (INF → cd/res)
In two hours, I called Joseph to ***carry*** **him up again**;

(adapted from Bronte-1848-2,2,287.311, Wuthering Heights)

- DCC labels: 'INF' for inferential satisfaction of PSPs, 'GIV'—overt satisfaction, 'NOA'—no antecedent (e.g. inference via world/historical knowledge)

1. Overt satisfaction > inferential satisfaction

- Competition of **multiple putative antecedents** and target predicate(/event):

Overt satisfaction overrules inferential satisfaction.

(7) Linton has not been feeling well lately and has spent the days up in his room.

Last night, John had to _**carry**_ **Linton up to his room**. (GIV → rep)

Today he //**had dinner** in the **downstairs**// dining room. (INF → cd/res)

In two hours, I called Joseph to ***carry*** **him up again**;

(adapted from Bronte-1848-2,2,287.311, Wuthering Heights)

1. Proximity antecedent – target-predicate

- If there is **equal status** of putative antecedents in terms of **GIV/INF satisfaction**:

Relative proximity of antecedent to target predicate/event **disambiguates**

(Proximity overrules distance)

(8) Linton has not been feeling well lately and has spent the days up in his room.

Last night, John had //**carried Linton up to his room**//. (GIV → rep)

Today, Linton _**came**_ **downstairs** to have dinner with us in the dining room. (GIV → cd/res)

In two hours, I called Joseph to ***carry* him up again**;

(adapted from Bronte-1848-2,2,287.311, Wuthering Heights)

1. Twist for proximity

- **Discursive proximity is not always enough** to reliably disambiguate.
- Here, at first glance: //*uniting*// makes for a putative antecedent for the repetitive reading.
- A closer look reveals _parting_ as the correct antecedent:

(9) The filaments of a cortical vessel are to be looked on […] as so many little bundles placed near together, and at first growing parallel to each other; but soon quitting this direction, **the filaments of one fascicle _parting_ from that to which they originally belonged**, and inclining more or less obliquely towards another, **sometimes //uniting// with it**, at others bending backwards, and *__uniting__* **again with that from which it proceeded** […]

(ADAMS-1787-2,663.157)

1. Expert annotation – Final Product — A sample

(1) [A]ll the plants then must be **_examined_**, […] and those which are planted in pots, should in the following year's bloom be **again** ***examined*** (FALLOWFIELD-1791-2,34.349, '*Gardening Calendar*')

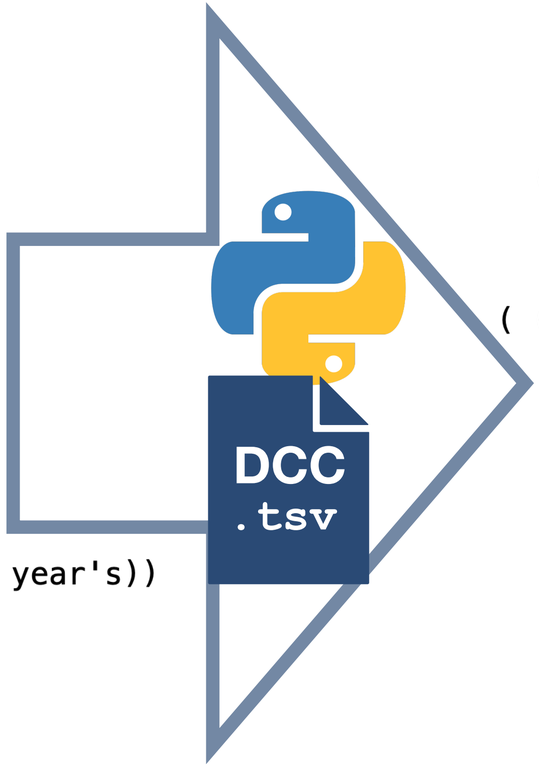
- **repetitive** reading presupposes an event of the same kind — tag as **rep**

i. classify uses of again	'rep'(etitive)	`@DC511:rep:GIV:TRG`
ii. mark the main verb of the again-predicate	*examined*	`@DC511:rep:GIV:PRD`
iii. mark the main verb of the antecedent	_examined_	`@DC511:rep:GIV:ANT`

→ 'DCC'-tags are aligned with the PPCHE as well as YCOE

1. Sample from fallowfield-1791-2.psd:

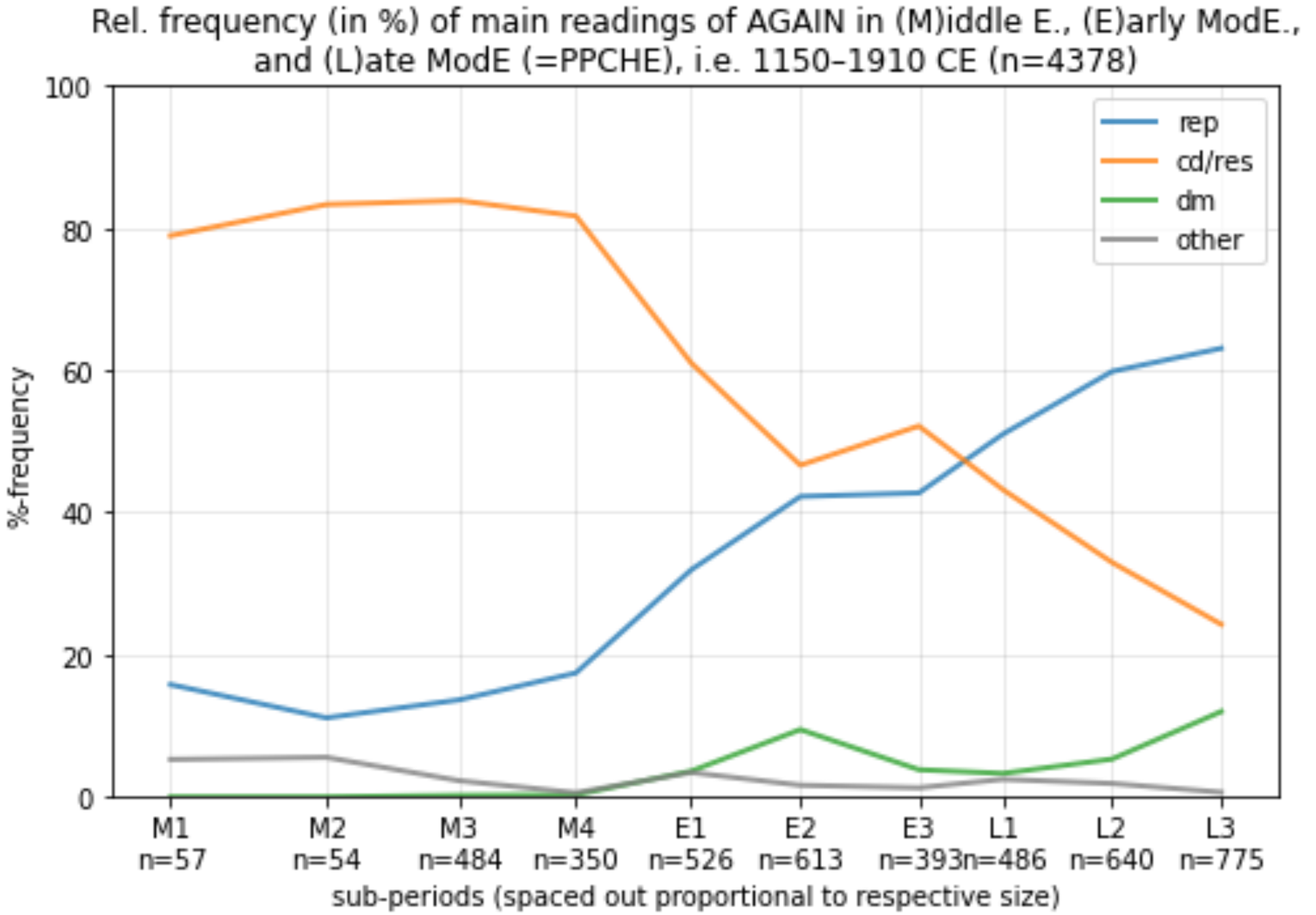
```
( (IP-MAT (NP-SBJ (Q all) (D the) (NS plants))
      (ADVP-TMP (ADV then))
      (MD must)
      (BE be)
      (VAN examined)
      (. .))
  (ID FALLOWFIELD-1791-2,33.345))

                          [...]

( (IP-MAT (CONJ and)
      (IP-MAT-1 (NP-SBJ (D those)

                          [...]

            (MD should)
            (PP (P in)
            (NP (NP-POS (D the) (VAG following) (N$ year's))
                (N bloom)))
            (BE be)
            (ADVP (ADV again))
            (VAN examined))

                          [...]

      (. .))
  (ID FALLOWFIELD-1791-2,34.349))
```




```
( (IP-MAT (NP-SBJ (Q all) (D the) (NS plants))
      (ADVP-TMP (ADV then))
      (MD must)
      (BE be)
      (VAN@DC511:rep:GIV:ANT:6 examined)
      (. .))
  (ID FALLOWFIELD-1791-2,33.345))

                          [...]

( (IP-MAT (CONJ and)
      (IP-MAT-1 (NP-SBJ (D those)

                          [...]

            (MD should)
            (PP (P in)
            (NP (NP-POS (D the) (VAG following) (N$ year's))
                (N bloom)))
            (BE be)
            (ADVP (ADV@DC511:rep:GIV:TRG:14 again))
            (VAN@DC511:rep:GIV:PRD:15 examined))

                          [...]

      (. .))
  (ID FALLOWFIELD-1791-2,34.349))
```

2. Development of main readings of *again* in PPCHE

- Main take-away: Decrease of *cd/res* readings clarifies and certifies previous accounts (Beck et al., 2009; Gergel and Beck, 2015).



2. Complex predicates, Visibility & Decompositionality

- **Diachrony** of morphosyntax of **complex predicates** (cps) is under-investigated.
- Fundamental **typological change can occur**, e.g. Persian has been developing from a simplex-verb language to a complex-verb language; complex verbs are dominant in PD Persian. (Karimi 2013, 2023, a.o.)
- **English** has **not generalized** such a **development** for verbs/verbal clusters

  (*pace* grammaticalization paths of *do* and auxiliaries; cf Garrett 1998 for confounding potential of infinitives/nominal during reduction of English inflections)
- Following Snyder (2001)—plus insight that acquisitionally relevant data show in diachronic developments (Cournane 2017, Kidney 2019)—cps include **goal PPs, particles, resultatives, double objects**, and possibly others.

2. Complex predicates, Visibility & Decompositionality

- Cps here—based on structural features in PPCHE:

  (no V+Adj resultatives as e.g. *hammer flat, wipe clean*, etc. (Snyder, 2001; Beck, 2005))

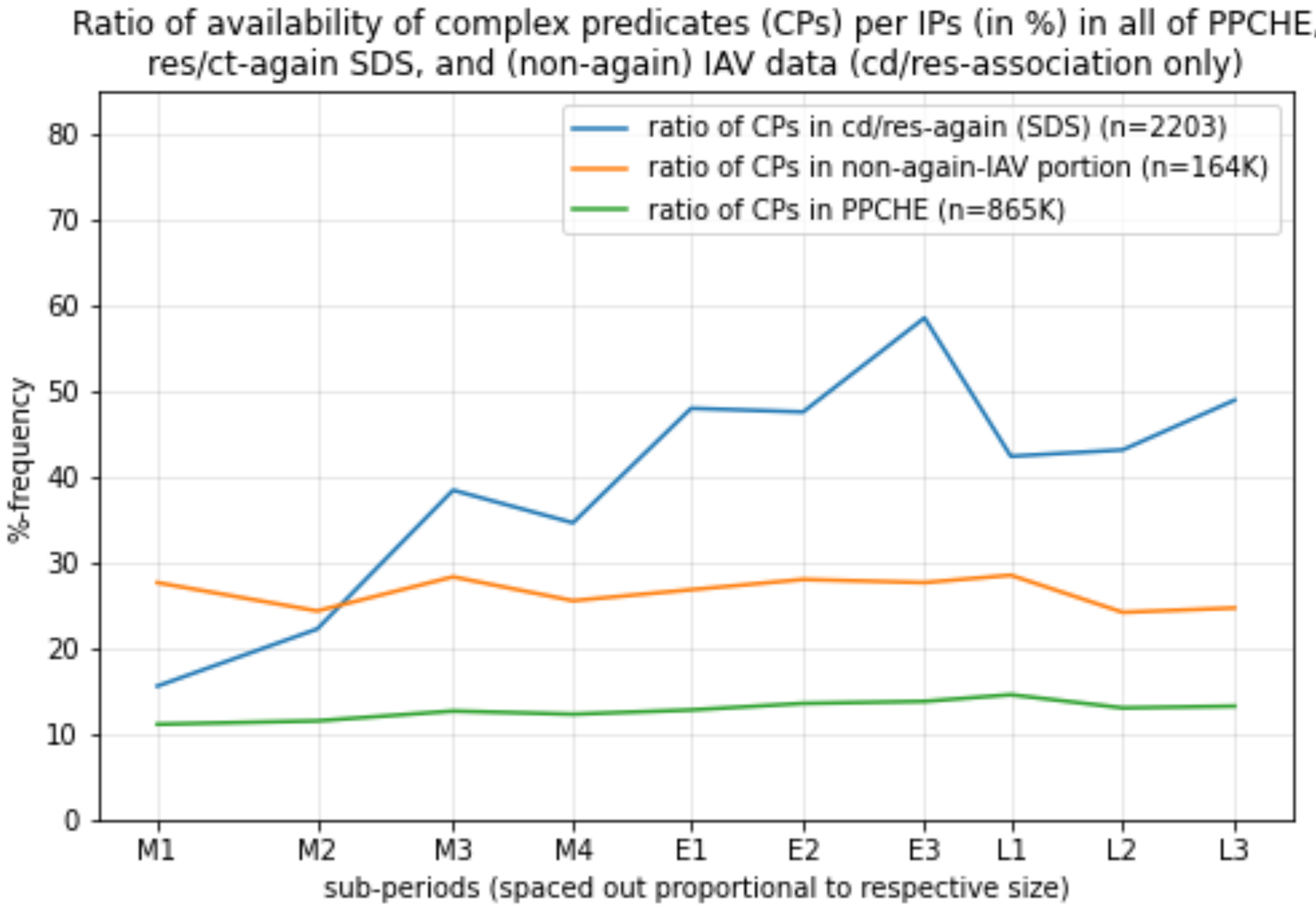
  - **Adverbial particles 'RP'** — *carry up, pull out, close in* …
  - **Directional phrases 'XP-DIR'** — *send* (NP-DIR *home*), *carry* (ADV-DIR *forth*), …
  - **Double-object structures '2OB'** — *give* (NP-OB2 *John*) (NP-OB1 *the book*)
  - **goal PPs** — PPs headed by P *to*/+*to*

  For all, result states are represented overtly; setting ii on visibility/accessibility parameter

  (Rapp & Stechow, 1999; Beck, 2005; Beck & Gergel 2015)

2. Cps in cd/res DCC vs. PPC and IAV-PPC

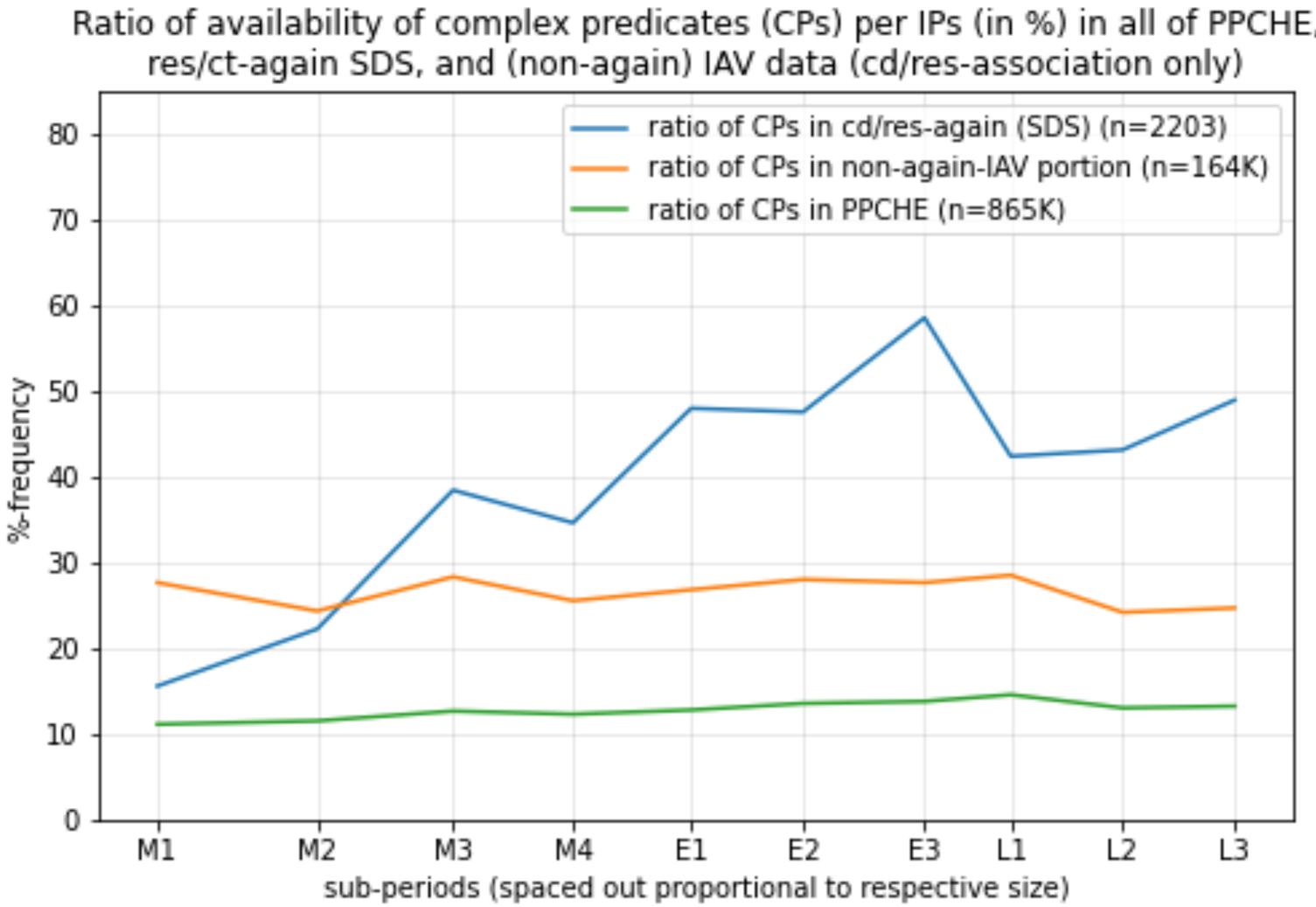
- cps increase with cd/res uses;
- cps remain stable in PPCHE (green) and ‘iterative-associating verbs’ (non-*again* cd/res IAVs; orange).



2. Cps in cd/res DCC vs. PPC and IAV-PPC

- cps increase with cd/res uses;
- cps remain stable in PPCHE (green) and ‘iterative-associating verbs’ (non-*again* cd/res IAVs; orange).
- Key finding: **Ratio of cps does not seem to be tied to a particular reanalysis point**;

($again_{cd}$ –> $again_{res}$ –> $again_{rep}$)



2. Conclusions: Overall modeling of entire development

- **Rather than discrete shifts** in cp-ratios—corresponding to respective processes of reanalysis (cd→res and res→rep)—**continuous development**
- Proposed modeling: **Grammar competition** (Kroch 1989) between a decompositional and non-decompositional version of *again*;
- Both available from the very beginning of $again_{adv}$

2. Conclusions: Overall modeling of entire development

- Driving forces:
  - A **single entry** is prima facie **most economical**; decompositional/Stechowian *again*; with setting iii of visibility/accessibility parameter; (‘*Jon picked up a rock and* ***dropped it again***’)
  - However, (proper) **decompositionality flies in the face of markedness**/visibility; Beck (2005) re. visibility:
    - **Result state modification** is **more likely in cps than in untransparent, lexical predicates**
- As a consequence, **cd/res-*again* requires** incrementally **more cps**; i.e. phonetically overt res.-state phrases; i.e. more setting-ii predicates, fewer setting-iii predicates (‘covert res.-state’)

2. Conclusions: Overall modeling of entire development

- Finally: Visibility strikes back!
- DCC data shows development **parallel to increase of cps** in cd/res uses:
    - **incidence of cd/res-*again* is decreasing**—next to increasing repetitive uses.
- Visibility strikes back: As cps increase—among overall decreasing cd/res uses, *again* is **increasingly hard to be used as a decompositional** in absence of cps, i.e. as truly decompositional.

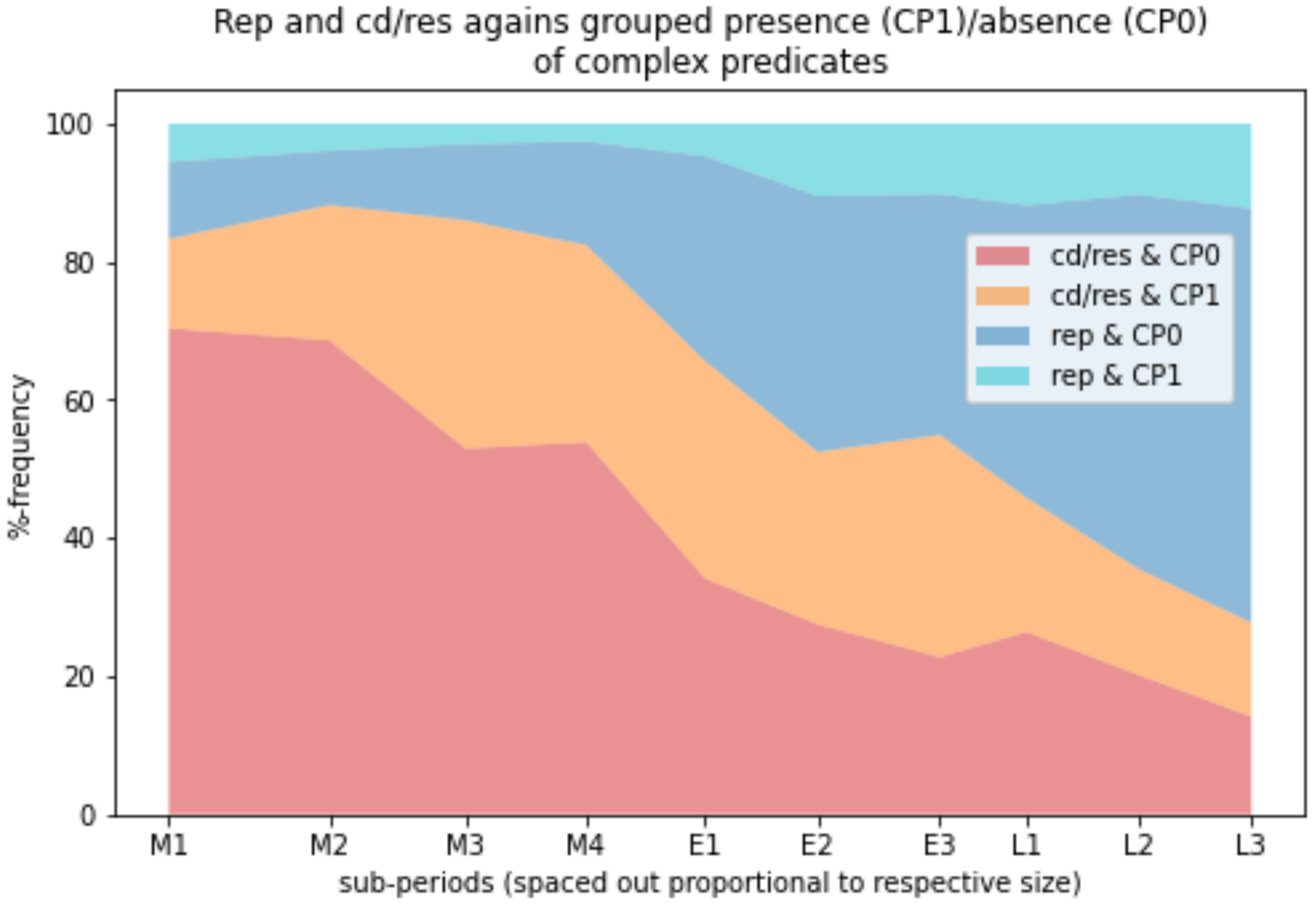
2. Conclusions: Overall modeling of entire development

Crucially:

From over ⅔ of *again*s occurring without cps in around 1150

↓

Towards the end of the LModE, less than 20% of all uses of *again* are decompositional (i.e. by approximation with our cp notion)



Outlook

- Complete the map of iterative PSP triggers:

  (graft annotations for *eft* and *re*-verbs onto YCOE and PPCHE)

- Closer look at **lemmas/semantic class** of verbs/predicates;

- Expand to other types of complex predicates, i.e. **resultatives**; cf. Snyder (2001), Beck (2005)

- Explore **cyclicity** and **presuppositional change** in iteratives and cps (*eft* and cps in OE data).

- Expand semantic annotations **beyond iteratives and decompositionality** to e.g. additives/scalar additive particles (*even*, *furdon*, *also*, *too*, …)

References

Beck, S. (2005). There and back again: A semantic analysis. Journal of Semantics, 22:3–51.

Beck, S., Berezovskaya, P., and Pflugfelder, K. (2009). The use of again in 19th-century English versus Present-Day English. Syntax, 12(3): 193–214.

Beck, S. and Gergel, R. (2015). The diachronic semantics of English again. Natural Language Semantics, 23,3: 157–203.

Cournane, A. (2017). In defense of the child innovator. In Micro-change and Macro-change in Diachronic Syntax. Oxford University Press.

Degano, M. and Aloni, M. (2021). Indefinite and free choice: When the past matters. *Natural Language and Linguistic Theory*, 40:447–484.

Garrett, A. (1998). On the origin of auxiliary do. English Language and Linguistics, 2(2): 283–330.

Gergel, R. and Beck, S. (2015). Early Modern English 'again': A corpus study and semantic analysis. English Language and Linguistics, 19(1):27–47.

Gergel, R. (2023). Yet another player in the *re* domain: the redirectional (*send-it-somewhere-else*) reading. Poster presented at: *Sinn und Bedeutung 28*; September 6, 2023; Bochum, Germany.

Karimi, S., editor (2013). Lingua SI: Complex Predicates, volume 135. Science Direct.

Karimi, S., Nematollahi, N., Kabiri, R., and Ngui, J. G., editors (2023). Advances in Iranian Linguistics II. John Benjamins.

Kodner, J. (2019). Estimating child linguistic experience from historical corpora. Glossa, 4(1).

Kroch, A. (1989). Reflexes of grammar in patterns of language change. Language Variation and Change, 1: 199–244.

Kroch, A., Santorini, B., and Delfs, L. (2004). The Penn-Helsinki Parsed Corpus of Early Modern English (PPCEME). Department of Linguistics, University of Pennsylvania, first edition. Release 3.

Kroch, A., Santorini, B., and Diertani, A. (2016). The Penn Parsed Corpus of Modern British English (PPCMBE2). Department of Linguistics, University of Pennsylvania, second edition. Release 1.

Kroch, A., Taylor, A., and Santorini, B. (2000). The Penn-Helsinki Parsed Corpus of Middle English (PPCMBE2). Department of Linguistics, University of Pennsylvania. Release 4.

Rapp, I. and von Stechow, A. (1999). Fast 'almost' and the visibility parameter for functional adverbs. Journal of Semantics, 16:149–204

Snyder, W. (2001). On the nature of syntactic variation: Evidence from complex predicates and complex word-formation. Language, 77: 324–342.

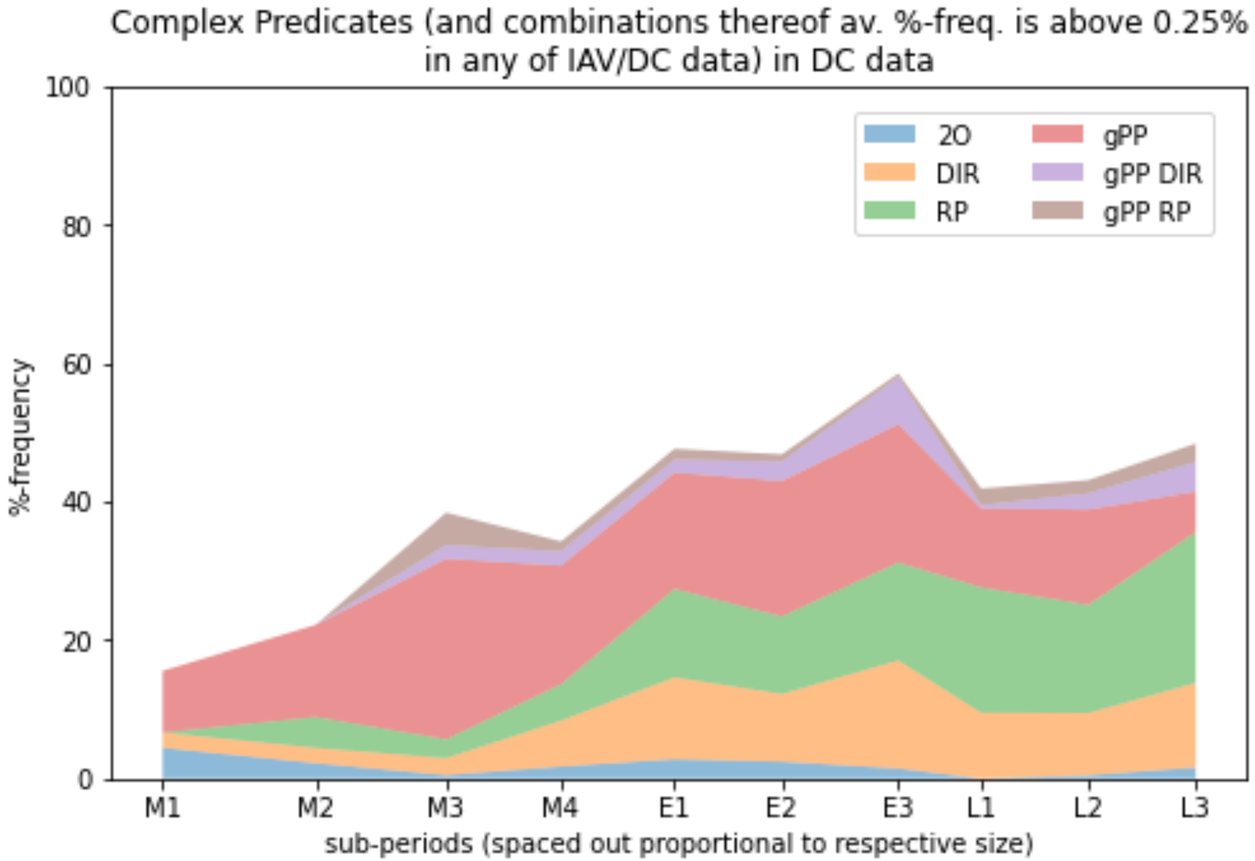
Warner, A. R. (1993). English auxiliaries: structure and history. CUP.

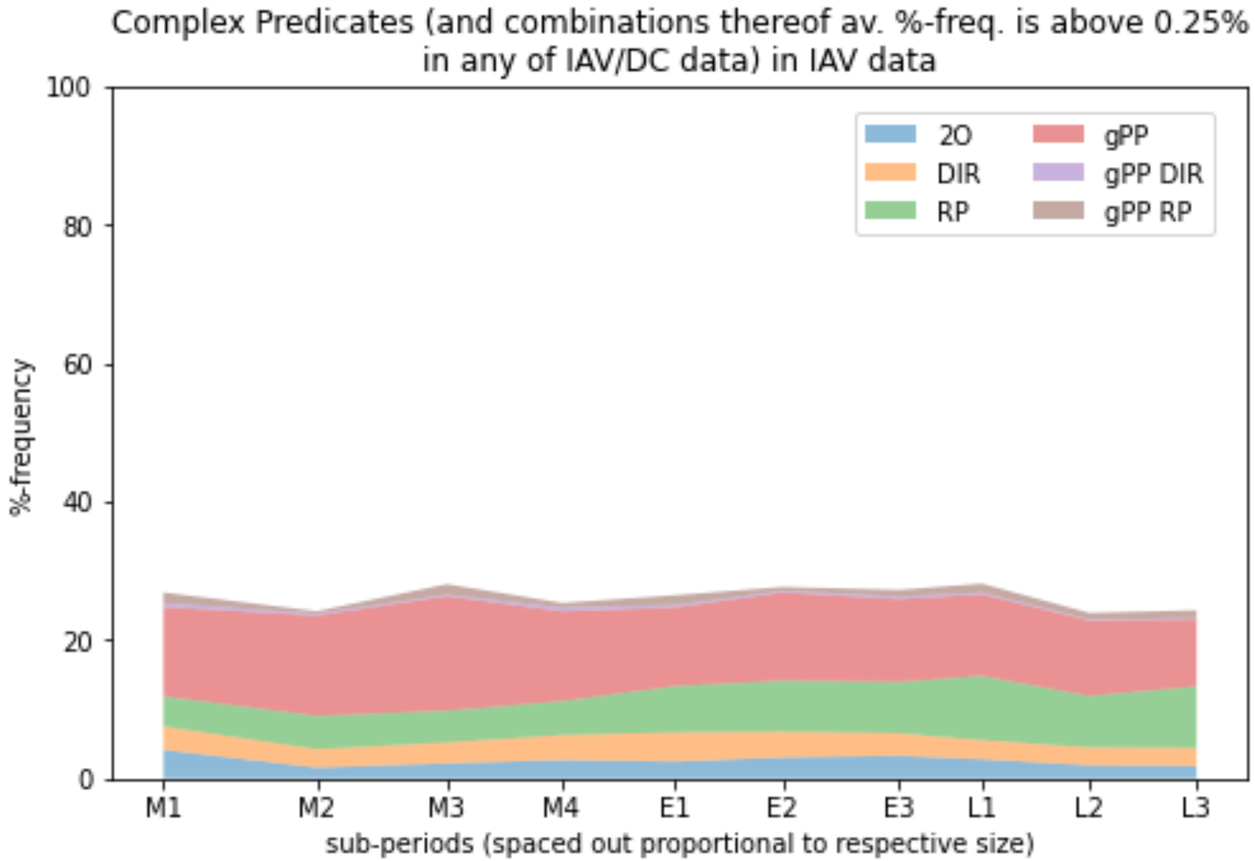
Thank you!

martin.kopf@uni-saarland.de, remus.gergel@uni-saarland.de

Appendix

Sup-types of complex predicates





Word counts PPCHE and %-frequency *again*