



Comparing annotation schemes across time

The problem of syntactic mapping

Eva Zehentner, Gerold Schneider, Marianne Hundt & Melanie Röthlisberger



Introduction

Prepositions and PPs:

- integral part of PDE language
- important role in system of verbal complementation
- abundance of previous research on individual aspects of prepositions
- BUT lack of a systematic, unified, longitudinal account of the diachrony of PPs in argument structure

PEAS (Prepositions in English Argument Structure Across Time and Space) **project:**

- bottom-up approach to PPs in argument structure across time and space (Middle English to PDE, World Englishes)
- annotation of data is a crucial prerequisite



Introduction

**PENN Parsed
Corpora**

PP-modification	1100-1500	1500-1569	1570-1639	1640-1710	1700-1800	1800-1914	Brown 1961	Penn TB 1981
verb-PP	59424	52115	53911	52599	58319	53815	49576	42151
noun-PP	27104	31036	29817	29460	33485	36953	36320	47629
noun-PP(of)	23897	27021	24495	23198	28428	30766	21514	22837
vPP / nPP	2.1925	1.6792	1.8081	1.7855	1.7417	1.4563	1.3650	0.8850

**ARCHER
automatically
parsed**

Rel	Direction	Early	Late	E	O/E(early)	Comment
obj	←	12213	8129	10171	1.502	<i>Bananas I like</i>
sentobj	←	2085	1458	1771.5	1.430	fronted subord. Clause
pobj	←	6681	5781	6231	1.156	Fronted PP (attached to verb)
pobj	→	61419	54457	57938	1.128	
...						



Outline

- background
 - the problem at hand: PPs and PP-attachment
 - syntactic annotation of historical corpora
 - different annotation schemes
- data and method
- results: extrinsic and intrinsic evaluation
- next steps



The problem at hand

PPs and PP-attachment: notoriously difficult, also for human parsers

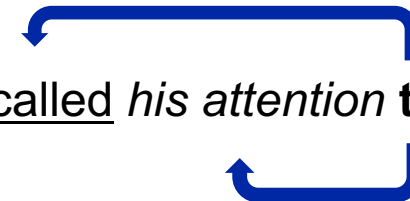
“PP (Prepositional Phrase) attachment ambiguity is one of the main ambiguities found in parsing” (Roh, Lee & Kim 2011: 559)

“What makes PP attachment particularly difficult is that the ambiguities can often not be solved using only structural preferences” (De Kok et al. 2017: 311)

(1) A man **in** the park saw a girl **with** a telescope



(2) President Bush called *his attention* **to** the matter



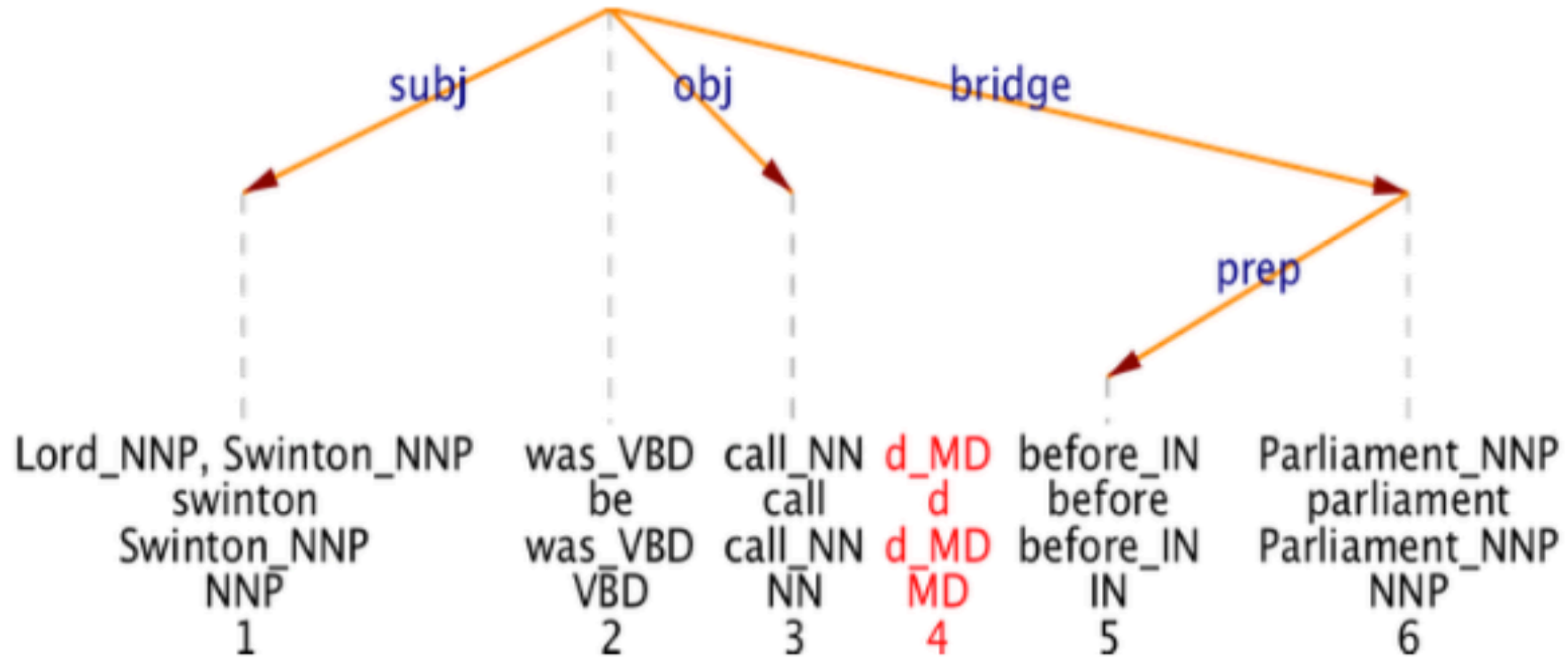


Syntactic annotation of historical corpora

- **Parsing is difficult: Ambiguity**
 - blind application of syntactic rules massively overgenerates
 - e.g. PP-attachment, leads to 100s of structures for long sentences
 - ambiguity is the dark side of collocation / Sinclair's two principles
- **Historical Parsing is even more difficult**
 - spelling variants → VARD
 - PDE expectations on lexical preferences
 - changes in more general features of linguistic system may have to be taken into account (e.g. increasing fixation of constituent order, loss of case-marking)



Syntactic annotation of historical corpora





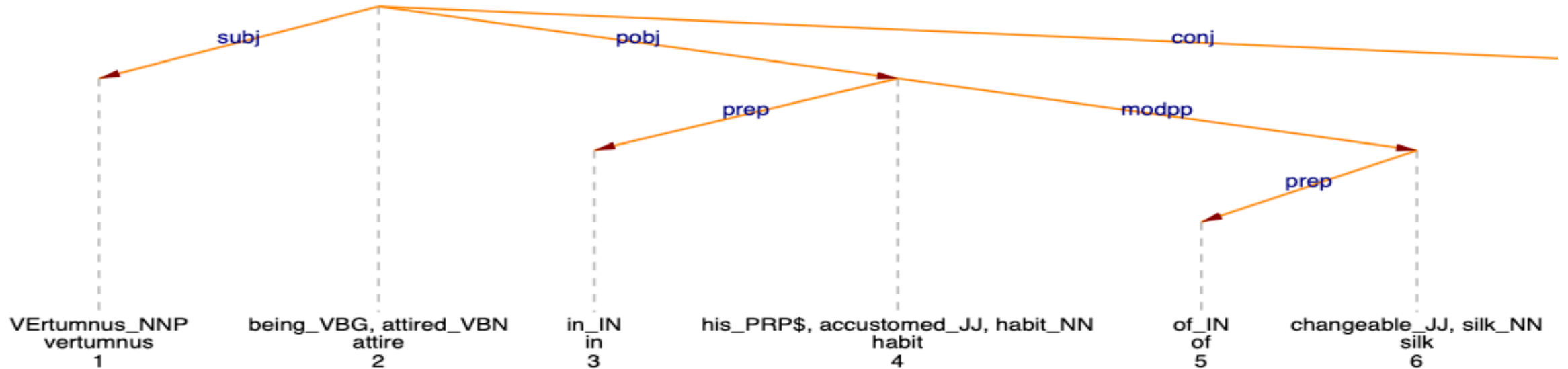
Different annotation schemes

Dependency vs. Constituency

- theoretically equivalent annotation subsets exist (Covington 1994), but many practical mapping problems
- dependency parser: dependency-based annotation scheme
 - word-based
 - constituents only a derived concept
 - no empty constituents
- Penn: constituency-based annotation scheme
 - syntactic functions are only partly annotated
 - CP/IP structure



Different annotation schemes



- subj = Subject
- pobj = PP attached to verb
- modpp = PP attached to noun=NP postmodification by noun
- conj = conjunction (the sentence continues ...)



Different annotation schemes

turning from Sam to Mr Pickwick with an air of some discomfiture

PENN output

```
(IP-PPL-SPE (VAG turning)
  (NP-MSR (PP (P from)
    (NP (NPR Sam)))
    (PP (P to)
      (NP (NPR Mr) (NPR Pickwick))))
    (PP (P with)
      (NP (D an)
        (N air)
        (PP (P of)
          (NP (Q some) (N
discomfiture))))))
  (. .))

(ID DICKENS-1837,558.363))
```

Dependency output

```
pobj('turn#20','Sam#22','from#21','(->)').
  prep('Sam#22','from#21','_','(<-)').
pobj('turn#20','Pickwick#24','to#23','(->)').
  prep('Pickwick#24','to#23','_','(<-)').
modpp('Pickwick#24','air#26','with#25','(->)').
  prep('air#26','with#25','_','(<-)').
modpp('air#26','discomfiture#28','of#27','(->))'.
  prep('discomfiture#28','of#27','_','(<-)').
```



Data and method

- periods covered: Early and Late Modern English (EModE, LModE)
- Penn-Helsinki corpora
- comparison of two annotation schemes (PENN Treebank vs. Pro3GRes)
- Pro3GRes (Schneider 2008) uses a manually written grammar (*competence*) and statistical disambiguation (*performance*)
- Pro3GRes needs preprocessing (spelling normalisation) to improve accuracy
- extraction of 250 sentences per sub-period from raw texts for evaluation of
 - recall (manually extracted all PPs that are complements of verbs)
 - precision (matched sentence IDs of sentences with manually extracted PP complements against PPs automatically retrieved from parsed corpora)



Data and method

Intrinsic evaluation

precision and recall of each parser

Extrinsic evaluation

compare the performance of the two annotation schemes with each other

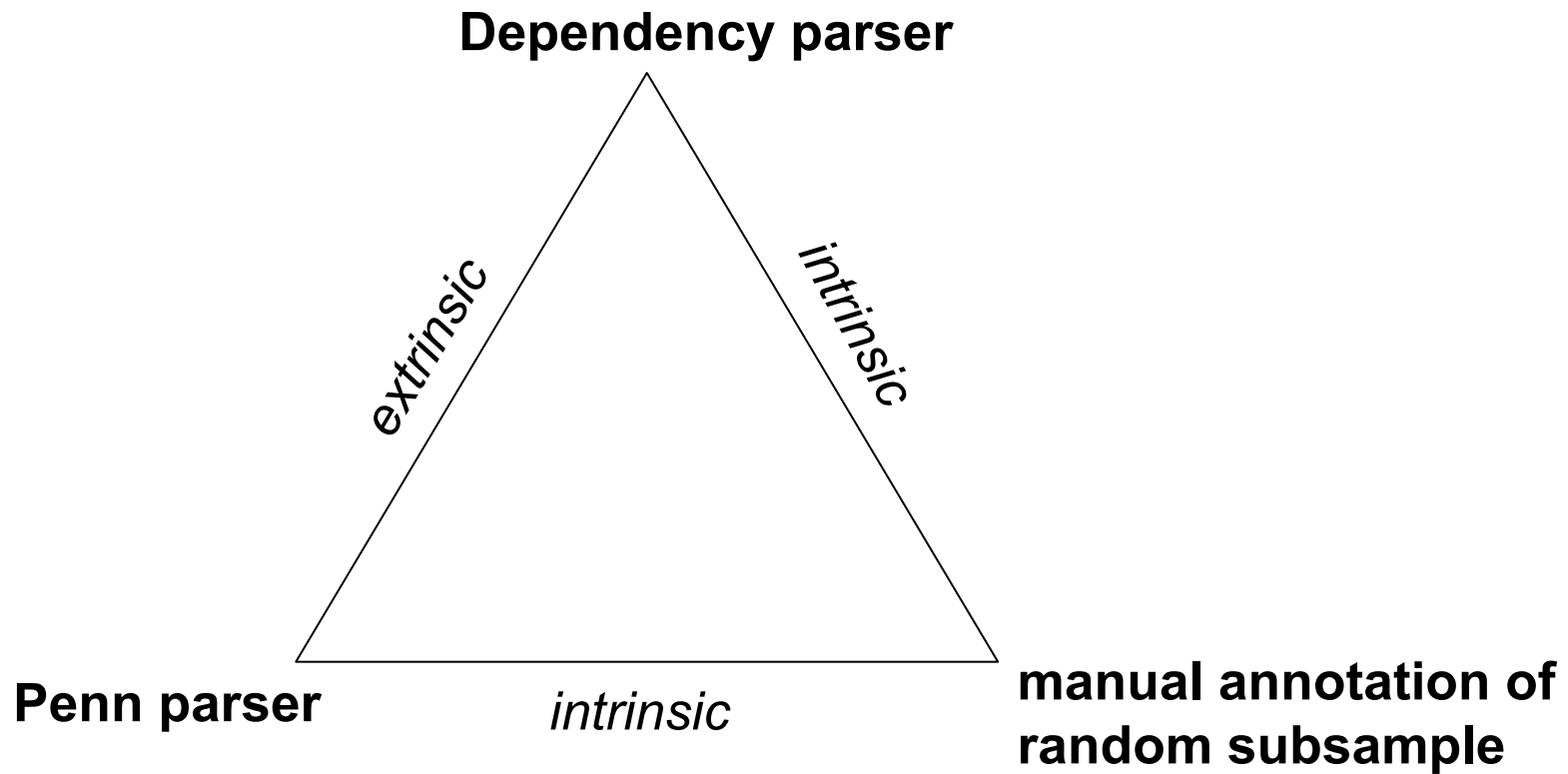
> parsers' performance with respect to a practical application

assess **the diachronic trends delivered by the two systems** in the application area of PP-attachment

(Mollá & Hutchinson, 2003)



Data and method



Data and method

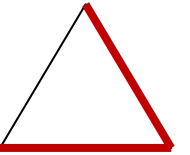
Case study on PP complements: retrieval from differently annotated data sets

- ❖ Dependency: *pobj* relation is part of annotation scheme
- ❖ Algorithm for extraction from PENN-annotated version

`/IP-?/=p<< (/^[VHB]|DA|DO/=verb $ (/PP/=pp < P=prep < (/NP/=descr<# /^(N|PRO|W)/=head))) : ID≠id`

Example:

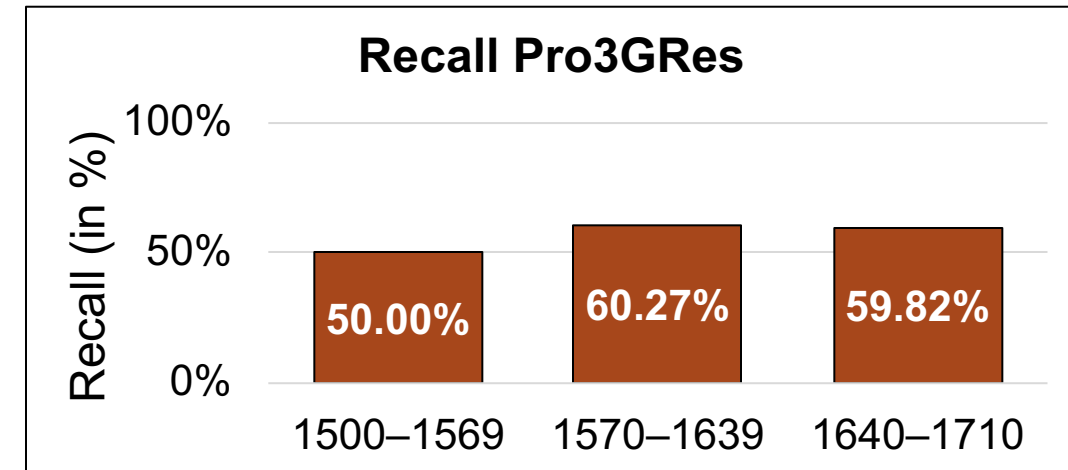
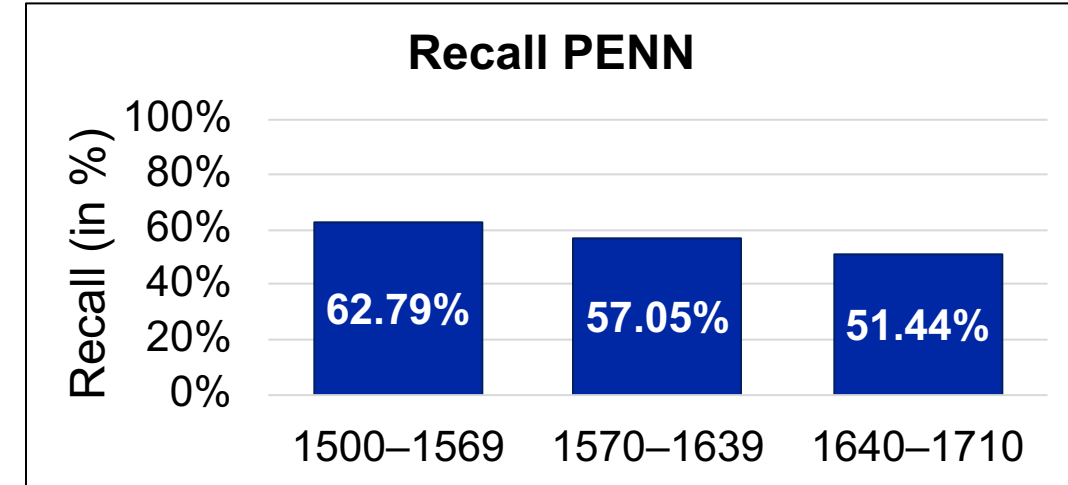
(**IP-SUB**=3 (NP-SBJ (N Judgment))
 (BE be)
 (NP-ADV (OTHER+N otherwise))
 (VAN given)
 (PP (P against)
 (NP (PRO him)))

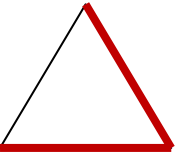


Intrinsic evaluation EModE

	PENN	Dependency
Recall overall*	57.0%	56.5%
Precision overall	98.9%	69.1%

* Recall is calculated without adjectival complements and without adjuncts

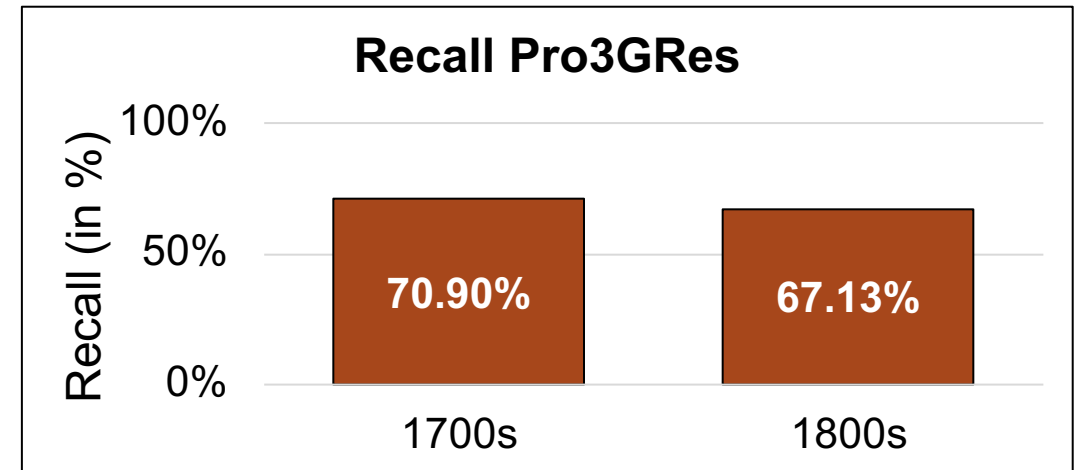
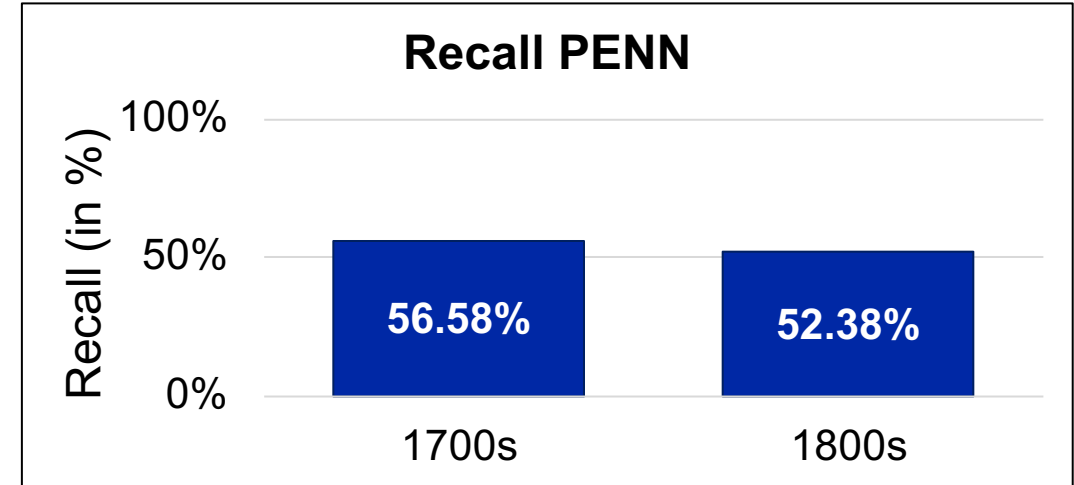


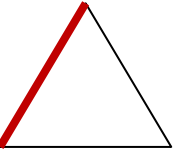


Intrinsic evaluation LModE

	PENN	Dependency
Recall overall*	54.5%	69.0%
Precision overall	89.0%	81.8%

* Recall is calculated without adjectival complements and without adjuncts



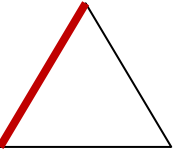


Results: Extrinsic evaluation

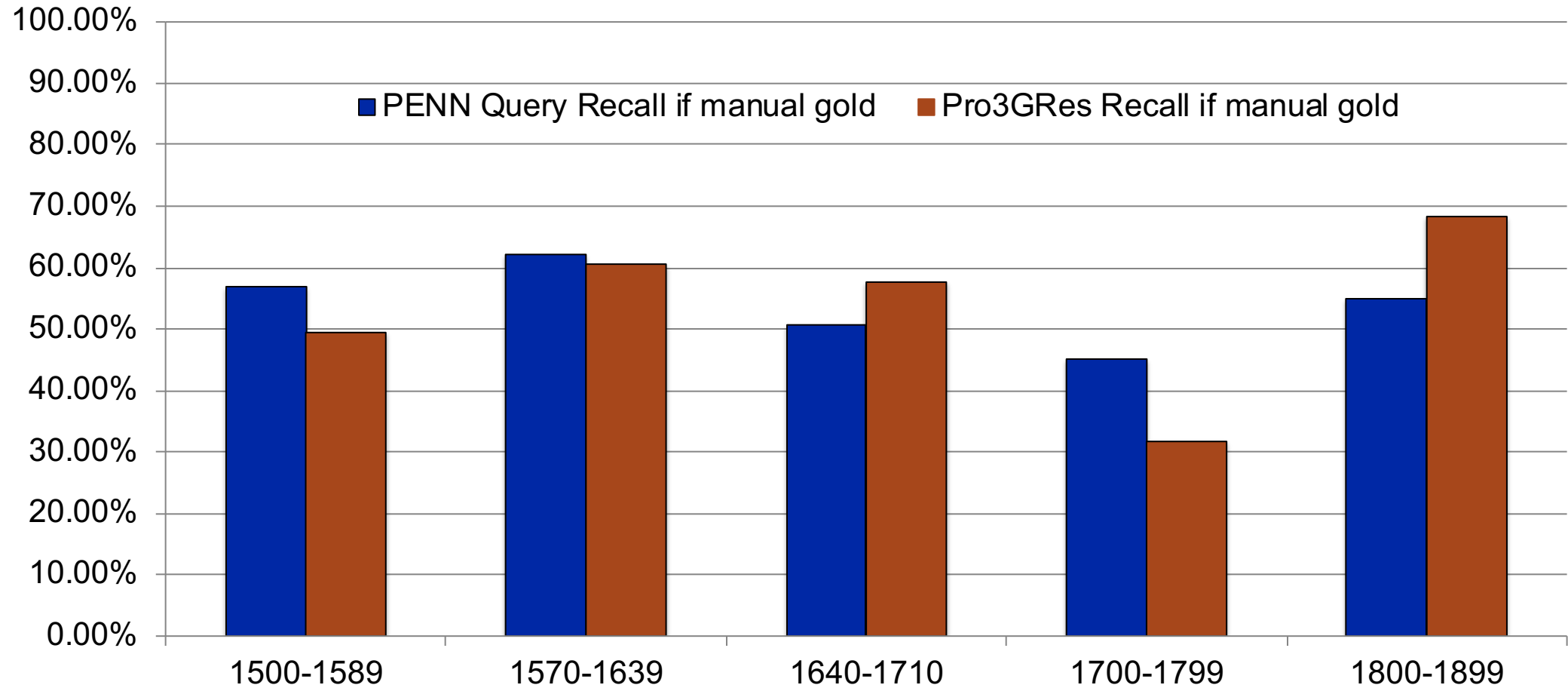
- Comparison in mapping between PENN annotation and Pro3GRes annotation

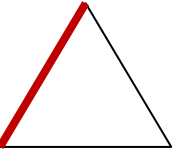
*If tried , it is a strange outlandish barbarity not **practiced** heretofore **in** England not to have you **advise of** counsel to **prepare for** a defence ,and , **in** order thereto , to **advise w=th=y=m=** . <CHATTON-E3-H,2,160.18>*

VERB	PREP	RECALL PENN	RECALL Pro3GRes	Pro3GRes vs. PENN	Meaning
<i>practice</i>	<i>in</i>	1	1	1-1	Both find it
<i>advise</i>	<i>of</i>	1	0	1-0	Only PENN: Pro3GRes Recall Error
<i>prepare</i>	<i>for</i>	1	1	1-1	Both find it
<i>advise</i>	<i>in</i>	0	0	0-0	Both fail
<i>advise</i>	<i>with</i>	0	1	0-1	Only Pro3GRes: PENN Recall Error

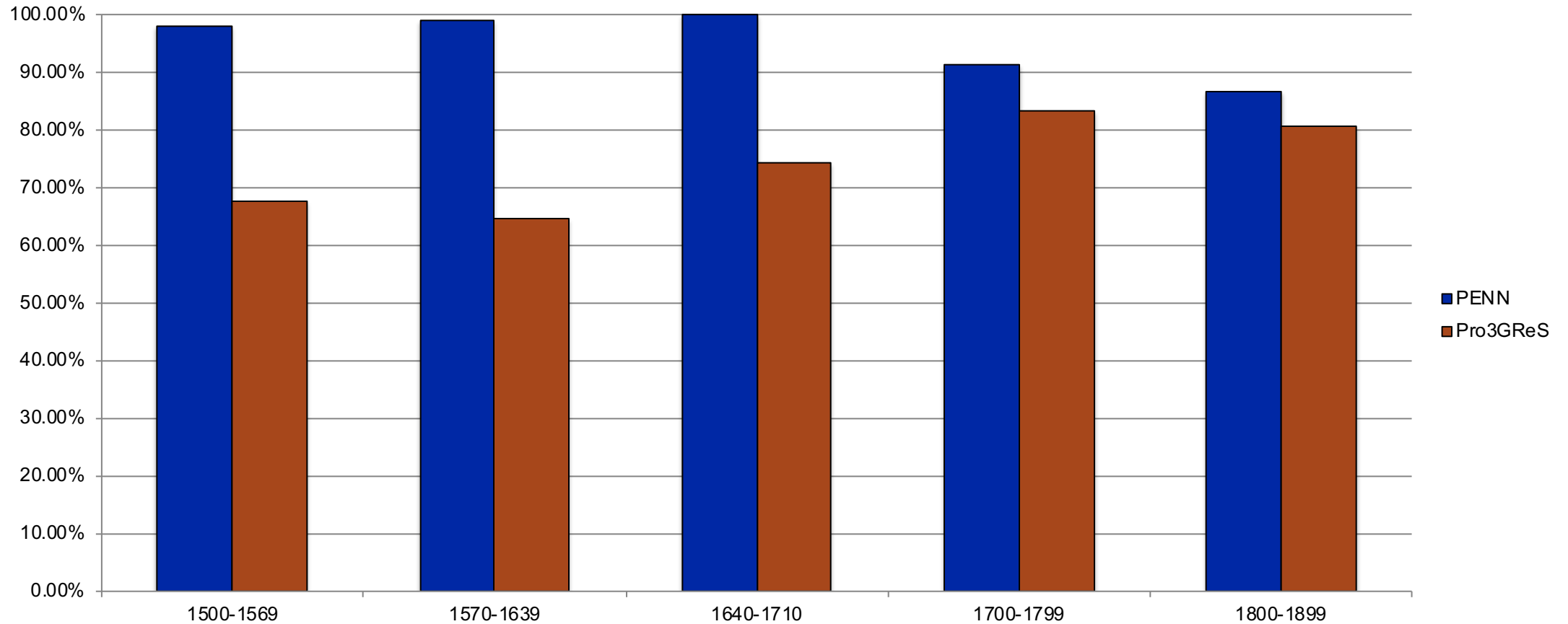


Results: Extrinsic evaluation – Recall





Results: Extrinsic evaluation – Precision





Potential problems of annotation sets

- **phrasal verbs**

- problem of category gradience (both synchronically and diachronically)

*You see how strangely different subjects for our consideration arise , when we come to **think over** these various matters ... <FARADAY-1859,11.109>*

- one of the most frequent tagging mistakes (low recall on _RP=verbal particle)

- **verbs with two prepositions**

*These go abroad young , either by the death of their parents , and nobody to look unto them , or else by some sharpe mistress that they serve , do **run away out of** service <HARMAN-E1-P1,75.386>*

→ Penn Parser always goes for the first preposition (*out*) but dependency parser usually goes for the first preposition-like constituent (*away*)



Potential problems of annotation sets

- **inter annotator agreement:** manual annotation as gold standard?

*And after he had humbly **asked** God **Pardon for it** , he desired me to call the Person to him , that he might ask him forgiveness <BURNETROC-E3-H,154.219>*

*The Lord Latimer **bouth** the **lands of the secund sister**. <LELAND-E1-P1,100.405>*



Conclusion & Outlook

- PENN queries have higher precision than the fully automatic dependency parser
- PENN queries are affected by recall problems
- comparison between annotation schemes can also be a forte (not just challenge) as it draws attention to categories missed by one annotation set (search query)
- fully automated approach delivers a roughly comparable signal
- improve the search query to allow for pre-posed PPs (i.e. verb list)
- annotate for all possible complementation patterns (not just PP) to show competition over time



Thank you!

www.prepcomp.es.uzh.ch



References

- Agire, E., Baldwin, T. & Martinez, D. 2008. Improving parsing and PP attachment performance with sense information. *Proceedings of ACL-08*: 317-325. [<http://www.aclweb.org/anthology/P08-1037>]
- Biber, D, Finegan, E. & Atkinson, D. 1994. ARCHER and its challenges: Compiling and exploring A Representative Corpus of Historical English Registers. In Fries, U., Tottie, G. & Schneider, P. Eds., *Creating and Using English Language Corpora*. Amsterdam: Rodopi, 1-14.
- Covington, M. A. 1994. *An Empirically Motivated Reinterpretation of Dependency Grammar*. Technical Report, University of Georgia.
- Kroch, A., Santorini, B. & Delfs, L. 2004. *Penn-Helsinki Parsed Corpus of Early Modern English* (PPCEME). <http://www.ling.upenn.edu/hist-corpora/> (13. August 2019).
- Kroch, A., Santorini, B. & Diertani, A. 2016. *The Penn Parsed Corpus of Modern British English* (PPCMBE2). <http://www.ling.upenn.edu/hist-corpora/> (13. August 2019).
- Mollá, D. & Hutchinson, B. 2003. Intrinsic versus extrinsic evaluations of parsing systems. *Evaluation Initiatives '03 Proceedings of the EACL 2003 Workshop on Evaluation Initiatives in Natural Language Processing: are evaluation methods, metrics and resources reusable?* Budapest, Hungary, April 14 2003, 43-50. (<https://www.aclweb.org/anthology/W/W03/W03-2806.pdf>) (20. November 2018).
- Santorini, B. 2016. Annotation manual for the Penn Historical Corpora and the York-Helsinki Corpus of Early English Correspondence. <http://www.ling.upenn.edu/~beatrice/annotation/> (13 August 2019).
- Schneider, G. 2008. *Hybrid Long-Distance Functional Dependency Parsing*. Zurich: University of Zurich PhD thesis.
- Schütze, C. T. 1995. PP attachment and argumenthood. *MIT Working Papers in Linguistics* 26: 95-151.
- Volk, M. 2001. Exploiting the WWW as corpus to resolve PP attachment ambiguities. In Rayson, P., Wilson, A., McEnery, T., Hardie, A. & Khoja, S. Eds. *Proceedings of the Corpus Linguistics 2001 Conference, Lancaster, 30 March – 2 April 2001*. Department of Linguistics. No pagination.



VARD (VARiant Detector)

<http://ucrel.lancs.ac.uk/vard/>

VAR D 2.5

undo/redo

Setup
/Users/setup folder VARD/developing/setup-eebo trained : EEBO Trained

Lucida Grande 12 text toolbar

F-Score
f-score weight
P 1/8 1/4 1/2 1 2 3 4 5 R

Method	F-Score	Precisi...	Recall
Known Variants	84.99%	93.76%	77.72%
Letter Rules	43.7%	43.7%	41.12%
Phonetic Matching	4.28%	4.28%	90.08%
Edit Distance	5.59%	2.05%	80.81%

method scores

Text Batch Train Setup Rules
Sidebar

Display
 Variants (357 tokens)
 Normalised (0 tokens)
 Not variants (4448 tokens)

Types List
 themselves (21)
 som (19)
 beeing (16)
 cours (12)
 neighbors (11)
 Countrie (10)
 hee (8)
 onely (8)
 anie (7)
 Incouragement (7)

Step Complete
 25/357
 themselves (94.15%)
 Normalise instance
 Normalise all
 Normalise to...
 Instance not variant
 All not variant

Auto Normalise
 Threshold (%): 50

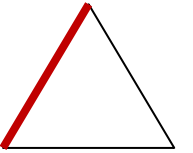
main text area

variant popup

current file and encoding

progress bar

THE ADVOCATE. ----- LONDON, Printed by William Du-Gard, Printer to the Council of State; and are to bee sold by Nicolas Bourn in Cornhil, at the Corner shop, at the entrance into the Exchange. 1652.
 TO THE RIGHT HONORABLE THE Council of State. Right Honorable! I Am often in very great doubt (if I may so speak), that the Goodness & Wisdom of God, & his thoughts of these, are very rarely met with in the Paths, which the scantling of Man's Reason and Judgment walk in; And as I dare not but own the Belief of the Coming of his Appearance, and the breaking forth, very shortly, of his Glorie: So I believ likewise, this will bee a sight very strange, and very unexspectd to men; and not onely greatly above, but in som measure even contrarie (and perhaps, very unwelcom) unto the most enlarged and raised thoughts wee have yet prepared our selvs with, to receiv it.
 The sign of which Coming, will bee the Detection, by little and little, of all Imposture, and the laying of all things low, naked, and mean before him; the stripping men of that Honor, Credit, and Repute, that they had by several means been gaining with themselvs, (and by themselvs) with others, either through a Not-knowledg, or through an artificial concealing of themselvs.
 In these very things (if well observed) lying, and within these indeed, beeing spent and consumed, the whole indeavors, Practise, Studie and Wisdom (if not Religion) of All States, Ages, Nations and Men, viz.
 Either in devising shifts (by acquisition of such and such Power, Habits, or otherwise) not to appear uncomly to themselvs, but to beget rather a greatness, esteem and satisfaction in themselvs, and others of themselvs;
 Or, in Concealing and keeping close (by several Glosses) the Imperfection apprehensions of others:
 The more exquisite main text area ferior onely in this, that the the other; And the most fortunate man in this world, having the Advantag covering of this fall's off a little sooner, and the other wrap's himself up in
 For thus all Lusts whatsoever (whether those of Acquisition) in the minde utmost Accomplishments, are still but either to give countenance unto, o The Assumption (on the other side) of a Repute and Esteemence to our se currancey (during it) to all Lusts. (These two, observing tune, and holding time each to the measures and motion therefore is concluded whole Humane Nature, with all the Parts, and the most perfect Actions of it.
 Upon these thoughts (Right Honorable!) I finding none not very ready to have asscribed to himself the name of beeing (in some measure) wise, becaus it is estimable, finde none that can endure the Manifestation (in any kinde) of their Follie; or that can bee willing to have this their Wisdom and Glorie to bee flatly and plainly called a Pageantrie, Mumming, or playing at bo-peep with themselvs and others; and yet (it really beeing no honest) if wee cannot bear the Judgments thus one of another; how shall wee abide the Trial, or bee able to stand before the Judgment of God, which seeth us, (and will shortly search us) with more pure and uncorrupt Eies? And how miserably then are the Imaginations of almost all men muffled up here in their own dressings? when the Truth most certainly is in our Beeing, at their best (natural, or acquir'd) Appearances, are yet such feign'd Things, and Propp'd up with such Lyes and fundamental mistakes of themselvs, that they would (Belshazzer like,) moulder and shatter themselvs to pieces, should they bee taken quite asunder from the disguises and patchings they have put upon themselvs, bare-facedly and undeceivedly to see themselvs but one small part of an hour.
 Beeing convinced, by these, therefore (Right Honourable!) of the low Condition wee are herein, I not knowing what the Councils of God intend to bring forth for the settlement of this Nation; Nor how hee hath resolved in his Wisdom to dispose of it, (as to its outward Condition,) whether Hee



Results: Extrinsic evaluation

TRIANGULATION OF RECALL in EModE	
156 11=both find manual gold	:-)
109 00=both fail manual gold	:-(
89 10=Pro3GRes but not PENN	PENN Query error if Pro3GRes gold
91 01=PENN but not Pro3GRes	Pro3GRes error if PENN gold
445 Σ	
20.00%	80.00% PENN Query error if Pro3GRes parser gold
20.45%	79.55% Pro3GRes Parser error if PENN gold
44.49%	55.51% PENN Query Error if manual gold
44.94%	55.06% Dependency Parser Error if manual gold
Error %	Correct %