### The spy saw a cop with a telescope: Who has the telescope? - An attempt to understand the basic building blocks of ambiguous PP-attachment sequences

Mengzhu Yan & Michael Hai Nguyen Victoria University of Wellington, Aarhus University mengzhu.yan@vuw.ac.nz, normhtn@cc.au.dk

#### Abstract

This paper explores the problem of ambiguous PP-attachment by extracting information from a PPattachment corpus using Python. Cases of ambiguous PP-attachment involve sequences of the head words of the following type: verb > noun > preposition > noun. The head nouns of ambiguous PPattachment sentences, as well as aspects beyond head words, are investigated by testing a number of hypotheses using a corpus of thousands of real-world examples. The hypotheses are partially based on theory and partially on empirical evidence. The results support some theoretical claims while discarding others. For instance, one finding that supports an existing claim is that of-PPs always attach to NPs whose heads are classifiers. This kind of knowledge can be put into practice when parsing natural language.

Keywords: PP-attachment; Python; ambiguous; parsing

#### **1. Introduction**

#### 1.1 The problem of PP-attachment

Disambiguation is an important issue in many applications such as "speech recognition, machine translation, information extraction, and human-machine interaction" (Roth, 1998, p. 806). Disambiguation refers to the process that narrows down the meanings of an expression. Clearly, disambiguation is especially crucial when a correct semantic interpretation is required. Ambiguity can give rise to different interpretations and occurs at the lexical and the structural level, as shown in the examples in (1).

(1) a. I went to the bank yesterday.

(lexical ambiguity: the edge of a river or a financial establishment?)

b. Visiting scholars can be funny.

(structural ambiguity: "to visit scholars can be funny" or "scholars who can be funny?")

This paper is solely concerned with a specific type of structural ambiguity: Preposition phrase attachment (henceforth PP-attachment). PP-attachment ambiguity refers to the ambiguity between

the two possible attachment sites of a PP, as in (1b) above. A PP can attach to either a noun phrase (henceforth NP) or a verb phrase (henceforth VP). A sentence can thus have (at least) two possible parses:

(2) The spy saw a cop with a telescope.

language works

(structural ambiguity: "It was with a telescope that the spy saw a cop" or "It was a cop with a telescope that the spy saw"?)

The sentence in (2) *The spy saw a cop with a telescope*, the PP *with a telescope* can modify the act of seeing or describe the cop who carries a telescope. Thus, the PP has two potential attachment sites, and accordingly, the sentence has two readings. PP-attachment ambiguity often arises when there is a sequence of the following head words<sup>1</sup>, as in (3):

(3) Sequence in the verb phrase:  $V-N_1-P-N_2$ (verb - noun 1 - preposition - noun 2)

There is a vast amount of literature on disambiguating PP-attachments. One line of research is represented by Hindle and Rooth (1993). This approach relied on the frequencies with which particular verbs and nouns occur with particular prepositions. For instance, the verb *send* is usually associated with the preposition *to*. Therefore, it is most likely that *to* attaches to the VP in the context of *send NP to*. However, N2 was not considered in Hindle and Rooth's study (1993). Similar to Hindle and Rooth (1993), Kawahara and Kurohashi (2005) also used the frequency of the co-occurrences of particular verbs, prepositions and N1s. They extracted unambiguous examples from a large corpus to get lexical preferences, i.e. which lexical elements typically co-occurred. Their study thus contributes to the disambiguation of PP-attachments by providing more information on lexical preferences of the verbs and nouns.

Both Kawahara and Kurohashi (2005) and Hindle and Rooth (1993) mechanically identified truly unambiguous examples containing the sequence in (4) and (5): (i) when a PP follows a noun that precedes the verb, then the attachment of the PP is unambiguously attached to the NP, as in (4), or (ii) when the PP follows a pronoun, the PP is attached to the VP, as in (5):

- (4) The man with the hat kissed the girl.
- (5) *He kissed her on the beach.*

In this paper, we do not attempt to disambiguate PP-attachments. Instead, we aim to get a better understanding of the nature of PP-attachments by investigating the factors internal and external to the head word sequence in (3). The internal factors concern what kind of words in the position of head words determines the attachment site. The external factors concern everything else but this kind of information. By investigating internal as well as external factors, we suggest what kind of difficulties must be handled when disambiguating PP-attachments. Thus, from a practical point of view, our findings are useful in practical disambiguation of PP-attachments. Furthermore, the results that we get can be used to support or discard pre-existing theoretical claims. Therefore, our study also has theoretical significance. To our knowledge, little research has so far examined the very basic building blocks of PP-attachment ambiguity and tested theoretical assertions by doing corpus work.

#### 1.2 The corpus

For our current inquiry, we use the PP-attachment corpus (nltk.corpus.ppattach) that is included in the Natural Language Toolkit (NLTK package) (Loper & Bird, 2002). Python (version 2.7.8), one of the most popular languages for scientific computing, is used to extract information from the corpus. The corpus contains 28,000 PP-attachment decisions. We are, however, only going to use the training set which contains 20,801 decisions. The percentages of VP-attachments and NP-attachments are 47.8% and 52.2%, respectively. So in this corpus, the PP occurs slightly more often attached to the NP than to the VP. These calculations enable us to normalize and thereby compare different numbers of VP- and NP-attachments in different contexts. In the corpus, each instance is encoded as a PP-Attachment object, which is organized in the following way:

(6) Sentence number, verb, noun 1, preposition, noun 2, attachment.

For instance, in the entry *u'1351'*, *u'put'*, *u'dollars'*, *u'in'*, *u'equity'*, *u'N'*, *1351* is the sentence number; *put* is the verb; *dollars* is N1, *in* is the preposition; *equity* is N2; the attachment site is NP. To make the data more legible, we therefore remove the delimiter 'u' (1351: put dollars in equity "N") in the examples from now on.

#### 2. Investigating internal and external factors

In this section, we first examine the head word sequence in (3) and extract information from PPattachment corpus to investigate which factor influences the site of PP-attachment. Then we test the hypotheses partially based on theory and partially on empirical evidence using the corpus.

The complexity which we are dealing with can be illustrated in the following examples, where each element can determine the site of attachment:

(7) *The spy saw a cop with a <u>telescope</u>*.(8) *The spy saw a cop with a <u>revolver</u>*.

(Schütze & Gibson, 1999, p. 409)

(9) *The spy saw a <u>cop</u> with a telescope.*(10) *The spy saw a <u>star</u> with a telescope.*

Example (7) is an ambiguous sentence because the PP *with a telescope* can attach to both the VP *saw a star* and the NP *a cop*. The PP can modify the act of seeing, or denote a property of the cop, namely that he carries a telescope. However, in (8), we can infer from our knowledge of the actual world that the PP is attached to the NP (revolvers are not used for looking at something or somebody). In a similar vein, the sentences in (9) and (10) show that changing N1 can affect the attachment of the PP. It does not make any sense in any normal world that a star carries a telescope. That would be the meaning of (10) if the PP *with a telescope* were to attach to the NP *a star*.

From the examples above, it should be clear that particular lexical elements may determine the interpretation of a given sentence. Our general method for investigating internal factors is thus the following: In the sequence VN1PN2, we keep all three lexical elements constant but one (as in (7) (8) and (9) (10)). In this way, we can encapsulate some of the factors that determine the attachment site. Furthermore, we also inquire whether the attachment site may differ if all elements in the mentioned sequence are kept constant (external factors).

Using the corpus, we test the following hypotheses:

- Hypothesis I: If the verb selects an NP and a PP, then the PP is likely to attach to the VP.
- Hypothesis II: If the PP denotes time, then it is likely that PP attaches to the VP.
- Hypothesis III: If N1 is a pronoun, the PP cannot attach to this NP.
- Hypothesis IV: If N1 is a classifier<sup>2</sup>, and the PP is headed by *of*, the PP is attached to this NP.
- Hypothesis V: If two sequences are identical, i.e. the lexical elements are the same, external factors (e.g. the subject) outside the sequence can influence the attachment.

The hypotheses are based on theoretical considerations and initial empirical observations that concern aspects of the elements VN1PN2 (and factors outside it). The next subsections are concerned with these aspects of PP-ambiguity and with testing of the above-mentioned hypotheses.

#### 2.1 The valency of the verb (Hypothesis I)

One step towards understanding the nature of PP-attachment is to examine the valency of the verbs in question. Some verbs select PP-arguments, and these PPs are therefore presumably attached to these VPs, whereas the attachment of PP-adjuncts cannot be determined in the same way. Adjuncts are not mandatory in a sentence, such as *in the morning* in (11). To select is to require the presence of a certain syntactic argument (Chomsky, 1965). One typical example of the verb that requires a PP as an obligatory argument is *put*:

(11) John put the book in the room.(12) \*John put the book.

(Merlo & Ferrer, 2006, p. 346)

Without the argument *in the room*, the sentence (12) is not grammatical. Therefore, the PP most likely attaches to the VP if a PP is an argument of the V. We have now provided the theoretical justification for the following hypothesis:

• Hypothesis I: In clauses that contain verbs that subcategorize for an NP and a PP, the PP is most likely to attach to the VP.

In order to test this hypothesis, we extracted the number of PP attachments with the verb *put*. The numbers of PP-attachments to the VP and the NP were 116 and 27, respectively. It was unexpected that 27 instances of PP-attachments to the NP were found. The first 10 instances of PP-attachment to NP are given in (13):

(13) [(1351: put dollars in equity "N"), (2038: put them on streets "N"), (2052: put brakes acquisitions "N"), (6403: put resistance from interests "N"), (6441: put all of properties "N"), (9191: put plug for donations "N"), (11024: put all of million "N"), (11045: put seven of 17 "N"), (11722: put money at risk "N"), (11962: put bit of damper "N")]

These results indicate that even though PP-arguments are required by the verb, the immediately following PP does not necessarily attach to the VP. Firstly, this could be due to the fact that there might be cases of multiple sequences of PPs: if two PPs follow N2, one of them might attach to the NP while the other attaches to the VP. The obligatory argument is present, but it is not contained in the corpus because it comes after another PP, e.g. *He put the book <u>about biology</u> on the table*. Therefore, the hypothesis does not apply to these cases. If only one PP is present, it is indeed odd why the PP should attach to the NP. This is one only way of accounting for the data. Although this analysis may seem implausible to some extent.

Secondly, some idiomatic expressions might overrule the valency of the verb. *put brakes on something* is an example of such an idiomatic expression. Therefore, the hypothesis that the PP of a verb which selects a PP as an argument, must necessarily attach to the VP, is problematic. However, we might speculate that if the PP does not attach to the VP, then the whole sequence VN1PN2 might be part of an idiom.

In the case of *put*, given that the PP attaches to the VP 116 times and the NP 27, what we actually can say based on a comparison of the pp-attachments to NP and VP in the corpus is that if *put* is used, then the PP in most cases is attached to the VP. This is compatible with hypothesis I, which is thus theoretically as well as empirically motivated.

We might very well generalize this to other verbs with the same selectional requirements; we therefore examine these verbs. These verbs were encoded as NOM-NP-PP in NomLex (Macleod, Grishman, Meyers, Barrett, & Reeves, 1998), which is a dictionary containing information about nominalizations of verbs (e.g. *destruction* is the nominalization of the verb *destroy*) and the verbs themselves. We looked through the dictionary and searched for the string NOM-NP-PP, meaning that the verb from which the noun is nominalized requires an NP and a PP. Then we created a list with these verbs (*consign, equate, transfer, entrust*) with this specific selectional frame and checked numbers of NP- and VP-attachments. There were 19 attachments on the NP and 9 on the VP. The numbers are too small to draw any generalizations.

#### 2.2 Temporal N2 (Hypothesis II)

language works

In the previous section, we examined the verb and its influence on the attachment of the PP. In this section, we attempt to encapsulate the influence of N2 on the attachment site of the PP. In the two sentences below, N2 varies, while all other elements are constant:

(14) The environmental agency offered some exemptions from the law.

(15) The environmental agency offered some exemptions from the start.

language works

(Boland & Blodgett, 2006, p. 389)

In (14), the PP is an argument of the noun *exemptions* and is therefore attached to the NP. In (15), the PP is an adjunct, denoting the time of the act of offering; therefore, it is attached to the VP. These examples show that the fourth element (N2) can determine the site of attachment. More specifically, the PP *from the start* indicates time. According to Dahlgren & McDowell (1986, pp. 589-590), if N2 denotes time, then the PP attaches to the VP because only events have temporal adjuncts, which leads to the hypothesis below:

• Hypothesis II: If the PP denotes time, then the PP attaches to the VP.

In order to support the above findings and further test the hypothesis, we created a list of nouns that presumably typically denote time (e.g. *day*, *month*, *January*) and extracted the number of NP-attachments (139) and VP-attachments (539). The results show that PPs with temporal N2s more often attach to the VP than to the NP. The normalized numbers of NP-attachments and VP-attachments were 1.4% and 5.8%, respectively. So the PP attached approximately four times as often to the VP than to the NP if N2 was a time noun. Even so, the claim made by Dahlgren and McDowell was challenged. Nevertheless, the tendency is clear: if N2 denotes time, then the PP is most likely to attach to the VP.<sup>3</sup> Note that we have not tested all temporal N2s; we have only tested some temporal nouns which presumably typically denote time.

PPs that attach to NP are typically fixed expressions such as *year by year*, *week to week* as shown in (16). Also, *of*-PPs seem to attach to the NP (this might be related its genitive use, see also its use as a classifier in Section 2.4)

(16) [(5323: dropping year by year "N") (12625: was one of months "N") (17180: happens week to week "N") (18120: unveil proposal of year "N")]

#### 2.3 N1 as a pronoun (Hypothesis III)

This section examines the alternation of the site of PP-attachments by varying N1, as exemplified in the following two sentences:

(17) I saw the man in the park.(18) I saw him in the park.

In (17), the PP *in the park* can be either attached to the VP or the NP. In (18), the PP *in the park* is unambiguously attached to the VP. Therefore, N1 can be a factor that determines the site of the attachment. We deduced the following hypothesis (which is also grounded in Hindle and Rooth's (1993, p.108) and Kawahar and Kurohashi's (2005, p.192) assumption):

• Hypothesis III: If N1 is a pronoun, the PP cannot attach to the NP.<sup>4</sup>

We thus extracted instances where N1 is a pronoun. There were 33 instances of NP-attachments, which conflicts with the assertion above. 507 instances of VP-attachments were found. The normalized percentage for NP-attachments was 0.051% and for VP-attachments was 16.80%. So according to the annotation, it was extremely rare that PPs attached to a pronoun, but nonetheless, it contradicted with the assertion that PPs cannot attach to pronouns. We then looked closer at what kind of pronouns can be attached by the PP. The numbers of NP-attachments and VP-attachments are given in Table 1:

Pronoun type	NP-attachment
Reflexive pronouns (e.g. myself, yourself)	0
Personal pronouns (e.g. you)	15
Absolute possessive pronouns (e.g. <i>mine</i> , <i>yours</i> )	1
Demonstrative pronoun (e.g. this)	17
Wh-pronouns (e.g. which)	0

Table 1. The numbers of NP-attachments and VP-attachments

According to the annotation, a PP can attach to personal, absolute possessive and demonstrative pronouns. The only instance of absolute possessive pronoun is the string *prove mine for agency*. It is not inconceivable that *mine* is used as the regular noun *mine* (for instance a gold mine), but it could also be used as a possessive pronoun in the construction. The question raises whether the PP can attach to *mine* if it is used as an absolute possessive pronoun. In principle it is possible for a PP to attach to an absolute possessive pronoun, e.g. *He proved his sympathy for this agency. I proved mine for that agency*. But those instances are presumably extremely rare.

In general, demonstrative pronouns seem to be able to host PPs. One of the 17 instances where the PP is attached to a demonstrative pronoun is *including those in programs*. In the following (invented) context, the PP would attach to the demonstrative pronoun:

(19) *He was including these people in his tv programs when he was talking about "a group* of *people". I was not including those in his tv programs, but those in his films.* 

In this subsection, we have provided some possible counter-examples to the hypothesis that an PP cannot attach to N1 if N1 is a pronoun. We have shown that PPs attaching to possessive and demonstrative pronouns is in principle possible, perhaps in contexts where the noun phrase is only implied.

#### 2.4 N1 as a classifier (Hypothesis IV)

NPs whose head is a classifier generally have the following form:

(20) (Det)  $N_1$ , [PP of (Det)  $N_2$ ]

(Lehrer, 1986, p. 111)

One example of this would be *a slice of salami*, where *slice* is a classifier. Thus, according to (20), the preposition *of* usually attaches to the classifier, i.e. to the NP. This leads to the following hypothesis:

• Hypothesis IV: If N1 is a classifier, and the PP is headed by *of*, the PP is attached to the NP.

Lehrer (1986) furthermore offered a number of classifiers (e.g. *bunch*, *slice* etc.), which we use to test the hypothesis. The corpus returned 106 instances of NP-attachments and 36 of VP-attachments.<sup>5</sup> We then tested the co-occurrence of classifiers with *of* to compare the number of *of*-NP-attachments with the number of *of*-VP-attachments.<sup>6</sup> The corpus returned 98 instances of *of*-NP-

Journal of language Works

attachments and **no** *of*-VP-attachments. This suggests that if the PP is headed by the preposition *of*, and if N1 is a classifier, then the PP attaches to the N1.

We also counted the number of instances of PPs headed by other prepositions when N1 was a classifier. The number of instances of PPs headed by other prepositions was 8. Therefore, classifiers co-occured much more frequently with the preposition *of* than any other prepositions.

#### 2.5 External factors (Hypothesis V)

This section tests the following hypothesis:

• Hypothesis V: If two sequences are identical, i.e. the lexical elements are the same, external factors (e.g. the subject) outside the sequence can influence the attachment.

We found 55 instances of identical sequences that appeared both in NP-attachments and VPattachments, which seems to support Hypothesis V. This suggests that *something* outside the VN1PN2 sequence exerts some influence on the attachment. In the following, we suggest which external factors these could be.

The subject that is not included in the corpus could be a factor. It is reasonably clear that the PP with the telescope attaches to the NP in (21), whereas in (22), there are two readings: (i) either I used the telescope as an instrument to hit the man or (ii) I hit the man who carried or owned the telescope. The syntactic ambiguity of (22) can be illustrated by the tree structures in Figure 1.

(21) The lightning hit the man with the telescope.(22) I hit the man with the telescope.

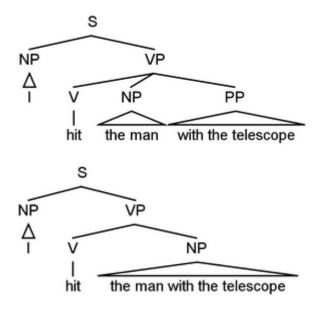


Figure 1. Tree structures of the ambiguous sentence *I hit the man with the telescope*. Another factor might be the overall context in which the sequence appears. See (23):

(23) The man fought the cat with a rapier.

If we assume that the cat in (23) is the fictive character *Puss in Boots* (who always carries a rapier), there are at least two interpretations: The man used a rapier to fight the cat, or the man fought the cat who carried a rapier. The latter interpretation is only possible in a fairy tale discourse.

A third factor concerns multiple PP-attachments (which was already discussed in Section 2.1.). Unfortunately, the corpus provides no information about the presence of multiple PP-attachments. Multiple PP-attachments introduce additional complexity in deciding the site of attachment:

## (24) ACET will shortly be opening a new office in the east end of London to serve clients in North and East London.

(Bharathi, Rohini, Vishnu, Bendre, & Sangal, 2005, p. 214)

The complexity of multiple PP-attachments can be illustrated by showing the different possible attachment sites in (24). The possible attachments of each PP are as follows (for expository reasons, we include only the head words of the relevant sequence):

open	office	in	end	of	L.	serve
	clients	in	L.			
V1	N1	P1	N2	P2	N3	V2
	N4	P3	N5			

On the face of it, the following PP-attachments are possible:

PP1 can attach to V1 or N1.

PP2 can attach to V1 or N1 or N2.

PP3 can attach to V1 or N1 or N2 or V2 or N4.

Bharathi et al. (2005) put forward two rules in order to reduce the number of possible attachment sites:

- A PP can only attach to elements within the same clause.
- If there are two following PPs after NP2, there are restrictions on where they can attach. If PP1 attaches to VP1, then PP2 cannot attach to NP1.

Applying the first rule, we can eliminate some possible attachment sites. PP3 occurs in the infinite clause headed by the verb *serve*. Therefore, it cannot attach to the elements outside this clause. We can reduce the complexity by ruling out the verb *open* as the possible attachment site. According to the second rule, if PP1 is attached to VP1, then PP2 cannot be attached to NP1. In other words, *cross-attachment* is not allowed:



Cross-attachment

Figure 2. Cross-attachment

To sum up, more information about the above-mentioned factors in this section would be of great help for more reliable results. It is to be kept in mind that some of the tests of the hypotheses involve a certain margin of error. The results may be contaminated by word forms that are spelled the same, but have different properties, for instance different meanings, selectional frames etc. For instance, we assumed that if N2 denotes time, then the PP also denotes time. This is only an assumption which is not true in all cases (even though it might be true in most cases).

#### Summary

In this paper, we examined the problem of ambiguity of PP-attachments. First, we introduced the problem of PP-attachment ambiguity. Then we generated a number of hypotheses that were partially grounded in theory or triggered by initial empirical evidence. We showed that PP-attachments were dependent on many different conflicting factors. We identified some factors that could increase the likelihood of VP- and NP-attachment, respectively (even though we could not fully test all hypotheses). From these factors we can create the most unambiguous and ambiguous sequences.

An unambiguous sequence in which the PP is attached to the VP could be created this way:

Choose a verb that requires a PP-argument.

Choose an N1 that is a pronoun.

Choose an N2 that denotes time.

Thus, two unambiguous sequences in which the PP is attached to the VP would be sentences such as:

(25) I put a book on the table.(26) I kissed him/a boy in December.

An unambiguous sequence where the PP is attached to the NP could be created in the following way:

Choose an N1 that is a classifier.

Choose *of* as the preposition (in fact it is the only preposition which goes together with the classifier, so in reality, there is no choice).

Thus, a very unambiguous sequence where the PP is attached to the NP would be a sentence such as:



(27) I bought a bunch of flowers.

#### 4. Implications

The results are theoretically interesting, but they are also applicable in practice. When parsing language, it is for instance useful to know which N1s are most likely to be attach to a PP. Despite the above-mentioned lack of information, this paper is a novel study that utilizes a corpus of thousands of real-world examples to explore PP-attachment ambiguity. In addition, this paper gives empirical knowledge and both lends support to existing claims and challenges some established claims. For example, the results seem to suggest that it is not totally true that some types of pronouns cannot be attached by PPs, as claimed by Hindle and Rooth (1993) and Kawahara and Kurohashi (2005). Furthermore, the results can be applied in parsing PP-attachment ambiguities. Our study is only one little step in disambiguating PP-attachments; more research is needed.

#### References

- Bharathi, A., Rohini, U., Vishnu, P., Bendre, S., & Sangal, R. (2005). A hybrid approach to single and multiple PP attachment using WordNet. In *International Conference on Natural Language Processing* (pp. 211-222). Berlin, Heidelberg: Springer.
- Boland, J. E., & Blodgett, A. (2006). Argument status and PP-attachment. *Journal of psycholinguistic research*, 35(5), 385-403. doi: 10.1007/s10936-006-9021-z
- Chomsky, N. (1965). Aspects of the Theory of Syntax. Cambridge, MA: MIT Press.
- Dahlgren, K., & McDowell, J. P. (1986). Using Commonsense Knowledge to Disambiguate Prepositional Phrase Modifiers. In *Fifth National Conference on Artificial Intelligence (AAAI-86)* (pp. 589-593). Menlo Park, California, USA: The AAAI Press.
- Hindle, D., & Rooth, M. (1993). Structural ambiguity and lexical relations. *Computational linguistics*, 19(1), 103-120.
- Kawahara, D., & Kurohashi, S. (2005). PP-attachment disambiguation boosted by a gigantic volume of unambiguous examples. In Dale, R., Wong, K.-F., Su, J. & Kwong, O.Y. (eds.) Natural Language Processing – IJCNLP 2005 (pp. 188-198). Berlin, Heidelberg: Springer. doi:10.1007/11562214
- Lehrer, A. (1986). English classifier constructions. *Lingua*, 68(2-3), 109-148.
- Loper, E., & Bird, S. (2002). NLTK: The natural language toolkit. In *Proceedings of the ACL-02* Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics-Volume 1 (pp. 63-70). Philadelphia, Pennsylvania, USA: Association for Computational Linguistics. <u>https://arxiv.org/pdf/cs/0205028.pdf</u>
- Macleod, C., Grishman, R., Meyers, A., Barrett, L., & Reeves, R. (1998). Nomlex: A lexicon of nominalizations. In *Proceedings of EURALEX* (Vol. 98, pp. 187-193), Liège, Belgium. (http://nlp.cs.nyu.edu/nomlex/ - accessed at 5<sup>th</sup> of January 2015).
- Merlo, P., & Ferrer, E. E. (2006). The notion of argument in prepositional phrase attachment. *Computational linguistics*, *32*(3), 341-378.
- Python Software Foundation. Python Language Reference, version 2.7. Available at <u>http://www.python.org</u>

Roth, D. (1998). Learning to resolve natural language ambiguities: A unified approach. In Proceedings of the Fifteenth National on Artificial Intelligence (pp. 806–813). Menlo Park, California, USA: The AAAI Press.

<sup>4</sup> This is not the case in Danish:

(i)	Ham	med det	lange	skæg	er	min	
	him	bedste ven with the friend	long	beard	is	my	best

'He who has a long beard is my beard is my best friend.'

<sup>5</sup> The list of classifiers includes: *bunch, bunches, dash, dashes, pinch, pinches, slice, slices, glimmer, gaggle, piece,* pieces, hundreds, clumps, species, herd, herds, flock, flocks, swarm, swarms, group, groups, pound, pounds, sprinkle, *sprinkles, swallow, swallows, cup, cups, mouthful, roomful, kind, kinds, type, types, sort, sorts.* <sup>6</sup> Note that some classifiers (e.g. *bunch, slice*) might be bare nouns, in which case they are not classifiers.

Schütze, C. T., & Gibson, E. (1999). Argumenthood and English prepositional phrase attachment. Journal of Memory and Language, 40(3), 409-431.

<sup>&</sup>lt;sup>1</sup> Other words might come in between, but we are not going to pursue this. The corpus does not deal with any other words than those mentioned in the above sequence.

<sup>&</sup>lt;sup>2</sup> A classifier is a measure word, occurring with a noun to classify the noun. An example of a classifier is the word bunch in a bunch of flowers in English.

<sup>&</sup>lt;sup>3</sup> It is not rare that the PP attaches to the NP in expressions like *year by year*, *week by week*.