

# A PP attachment test suite

Patricia Fischer

October 7, 2020

## 1 Factors affecting PP attachment

### 1.1 Number of attachment candidates

How many attachment candidates does a sentence contain in addition to the correct head? The more candidates, the more difficult will the attachment be. In the examples, the correct head is labeled as *word<sub>h</sub>*, the incorrect head candidates as *word<sub>c</sub>*.

*Test suite variations:*

- 2 head candidates

e.g. *Der Vorsitzende<sub>c</sub> kündigte<sub>h</sub> bei der Prüfgesellschaft.*

- 3 head candidates

e.g. *Der Vorsitzende<sub>c</sub> kündigte<sub>c</sub> die Stelle<sub>h</sub> bei der Prüfgesellschaft.  
Der Hüfensenator<sub>c</sub> überzeugt<sub>h</sub> die Skeptiker<sub>c</sub> in den Wirtschaftsförderausschüssen.*

- 4 head candidates

e.g. *Die Museumsbesucher<sub>c</sub> entfalten<sub>c</sub> nach dieser Entdeckung<sub>c</sub> eine neue Beziehung<sub>h</sub> zum umgebenden Raum.*

Subjects tend to occur as attachment candidates under the following conditions:

- The PP directly follows the subject

e.g. *..., dass sie<sub>h</sub> als ehrenamtliche Vorsitzende ein dienstliches Handy hat.  
Die Nachfrage<sub>h</sub> unter Jugendlichen ist groß.*

- Verb between subject and PP

- Preposition *als*

- e.g. *Sie<sub>h</sub> stehen als Datenbanken unter Schutz.*

- Preposition *wie*

- e.g. *Sie<sub>h</sub> will das ausprobieren, wie so viele.*

- Preposition *mit*

- Separation of *was für*

- e.g. *Was<sub>h</sub> hampeln denn da für geometrische Formen über die Bühne?*

- Extraposition of heavy PP

- e.g. *In Bonn ist ein Buch<sub>h</sub> erschienen mit dem Titel “Kultur und Entwicklung”.  
Sämtliche Delegationen<sub>h</sub> haben dem Beschluss zugestimmt, mit Ausnahme Frankreichs.*

- e.g. *Es muss auch mal Schluss<sub>h</sub> sein mit dem Gesülze.*

- Ortspolitiker<sub>h</sub> haben erhebliche Probleme, sowohl in der SPD wie in der CDU.*

- Was<sub>h</sub> bleibt, außer dem Begriff Soziale Marktwirtschaft?*

In 4.15% of the cases where a PP attaches to a subject in TüBa-D/Z v11 UD, subject and PP are separated by the verb.

## 1.2 Length between left and right bracket of PP

Complexity here refers to the number of phrases within a PP. An example of complex PPs would be embedded PPs where one PP modifies the other. If the PP-internal noun occurs without a determiner, this would be considered less complex.

Increasing complexity is not necessarily correlated with more difficult attachment. More information can be beneficial for PP attachment, e.g. when the PP-internal noun is modified by an adjective and the adjective provides additional cues for the attachment of the PP.

*Test suite variations:*

- PP = preposition + noun  
e.g. *Details über Transaktionen*
- PP = preposition + determiner + noun  
e.g. *in einer Anzeige mitteilen*
- PP = preposition + determiner + adjective + noun  
e.g. *in einer anonymen Anzeige mitteilen*
- Possessive modifier  
e.g. *Geld floss über das Konto des Landesverbandes*
- Embedded PPs  
e.g. *eine Geldsammlung für die Flutopfer in Südpolen  
Sie sitzt in dem Kuratorium für das Altenheim*

## 1.3 Distance between PP head and PP attachment head

The attachment head of the PP is the token to which the PP as a whole attaches. The PP head is the token within the PP which attaches to a node higher up in the dependency tree. Since long-distance dependencies have been shown to be more difficult than dependencies across shorter distances, a negative effect of increasing dependency distance and PP attachment accuracy is expected.

As an increase in PP complexity may also increase the attachment distance, the two variations in Sections 1.2 and 1.3 can produce two conflicting hypotheses: Increasing complexity may result in easier attachment due to more information available for disambiguation, increasing dependency length can result in more difficult attachments.

*Test suite variations:*

- $d < 5$  intervening tokens  
e.g. *Im Januar<sub>d</sub> entließ<sub>h</sub> die Arbeiterwohlfahrt ihren Geschäftsführer.*
- $5 \leq d < 10$  intervening tokens  
e.g. *Deshalb entließ<sub>h</sub> die Arbeiterwohlfahrt ihren Geschäftsführer im Januar<sub>d</sub>.*
- $10 \leq d$  intervening tokens  
e.g. *Deshalb entließ<sub>h</sub> die Arbeiterwohlfahrt, die einen Neuanfang versuchte, ihren Geschäftsführer im Januar<sub>d</sub>.*

## 1.4 PP position

Qualitative analyses on the TüBa-D/Z revealed a noticeable number of errors for fronted PPs. A broad-coverage test suite should therefore include PPs in different positions. What is particularly important here is the position of the head and the resulting distance between PP and head (cf. Section 1.3).

*Test suite variations:*

- Fronted PP, topological field: Vorfeld  
e.g. *Im Januar entließ die Arbeiterwohlfahrt ihren Geschäftsführer.*
- PP occurs after its head, topological field: Mittelfeld  
e.g. *Die Arbeiterwohlfahrt entließ ihren Geschäftsführer im Januar.*

## 1.5 Argument structure

Results on the TüBa-D/Z with Hamburg dependency annotations showed that the distinction between prepositional phrases and prepositional objects is difficult for parsers. While this distinction is not made in the Universal Dependencies scheme, the test suite should still provide examples of this distinction to cover a wider range of PP-related attachment issues. It is possible in UD to mark prepositional objects by the enhanced, language-specific label `obl:arg`.

*Test suite variations:*

- Prepositional phrase (adjunct) with verbal/nominal head  
e.g. *Gelder flossen in ein Altenheim.*  
*Im Januar entließ die Arbeiterwohlfahrt ihren Geschäftsführer.*
- Prepositional object  
e.g. *Er ist verantwortlich für die Finanzen.*  
*Die Rechnung geht an die AWO.*  
*Er verweist auf das Dokument.*  
*Es kam zu Warnstreiks.*

## 1.6 Degree of compositionality

PP attachment is expected to be easier if the PP and its head are both part of an idiomatic expression, i.e. if the meaning of the phrase is highly non-compositional.

*Test suite variations:*

- Non-idiomatic PP-head combination  
e.g. *Sie gehen nach Hause.*  
*Die Pflanze wächst auf dem Balkon.*
- Idiomatic PP-head combination  
e.g. *Sie gehen auf die Barrikaden.*  
*Sie wächst dem Mann ans Herz.*

## 2 Possible enhancements

The following factors affecting PP attachment may be included in the test suite after further investigations.

## 2.1 PP attachment head class

Many prepositions have a preference for verbal or nominal heads, i.e. they occur more frequently with a verbal/nominal head. PP attachment should be easier if the correct head of the PP belongs to the word class to which the preposition attaches more frequently. Prepositions will be clustered by meanings (cf. Section 2.2).

*Test suite variations:*

- To be determined

## 2.2 Preposition meaning

Possible usages of prepositions include local, temporal, causal, abstract, propositional and case marking prepositions. Further analyses are needed to determine which preposition usages should be used in the test suite and which ones should be excluded.

*Test suite variations:*

- To be determined

## 2.3 PP attachment head position

- To be determined

## 2.4 PP noun synonym

Replacing the noun within the prepositional phrase by one of its synonym should not change the attachment.

*Test suite variations:*

- 

## 2.5 Semantic plausibility

If there is more than one head candidate for a PP, how semantically plausible are the different candidates? The more plausible the correct head of the PP is, the easier the attachment. Plausibility needs to be based on human judgements which will be rated on a 1–7 Likert scale.

*Test suite variations (reduced inventory as a basis for discussion):*

- Highly plausible attachment  
e.g. *die Antwort auf die Frage*
- Neutral attachment  
e.g. *Potentiale bietet die Perspektive aus Frankreich.*
- Implausible attachment  
e.g. *eine Hauptstadt ohne Macht*

# 3 Format

## 3.1 Metadata

- Number of intervening clauses between PP and its attachment head
- Distance between PP and attachment head