

Retrieving prepositional arguments from ICE

1. Introduction

- “Finding the correct attachment site for prepositional phrases (PPs) is one of the hardest problems when parsing natural languages” (Volk 2001: 601)
- Firstly, the present paper reports on the intrinsic and extrinsic evaluation (Sparck Jones & Galliers 1995: 19) of the dependency parser Pro3Gres (Schneider 2008; Lehmann & Schneider 2011; Schneider & Lehmann 2012), which is available in the Dependency Bank.
 - Intrinsic evaluation: how well does the parser perform (according to its own grammar) in recognising verb-dependent PPs (precision/recall)?
 - Extrinsic evaluation: where does the parser pose problems for the linguistic analyses carried out by the PEAS project (e.g. the adjunct/complement distinction)?
- Secondly, Goldberg’s (2002) and Hoffmann’s (2007) categorisation of PPs as adjuncts or complements will be combined into one classification scheme, compared to the parser’s distinction, and evaluated.

2. Intrinsic evaluation

Variety	Register	Precision		Recall (prep = at)
		PP	VCOMP	
Nigerian English	Written	84%	92%	81%
Indian English	Spoken	72%	85%	72%
	Written	76%	95%	74%
Philippine English	Spoken	78%	85%	72%
	Written	76%	94%	75%

Table 1. Precision for PP attached to verb (PP), its subtype, PP attached to verb as complement (VCOMP), and recall for verb-dependent PPs headed by *at*

3. Extrinsic evaluation: problems

- Syntactic ambiguity: the parser chooses between noun or verb complementation, even when the instance is syntactically ambiguous.
 - *In a programme of that kind what is **happening** to people in different parts of India I think uh the important thing is not that you tell people who know about themselves.* (ICE IND S1b-025:1:145:B)
- Predicational vs. sentence adjuncts (Fig. 1):
 - The dependency bank uses dependency grammar, which means that the verb is the highest element of the sentence.
 - This entails that there is no distinction between sentence adjuncts and predicational adjuncts.
- Subjuncts, disjuncts, conjuncts: the parser does not have a separate category for subjuncts, disjuncts and conjuncts (vs. adjuncts) (Fig. 2).
- Adjuncts vs. complements (Fig. 3): the parser does not make a fine-grained distinction between adjuncts and complements.

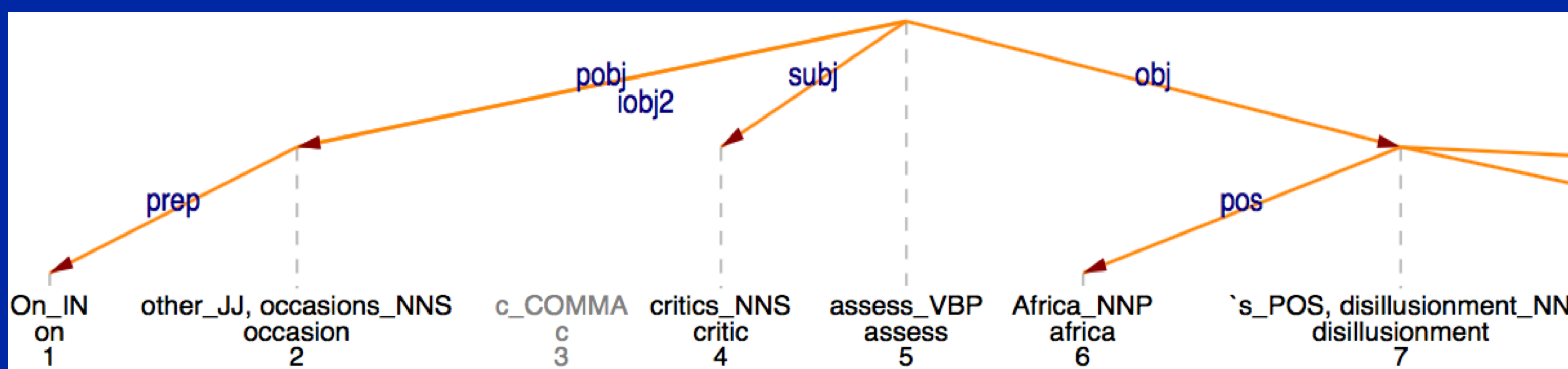


Fig. 1: Predicational vs. sentence adjuncts

On other occasions, critics assess Africa's disillusionment with its independences, which leave the continent in a dangling state [...] (ICE NIG: AHum_02:W2A-001:1:1)

- *On other occasions* is a sentence adjunct, but since the parser is a dependency parser, the PP attaches to the verb *assess* as the highest element in the sentence.

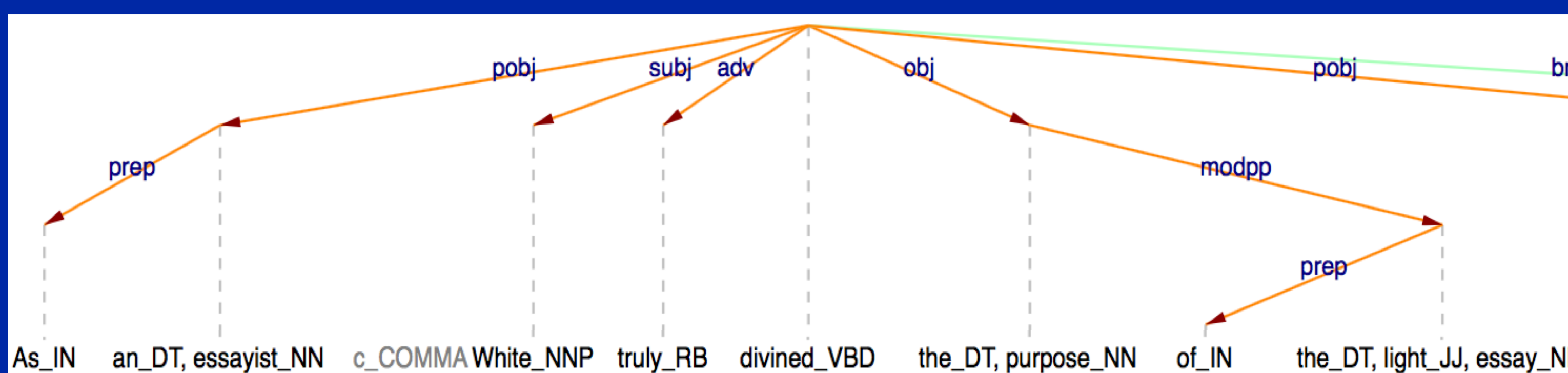


Fig. 2: Example of a subjunct

As an essayist, White truly divined the purpose of the light essay as an evocation of "the small things of the day, the trivial matters of the heart." (ICE IND: W2b-001:1:34)

- *As an essayist* is a item subjunct with subject-orientation, which “effects a characterization of the referent of the subject with respect to the process or state denoted by the verb” (Quirk et al. 1985: 574).
- There is no separate annotation for this class, which is furthermore not verb-dependent.

Obligatory	Optional
A) Complement of construction and verb Subcategorised P-type - Verbs that only license a specific prep e.g. I <u>rely</u> on you. - Verbs with an IO as prep e.g. I <u>gave</u> it to you. - Verbs that mean sth. else without a prep e.g. He <u>took</u> me for a teenager. Subcategorised PP-type e.g. She <u>put</u> the package on the table. Obligatory subject complement e.g. I <u>live</u> on the moon.	B) Complement of verb “Mixed adjuncts” when needed by the verb e.g. I <u>loaded</u> the wagon with hay. Optional complements e.g. We <u>talked</u> about everything.
C) Complement of construction Obligatory complement of the construction e.g. She sneezed the foam <u>off</u> the cappuccino.	D) Traditional adjunct Sentence adjuncts e.g. John <u>died</u> in Rome. Mixed PPs when not needed by the verb e.g. I <u>killed</u> the cat with a knife.

Fig. 3: Classification scheme of adjuncts and complements, combining Goldberg (2002) and Hoffmann (2007)

4. Goldberg and Hoffmann: towards an adjunct-complement distinction (Fig. 3)

Similarity between parser and Hoffmann-Goldberg (IndE)	Inter-rater reliability for Goldberg's classification (IndE)
72.67%	54%

- Problems with the Goldberg-Hoffmann categorization:

1. **Variety-specific/transcription error?**
 - We have taken *in action* on several persons [...] (ICE IND S1b-060:1:106:C)
2. **Adjunct or complement?**
 - [...] booking has been done *for you* (ICE IND W1b-026:1:235)
 - *From the corder* they are transported *to vasaeferece* [...] (ICE IND S2b-030:1:61:A)
3. **Obligatory?**
 - [...] for the purpose of extorting money *from him*. (ICE PHIL S2B-045:1:68:A)
4. **What kind of complement?**
 - [...] this definition is widely accepted *by researchers* [...] (ICE IND S2b-050:1:25:A)
 - [...] there is some reluctance in a part of them to go *to rural areas*. (ICE IND S1b-041:1:123:B)

5. Conclusion

- An intrinsic evaluation of the parser output with respect to PP-attachment (2.) is of key importance to the PEAS project. For the Pro3Gres parser, both precision and recall are quite high.
- The extrinsic evaluation of the parser pointed to some difficulties (3.), e.g. the distinction between complements and adjuncts. These are also difficult to distinguish manually (4.), mostly because theoretical distinctions (e.g. those found in Goldberg 2002 or Hoffmann 2007) are difficult to operationalise.
- A collocational analysis will be more useful to the project than a combination of Goldberg (2002) and Hoffmann (2007).

6. References

1. Goldberg, Adele E. 2002. Surface generalizations: An alternative to alternations. *Cognitive Linguistics* 13(4). 327-356.
2. Hoffmann, Thomas. 2007. Complements versus adjuncts? A construction grammar account of English prepositional phrases. *Occasional Papers in Language and Linguistics (University of Nairobi)* 3. 92-119.
3. Lehmann, Hans Martin & Gerold Schneider. 2011. A large-scale investigation of verb-attached prepositional phrases. *Studies in Variation, Contacts and Change in English 6* (Methodological and Historical Dimensions of Corpus Linguistics). 2007-VARIENG.
4. Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London/New York: Longman.
5. Sparck Jones, Karen & Julia R. Galliers. 1995. *Evaluating natural language processing systems: an analysis and review* (Lecture Notes in Computer Science; Lecture Notes in Artificial Intelligence 1083). Berlin/New York: Springer.
6. Schneider, Gerold. 2008. Hybrid Long-Distance Functional Dependency Parsing. Zurich: University of Zurich PhD thesis.
7. Schneider, Gerold & Hans Martin Lehmann. 2012. Dependency bank. *LREC 2012 Conference Workshop "Challenges in the Management of Large Corpora"*, 23-28. Istanbul, Turkey, 22 May 2012.
8. Volk, Martin. 2001. Exploiting the WWW as corpus to resolve PP attachment ambiguities. In Paul Rayson, Andrew Wilson, Tony McEnery, Andrew Hardie & Shereen Khoja (eds.), *Proceedings of the Corpus Linguistics 2001 Conference*, 601-606. Lancaster, 30 March - 2 April 2001: Department of Linguistics.

7. Contact

laetitia.vandriessche@es.uzh.ch
<https://www.es.uzh.ch/en/aboutus/team/lvandriessche.html>
<https://www.prepcomp.uzh.ch/en.html>