

# Individuation and Semantic Role Interpretation in the Adpositional Domain

Sander Lestrade and Peter de Swart

## **Abstract**

In this chapter we discuss the role of Individuation on semantic role interpretation in the adpositional domain. Taking the findings of Aristar (1996, 1997) for the case domain as our starting point we examine whether similar observations can be made for the adpositional domain. On the basis of a corpus study in Dutch we determine whether adpositions show restrictions on the animacy and semantic roles of their complements and whether there is a correlation between the two. Our results suggest that only low-frequency adpositions show typing restrictions whereas we observe much variation with high-frequency ones.

## **1 Introduction**

Since the seminal work of Hopper and Thompson (1980) it is well-known that the transitivity of a clause and corresponding transitivity alternations interact with a number of parameters. Almost at the

same time Tsunoda (1981, 1985) has shown that the distribution of case frames over verb classes – so-called verb splits– is guided by similar principles. A much discussed parameter is that of Individuation of the patient argument of a transitive verb (O) which refers to argument properties such as animacy, definiteness, and number. Under the Hopper and Thompson approach a sentence with a highly individuated O, e.g. one that is animate and definite, is more likely to exhibit marking for Transitivity. Indeed we find that in many languages the presence of such a highly individuated argument results in a transitive (nominative/accusative) case frame or transitive object agreement, cf. the phenomenon of differential object marking (Bossong 1985; Aissen 2003; de Swart 2007; Malchukov 2008).

In the nineties Aristar (1996, 1997) has demonstrated that this notion of Individuation can be transferred from the domain of verbal core arguments into the realm of spatial cases. On the basis of data from a number of different languages he illustrates that spatial cases often show animacy restrictions on their arguments. Due to these restrictions the combination of animate referents and spatial cases may result in a morphologically more marked configuration than one with inanimate referents. Although the underlying motivation may be different, this pattern is reminiscent of the situation found in the verbal domain as discussed above. In his papers Aristar only briefly touches upon the question whether his findings can carry over from the spatial case domain to the adpositional domain. It is this question that is central to the present chapter.

Adpositions and spatial cases are two of the most commonly used

constructions to express spatial (locational and directional) meaning. A reasonable assumption seems to be that in languages lacking a spatial case system, like English and Dutch, adpositions fulfill a function similar to that of spatial cases, e.g. instead of using allative case to express direction one uses an adposition like *to*. Indeed, the expression of spatial meaning is often said to be the basic function of adpositions (cf. Levinson and Wilkins 2006, 5). Given this similarity between the two coding mechanisms, one may expect a similar interaction with individuation features in both situations. In this chapter we will try to give some ground to this expectation through a corpus study of Dutch investigating the interaction between animacy, semantic role interpretation, and adpositions. We will show that although low-frequency Ps seem to strongly select for inanimate complements such a restriction does not exist for high-frequency adpositions.

The chapter is structured as follows: in the next section we provide a more detailed discussion of the work of Aristar which will function as the background to our corpus study. Section 3 presents our research questions and their corresponding hypotheses. In section 4 the methodology of our corpus study is introduced together with its results. Section 5 provides a general discussion of our findings and section 6 concludes this chapter.

## 2 Animacy, Core Cases, and Spatial Cases

In a study on case-marking patterns in Australian aboriginal languages, Silverstein (1976) found that the split-ergative systems of these languages follow a hierarchy of features, which later appropriately became known as the Silverstein Hierarchy:<sup>1</sup>

- (1) 1,2 pro > 3 pro > PN > hum CN > anim CN > inan CN

The languages in his sample behave in such a way that an accusative case system is found with elements high in the hierarchy (i.e., members high in the hierarchy are marked when they act as direct objects, but not when they act as transitive subjects), and an ergative one with those low in the hierarchy (i.e., members low in the hierarchy are marked when they act as transitive subjects, but not when they act as direct objects). In other words, accusative case proceeds through the hierarchy top down, whereas ergative does so bottom up. This distribution can be seen as establishing a markedness pattern such that elements which are marked as an object or as a subject are morphologically marked with accusative and ergative case respectively. Note that in such a view nominative and absolutive case are in fact analyzed as the absence of case (de Hoop and Malchukov 2008). Indeed, these cases often have a zero expression whereas accusative and ergative case are nearly always overtly expressed.

---

<sup>1</sup>Abbreviations used in this paper: ABL ablative case, CN common noun, PN proper noun.

Aristar (1996, 1997) argues that similar hierarchical patterns can be used to describe the distribution of non-core cases in many languages: “A nominal’s degree of animacy influences its marking and interpretation in a non-grammatical case – a case such as the causal, instrumental, locative or ablative – just as it affects the marking and interpretation of A’s and O’s” (Aristar 1997:316).<sup>2</sup> Aristar studies the distribution of what he labels “non-grammatical cases”, also called semantic or local cases, such as locative and instrumental case, in comparison to the “grammatical” dative case. Semantic cases index or identify the semantic role of the argument they occur on. For instance, the locative case is used on noun phrases with the semantic role of Location and the instrumental case is used for those noun phrases with the semantic role of Instrument. The dative case, on the other hand, is most often used for arguments fulfilling the grammatical function of indirect object, which generally bear one of the semantic roles subsumed under Primus’ (1999) concept of Proto-Recipient, i.e., Recipient, Addressee, or Benefactive.<sup>3</sup>

Aristar reports that when we consider the relation between animacy and the cases mentioned above, we find that dative case is cross-linguistically oriented towards animacy whereas locative and instrumental cases are oriented towards inanimacy. In the light of the semantic roles associated with these cases, the observed pattern falls

---

<sup>2</sup>The present discussion is mainly based on Aristar (1997).

<sup>3</sup>Not for every case in every language such a unique function can be determined. Often, a case represents a blend of functions, which in many situations can be related to historical developments. For instance, the Latin ablative is a merger of the Indo-European ablative, locative, and instrumental case. A similar thing can be observed for the dative case in Classical Greek which also shows an instrumental use. See Kulikov (2006) for a discussion of the historical development of case systems.

out naturally: the Proto-Recipient role associated with the dative is typically fulfilled by animate entities, whereas locations and instruments are typically inanimate. When we translate this to the Silverstein hierarchy in (1), we can state that the dative case naturally extends from the top of the hierarchy to its bottom, whereas the non-grammatical cases do so the other way around. Thus, although the hierarchy can be used in describing the distribution of both core and non-core cases, it is involved in a different way: the former cases signal the unnaturalness of elements in the hierarchy with respect to grammatical functions, whereas the latter ones signal their naturalness.

It should be noted that although Aristar shows how these case-marking patterns can be described in terms of hierarchies, he actually claims that this hierarchical patterning is an epiphenomenon of selectional restrictions on case-arguments. He refers to this preference of cases to combine with certain referents as *typing*.

Cross-linguistically we can observe three strategies when a mismatch occurs between the type requirement of a case and the animacy feature of its complement. The first is the exclusion strategy which is found when cases are in complementary distribution. For instance, as Aristar (1997:346) reports, in Old Hittite the dative case only appeared with animate nouns, while the locative case only appeared with inanimate nouns. Likewise, in Basque benefactive case is reserved for animates, whereas instrumental case can only occur with inanimates. In other words, in these languages a type mismatch results in ungrammaticality.

A second strategy is to allow the co-occurrence of the mismatching entity with the case marker but to reinterpret the associated semantics, e.g. from directional to causal. This is referred to as *meaning extension* by Aristar and is very related to the notion of sense transfer and coercion (cf. Nunberg 1995; Pustejovsky 1995). Such meaning extensions can be viewed as functions from animacy values into semantic role interpretations: if a case marker which selects for a certain animacy value is confronted with an alternative animacy value it returns a role interpretation which is close to its lexical meaning but can accommodate the semantics of its argument. A clear case is when an instrumental marker is combined with an animate argument resulting in a comitative interpretation.

This second strategy often goes together with the third one, in which a type mismatch is resolved by the use of some additional morphology ('bridges' in Aristar's terminology). That is, in order to use dative case on an inanimate noun first some additional marking has to be added to the noun before the dative case ending can be attached and the same holds when a non-grammatical case is used on animate nouns. The latter pattern is illustrated in (2) and (3) for the ablative case in the Australian language Yidiny:

YIDINY (Australian; Aristar 1997:317)

(2)    mandi-m  
           hand-ABL  
           'from the hand'

(3)    buɲa:-ni-m  
           woman-MARKING-ABL

‘because of the woman’

The inanimate noun *mandi* ‘hand’ in (2) takes the ablative case ending directly. The animate noun *buna*: ‘woman’ in (3), by contrast, because it is a non-typical nominal for the ablative case, must first be marked with a special suffix before the ablative ending can be added. Moreover, an extension to a causal meaning can be observed in this example as well. Aristar takes this final pattern to be an instantiation of a general markedness pattern in which marked configurations, i.e., dative with inanimate/spatial case with animate, produce marked morphological expressions. It is this pattern that is identical to the original pattern found for core cases by Silverstein.

### 3 From Cases to Adpositions

In his work Aristar is mainly concerned with the interaction between Individuation and case forms. Our goal in this chapter is to investigate whether his findings for case-marking systems can be extended to adpositional systems. To us this seems a very natural extension: spatial case markers and adpositions show a strong functional similarity. This holds particularly when we compare languages with a pure adpositional system lacking spatial cases to languages with a large inventory of spatial cases. Relations expressed in the latter type by means of case forms will be coded in the former type by means of adpositions. It should be noted that the situation may be different for languages which combine a substantial inventory of adpositions with a number of spatial cases as in those languages we may find a specific



division of labor between the two coding mechanisms (see Lestrade 2008 for in-depth discussion).

Aristar only briefly touches upon adpositions in his discussion of the different strategies for dealing with a type mismatch between a case and its argument (see previous section). Referring to English, he argues that the strategy of ungrammaticality is also found in this language, something he illustrates with the following example:

(4) ??I killed Peter with Susan.

Aristar claims that the sentence can only be grammatical under a reading in which Susan acts as a ‘co-agent’ (Comitative), but not when ‘with’ is interpreted as Instrument. In this respect English follows the trend that ‘use of an instrumental preposition with a nominal of high animacy status is in many languages disallowed’ (Aristar 1997:317).

Also the strategy of meaning extension is argued to be found in English:

(5) He heard a sound from the house.

(6) He heard a sound from Peter.

In these two examples the NP following the preposition ‘from’ is interpreted as directional when inanimate and causal when animate. Under the assumption that the core meaning of this preposition is directional, we can analyze the causal interpretation found in (6) as a meaning extension due to the animacy of its complement.

As for the bridging strategy, Aristar claims that it does not oc-

cur in English. Indeed, it seems to be the case that English has no morphology dedicated to this specific phenomenon. The only element that comes close is ‘together’, as illustrated in (7):

(7) I killed Peter together with Susan.

Here, the use of ‘together’ opens up a strong preference for a co-agent or co-patient (both Peter and Susan are killed or Susan helped me kill Peter) reading which requires an animate argument. An inanimate NP like ‘hammer’ is infelicitous in this construction:

(8) #I repaired the table together with a hammer.

Aristar’s discussion shows that the patterns found in case systems are partially mirrored in an adpositional system like that of English. Due to the limited nature of the data discussed this connection can only be taken to be very tentative. In the remainder of this chapter we want to examine whether more ground can be given to the observations found in Aristar’s work. On the basis of a corpus study of adpositions in Dutch we want to find the answers to the following specific questions:

1. Do adpositions like case markers show typing restrictions with respect to animacy? That is, do we find adpositions that only occur with animate or inanimate complements?;
2. Do adpositions show typing restrictions with respect to the semantic roles with which they can occur? That is, do certain adpositions substantially limit the number of semantic roles their

complements can fulfill?;

3. Can the semantic role interpretation associated with an adposition be modeled as a function of the animacy value of its complement? The best described instance of such a scenario is the relation between inanimate Instruments and animate Comitatives (cf. Stolz et al. 2006; Narrog and Ito 2007), as found for instance with the English adposition ‘with’. The question is whether this is a one-time-only situation or whether we can find other such neatly distributed pairs of animacy-modulated semantic roles.

By answering these questions we get a better view on the role of individuation in the adpositional domain. Moreover, the answers will help us to establish whether Aristar’s findings for the case domain can indeed be transferred to the adpositional domain. In order to come to the answers of these questions we have conducted a corpus study on Dutch adpositions which is detailed in the next section.

## 4 A Corpus Study of Dutch Ps

### 4.1 The Corpus

For our corpus study we extracted data from the Corpus Gesproken Nederlands (CGN, Spoken Dutch Corpus).<sup>4</sup> The CGN is based on roughly 1000 hours of contemporary Dutch from the Netherlands and Flanders. The speech is composed of different genres, ranging from face-to-face and telephone conversations to interviews, debates, radio

---

<sup>4</sup><http://lands.let.kun.nl/cgn/ehome.htm>

shows and read aloud books. The speech files amounting to roughly 10M words have been orthographically transcribed, lemmatized, and tagged for part-of-speech information. Moreover, about 10% of the corpus has been syntactically annotated (van der Wouden et al. 2002).

From this syntactically annotated part of the corpus we have extracted all prepositional phrases. This amounted to 57,287 PP instances containing 139 unique adpositions and 12,947 unique heads in the adpositional complements. From this set we extracted all heads of the adpositional complements with a frequency higher than 10 occurrences. These 766 unique words were subsequently annotated by the two authors for their animacy using the coding scheme of Zaenen et al. (2004) which provides a 9-way classification. Where possible, disagreement was resolved by discussion. Of these 766 words, 154 were left out due to unresolved disagreement between the two annotators and 53 because they contained context-dependent elements, e.g. pronouns. Together the 559 words annotated for animacy (4% of the total number of unique complements heads) accounted for 23,369 (41%) of the adpositional phrases extracted from the corpus and combined with 103 unique Ps. This is in line with the observation that most of the complement heads only occurred once in our data set. In the final stage Zaenen et al.’s 9-way classification was reduced to a binary distinction between animate (human, organization, animal) and inanimate (place, time, concrete, non-concrete, machine, vehicles) in order to avoid problems with data sparseness.

## 4.2 Results

### 4.2.1 Do Dutch adpositions show restrictions with respect to animacy?

In order to find out whether Dutch adpositions show typing restrictions with respect to animacy, we counted the number of animate and inanimate complements of two sets of adpositions, high-frequency Ps and less frequent ones. For our corpus study, we understand typing as a manifestation of strong cooccurrence limitations in natural language use. For a P to exhibit animacy typing, it should exclusively occur with one type of complement, either animate or inanimate. The category of low-frequency Ps consists of all Ps with a frequency of occurrence between 9 and 30, the group of high-frequency Ps consists of the 12 most frequent Ps in our corpus ranging from 5474 occurrences for *in* ‘in’ to 408 four *mee* ‘with’.<sup>5</sup> In Figures 1 and 2, the counts of these two classes are given. As these figures show, Dutch adpositions indeed seem to show restrictions with respect to animacy. However, typing seems to be restricted to low-frequency Ps which almost exclusively select inanimate complements. That is, in natural language use they exhibit a strong tendency to occur with inanimate complements. This does not mean that they could not combine with animates at all; highly frequent Ps show less clear animacy restrictions.<sup>6</sup>

---

<sup>5</sup>Some of these Ps can only be used as postpositions in combination with words like *daar* ‘there’, as in *daarheen* ‘to that’, *daartoe* ‘to/for that’, *daarvandaan* ‘from that’, *daar doorheen* ‘through that’.

<sup>6</sup>This difference in typing behavior between the two categories is highly significant: Fisher Yates test,  $p \approx 0$ .

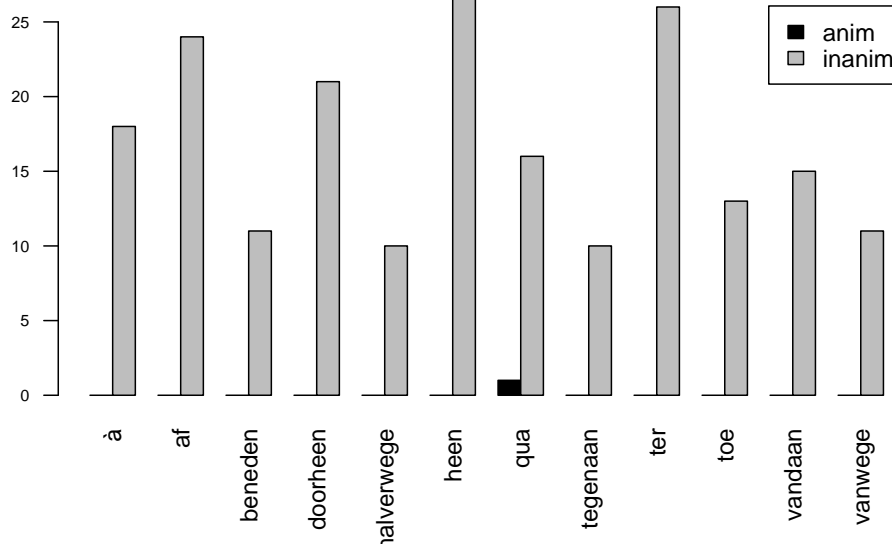


Figure 1: Animacy counts for less frequent Ps

#### 4.2.2 Do Dutch adpositions show restrictions with respect to semantic role?

To answer our second question we selected 10 Ps that occurred more than one hundred times in our corpus. These were *aan* ‘to, at’, *bij* ‘at’, *door* ‘through’, *in* ‘in’, *met* ‘with’, *op* ‘on’, *over* ‘over’, *per* ‘per’, *tegen* ‘against’, and *voor* ‘for, in front of’. The choice for this selection of adpositions was mainly informed by the authors’ expectations that interesting results could be obtained from it. For each P we randomly selected one hundred examples (approximately 50 animate and 50 inanimate ones) that were annotated for their semantic role by the two authors.<sup>7</sup> In our annotation we used the 22 role distinctions

<sup>7</sup>For some adpositions, the corpus did not contain 50 examples with animate complements. In this case we included all of the materials. Number of tokens per P: *aan* 91, *bij* 89, *door* 92, *in* 72, *met* 98, *op* 85, *over* 98, *per* 69, *tegen* 102, and *voor* 87.

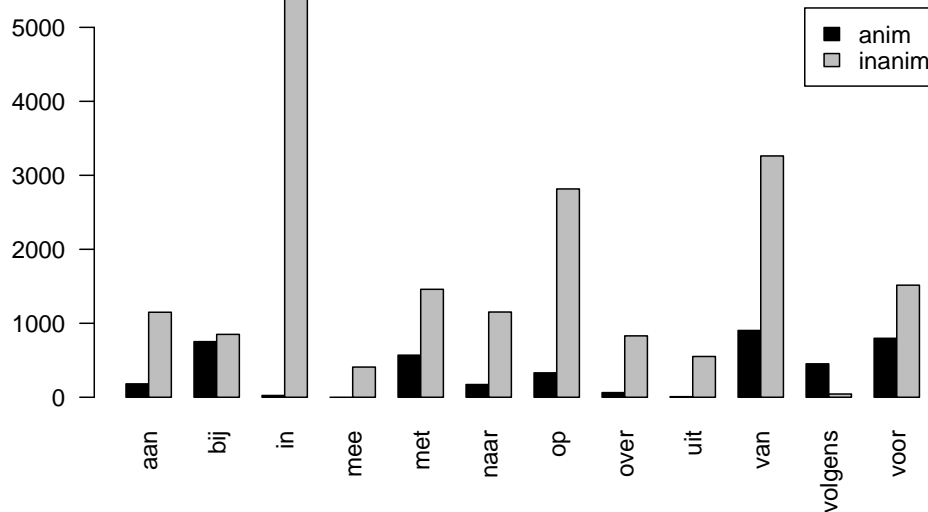


Figure 2: Animacy counts for highly frequent Ps

as given in Table 1 in the Appendix. The  $\kappa$  score for interannotator agreement was .76 (78%), which is reasonably good and allows us to draw conclusions (Carletta 1996).<sup>8</sup> Disagreement between the annotators was solved through discussion, such that each example ended up with a unique semantic-role tag.

In Figure 3 the distribution of semantic roles for the prepositions *aan* ‘to’, *bij* ‘at’, *door* ‘through’, and *voor* ‘for’ is given. As can be seen, some adpositions cover a wide range of semantic roles, others are more specific.

*Door* ‘through’ most often expresses the semantic role of Agent. This role is by definition restricted to animate complements (cf. Table 1 Appendix). In combination with inanimate complements, *door* ei-

<sup>8</sup>The kappa statistic is used to correct for expected agreement.

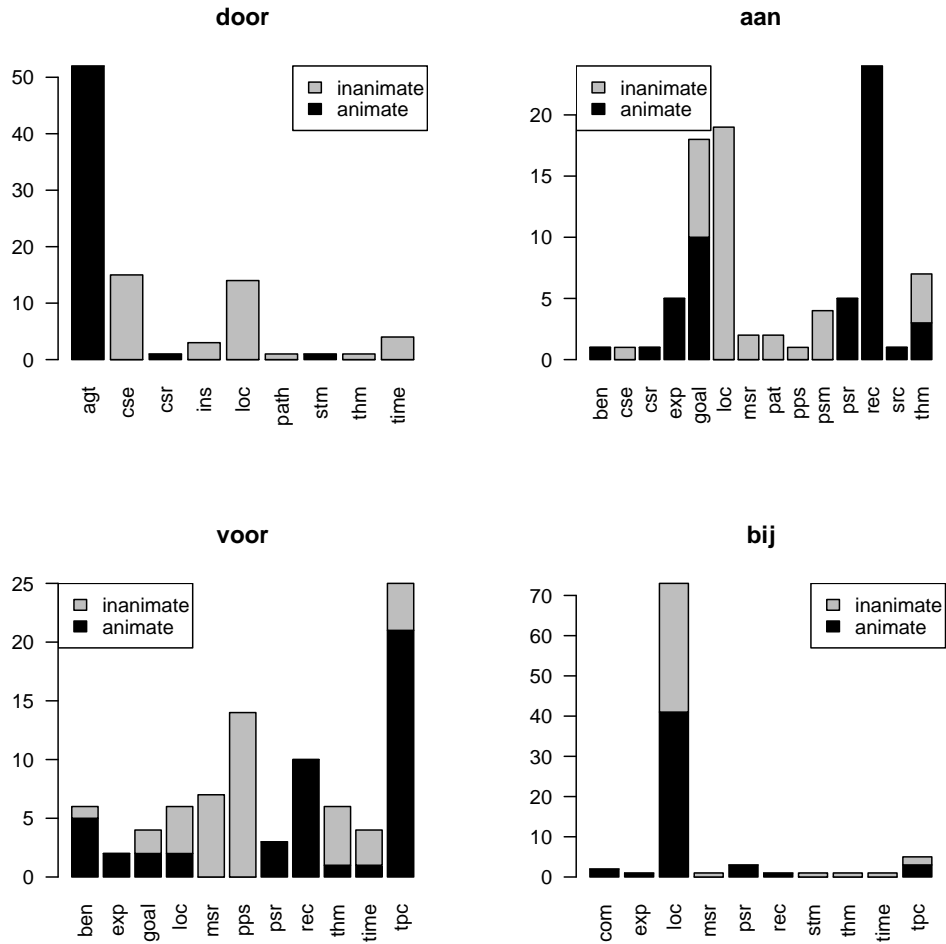


Figure 3: Semantic role distribution per P

ther expresses Location or Cause. An example of *door* ‘through’ with an animate and an inanimate complement is given in (9), semantic roles are given between parentheses:

- (9) a. zal daar nog sneller **door** worden  
 will there even faster through become  
 ‘will become even faster because of that’ (Cause)



- b. op het juiste moment gegeven **door** Bergkamp  
 on the right moment given by Bergkamp  
 ‘passed on the right moment by Bergkamp’ (Agent)

*Aan* ‘to’ shows a broader distribution of semantic roles. The three most prominent ones are Goal, Location, and Recipient. Less frequent roles of *aan* are Experiencer, Theme, Possessor, and Possessum. An example of *aan* ‘to’ with an animate and an inanimate complement is given in (10):

- (10) a. dat kan je wel **aan** 'm toevertrouwen  
 that can you indeed to him entrust  
 ‘you can leave that to him’ (Recipient)
- b. **aan** de telefoon vroeg zij van of uh...  
 on the telephone asked she of if uh...  
 ‘on the phone, she asked like...’ (Location)

Just like *aan* ‘to’, *voor* ‘for’ has a rather broad distribution. Its three most important roles are Purpose, Recipient, and Topic. An example of *voor* ‘for’ with an animate and inanimate complement is given in (11):

- (11) a. artsen deden het werk dan ook **voor** niets  
 doctors did the work then also for nothing  
 ‘doctors did the job for nothing’ (Purpose)
- b. dat was wel belangrijk **voor** 'm  
 that was indeed important to him  
 ‘for him, that was quite important’ (Topic)

The preposition *bij* ‘at’ almost exclusively expresses Location. An example of *bij* with an animate and inanimate complement is given

in (12):

- (12) a. blijf nog 'ns een lang weekend **bij** jou  
stay again once a long weekend at you  
'stay again a long weekend with you' (Location)
- b. uit de rij bootjes die dicht **bij** het strand  
from the row boats that close at the beach  
dobberen  
float  
'from the row of boats that float close to the beach' (Location)

In Figure 4, the distribution of semantic roles for the prepositions *in* 'in', *met* 'with', *op* 'on', and *per* 'per' is given. We find the same two patterns here, with some adpositions having a more specific meaning than others.

*Met* 'with' predominantly expresses Comitative. Instrument, Possessum, and Theme are some of its minor roles. An example of *met* 'with' with an animate and inanimate complement is given in (13):

- (13) a. waar zouden we naartoe kunnen gaan **met**  
where should we to can go with  
mijn ouders?  
my parents  
'where could we go with my parents?' (Comitative)
- b. voel maar 'ns **met** je tong  
feel but once with your tongue  
'try with your tongue' (Instrument)
- c. en ze kwamen steeds maar niet **met** de auto  
and they came always but not with the car  
'all that time they didn't come by car' (Vehicle)

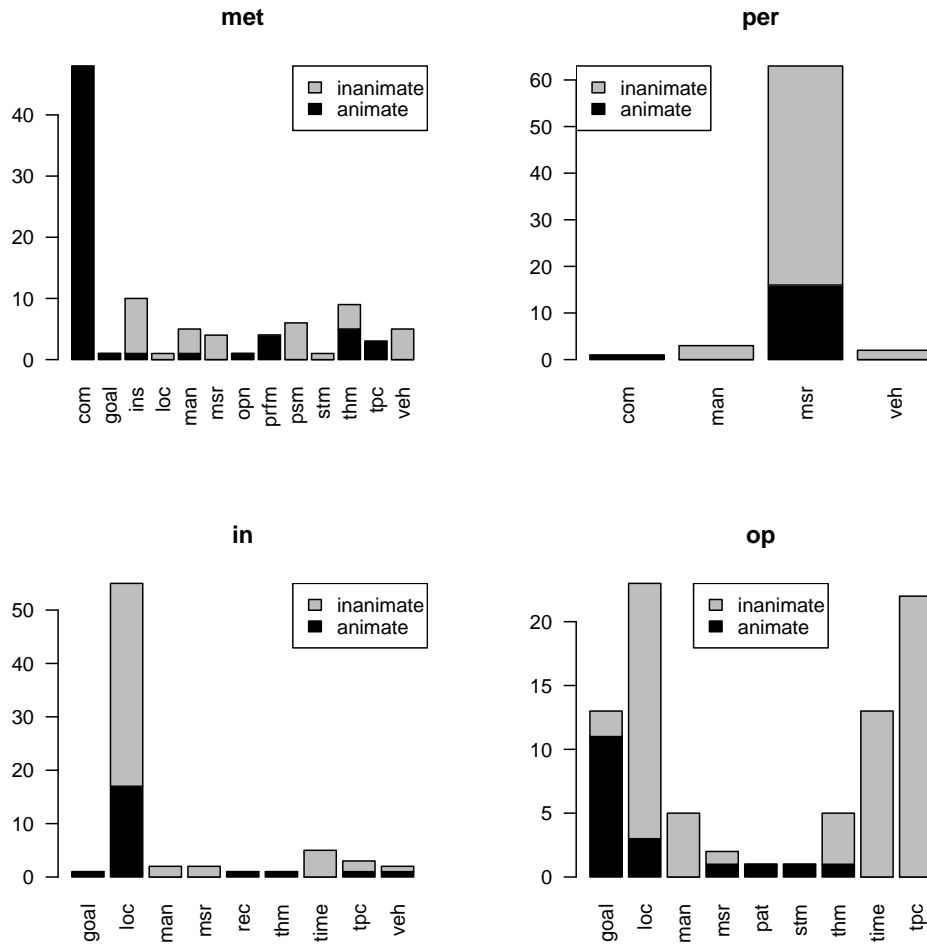


Figure 4: Semantic role distribution per P

*Per* ‘per’ almost exclusively expresses Measure. An example of *per* ‘per’ with an animate and inanimate complement is given in (14):

- (14) a. vijfhonderd achtentachtig **per** persoon ja  
 500 88 per person yes  
 ‘588 per person yes’ (Measure)
- b. zeg maar ruim honderd euro per uh **per** dag  
 say but more.than hundred euro per uh per day  
 ‘let’s say, more than 100 euro per day’ (Measure)

Also *in* ‘in’ has a very specific function in the expression of Location. An example of this preposition with an animate and inanimate complement is given in (15):

- (15) a. ze zocht naar iets anders **in** mij  
 she looked for something different in me  
 ‘she was looking for something different in me’ (Location)
- b. zij zijn **in** Brussel geweest om dien auto komen  
 they are in Brussels been to the/his car come  
 te halen  
 to pick.up  
 ‘they have been in Brussels to go collect that car’ (Location)

*Op* ‘on’ is again more broadly distributed, its four most important roles being Goal, Location, Topic, and Time. An example of *op* ‘on’ with an animate and inanimate complement is given in (16):

- (16) a. Numan gepast **op** Kluivert  
 Numan passed to Kluivert  
 ‘Numan to Kluivert’ (Goal)
- b. maar uh ja i\*a **op** zich vond ’k dat ook best leuk  
 but uh yes ## on itself found I that too quite nice  
 ‘but uh yes, I found that quite OK in itself’ (Topic)

Finally, in Figure 5 the distribution of semantic roles for the prepositions *over* ‘over’ and *tegen* ‘against’ is given. The latter adposition most often expresses Goal meaning, but is used for roles like Location, Opponent, and Theme as well. The adposition *over* ‘over’ almost exclusively expresses Theme.

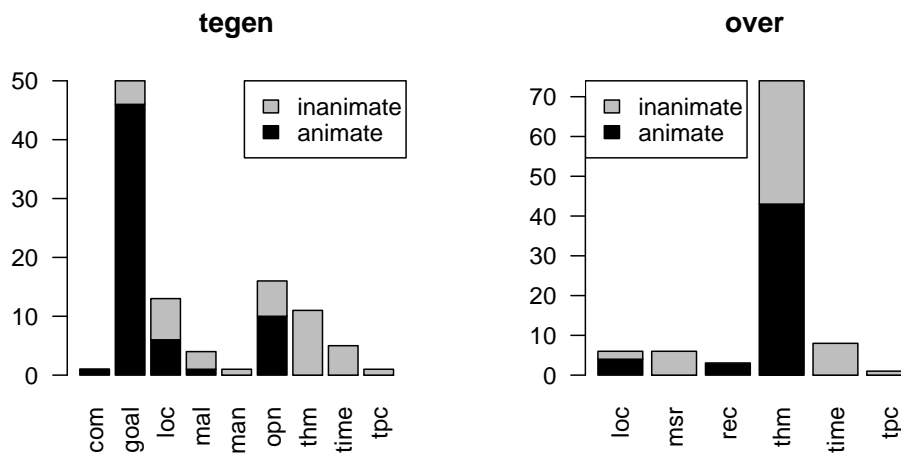


Figure 5: Semantic role distribution per P

An example of *over* ‘over’ with an animate and inanimate complement is given in (17); an example of *tegen* ‘against’ with an animate and inanimate complement is given in (18):

- (17) a. ik kan daar niet **over** oordelen  
 I can there not over judge  
 ‘I cannot judge that’ (Theme)
- b. en gisteren hebben er nog **over** u gesproken  
 and yesterday have there even over you spoken  
 ‘and yesterday (we) talked about you’ (Theme)
- (18) a. moet je **tegen** haar zeggen hoor  
 must you against her say PRT  
 ‘do say (that) to her, OK’ (Goal)
- b. kan niet **tegen** z’n verlies  
 cannot against his loss  
 ‘can’t stand losing’ (Theme)

In answer to our second question, we can say that most Dutch

prepositions are rather promiscuous. Instead of selecting a single or a strongly delimited subset of roles, most prepositions take a range of semantic roles, the semantic relation between them not being immediately clear.

### 4.2.3 Can semantic role interpretation be modeled as a function of animacy?

As discussed above (cf. sections 2 and 3), semantic role interpretation sometimes seems to be a function of animacy. This is also illustrated with the following example:

- (19) Jan repareert de tafel met... Marie/een hamer  
John repairs the table with... Mary/a hammer

In (19), the interpretation of the prepositional phrase is dependent on the animacy value of its complement. When combined with a human (Mary) it is interpreted as a Comitative; when combined with an inanimate, an Instrument interpretation arises.

In order to claim that semantic role interpretation can be modeled as a function of animacy a number of requirements should be fulfilled. First, Ps should combine with both animate and inanimate complements. Second, an adposition must be able to occur with more than one semantic role. Only when this is the case, the actual semantic role interpretation may follow from the animacy value of the prepositional complement. This of course only holds when the different semantic roles of a particular adposition are taken up exclusively by either animates or inanimates. In case the animate and inanimate role can

be semantically related to one another, as with the Comitative and Instrumental roles, we can treat semantic role interpretation as a function of animacy. Trivially, if both animate and inanimates are able to perform the same role of a particular preposition, the semantic role interpretation cannot be the result of the animacy value of its bearer.

In all the figures above, animate complements are represented in black, inanimates in grey. All adpositions in our data set allow for both animate and inanimate complements and occur with more than one semantic role, which means that the first two requirements stated above are fulfilled. Also, very often it is indeed the case that semantic roles are of one color only, suggesting that semantic role interpretation could be a function of animacy. But sometimes, animates and inanimates can perform the same role which disallows such a conclusion.<sup>9</sup> Examples of this latter pattern are given above for *over* ‘over’, *in* ‘in’, *per* ‘per’, and *bij* ‘at’.

A more sophisticated way to address the question whether semantic role interpretation is a function of animacy is to see whether semantic roles cluster on the basis of animacy. Cluster analyses translate similarity in form into vicinity in a tree. The more often two functions share a similar expression, the closer they are plotted in a cluster diagram.<sup>10</sup> In the cluster analysis given in Figure 6, we compared how often semantic roles were expressed by the 10 selected prepositions

---

<sup>9</sup>Of course, this is partially dependent on the definitions of the semantic roles under discussion which themselves may or may not include an animacy restriction, cf. Table 1 in the Appendix. When no such restriction is present, the semantic role will probably not be relevant for our question unless it is only taken up by either animates or inanimates. The interesting cases then are those in which an adposition occurs with two semantic roles with an animacy restriction (either by definition or in use) in which case we want to determine whether the two roles can be linked semantically on the basis of animacy.

<sup>10</sup>For a more detailed introduction to cluster analysis in linguistics, cf. Baayen (2008).

discussed in 4.2.2. We have excluded semantic roles that occurred less than five times. If different semantic roles show the same distribution over the prepositions, they will cluster together. For example, if the semantic roles Agent and Measure are expressed by the same set of prepositions they will cluster together. If they are expressed by a completely different set of prepositions, they will not cluster together. For each preposition, we determined the number of times it expresses a certain semantic role. If semantic roles come in pairs, we expect these pairs to be expressed by a completely different set of prepositions in the same way. That is, either both members of the pair are expressed by one preposition only or they are expressed by the same set of prepositions. We do not expect one member of an alleged pair to be expressed by completely different prepositions than the other member. In other words, if there exists such a thing as animacy-driven pairs of semantic roles, we expect these roles to be expressed in the same way by prepositions with the meaning of the PP being dependent on the animacy of the role bearer. In a cluster analysis as illustrated in Figure 6 this should then become clear.

As can be seen in Figure 6, the cluster analysis suggests that semantic role interpretation is not a function of animacy. Most semantic roles do not come in pairs. The only ones that do are Benefactive–Purpose (illustrated with *voor* ‘for, before’ in (20)), Experiencer–Possessor (illustrated with *aan* ‘to, at’ in (21)), and Possessum–Vehicle (illustrated with *met* ‘with’ in (22)).

- (20) a. artsen deden het werk dan ook **voor** niets  
doctors did the work then also for nothing



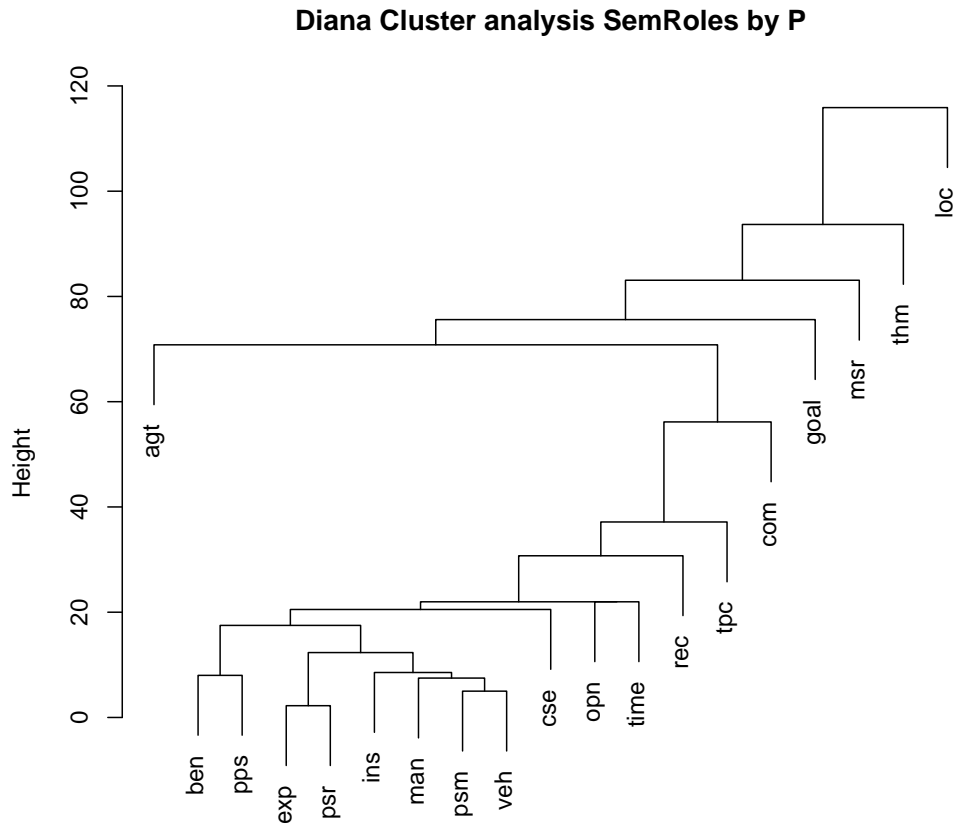


Figure 6: Cluster analysis of semantic roles by preposition (Diana)

‘doctors did the job for nothing’ (Purpose)

- b. hij zou dat **voor** mij ook mogen doen  
 he should that for me also may do  
 ‘he should do that for me too’ (Benefactive)

- (21) a. want anders zou ik die ’ns mee kunnen  
 because otherwise could I that once with could  
 nemen om **aan** ons jongens te laten zien  
 take to to our boys to let show  
 ‘because otherwise, I could take that with me once to

- show to our boys' (Experiencer)
- b. dan is 't woord **aan** de heer Schutte  
 than is the word to the gentleman Schutte  
 'than, the floor is to Mr Schutte' (Possessor)
- (22) a. tenminste ze lopen te veel **met** de bal  
 at.least they walk too much with the ball  
 'at least, they walk with the ball too much' (Possessum)
- b. en ze kwamen steeds maar niet **met** de auto  
 and they came always but not with the car  
 'And never once, they came by car' (Vehicle)

Of these three pairs, only the first can be analyzed as a pair of animacy driven meaning variants. The Purpose and Benefactive roles in this pair are almost exclusively expressed by *voor* 'for'. In our data, animates that combine with this preposition are almost always Benefactives, whereas inanimates are always Purpose. In contrast, the roles of the second pair, Experiencer and Possessor, are both animates and the roles in the third pair, Possessum and Vehicle, both inanimates. Therefore, these pairs are not distinguished by animacy.

An alternative way to analyze semantic-role patterns is by means of a semantic map. We constructed the semantic map in Figure 7 on the basis of the same set of adpositions mentioned above.<sup>11</sup> If two semantic roles are expressed by the same preposition in more than 15% of its occurrences, the roles are connected to each other on this map. Note that not all semantic roles are represented, as some of them did not meet this requirement (cf. Table 2 Appendix).

---

<sup>11</sup>As argued for by Malchukov and Narrog (2009), we show that it indeed is possible to draw a semantic map on the basis of polysemous constructions of one language.

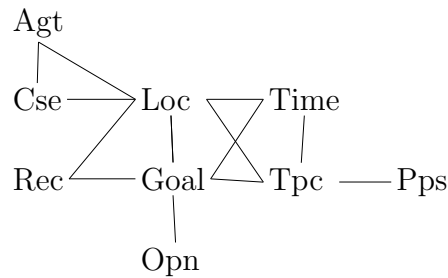


Figure 7: Semantic map of semantic roles

As the semantic map illustrates, spatial meaning (Loc and Goal) is at the core, from which nonspatial meaning is derived. Also, two semantic-role pairs that were not identified in the cluster analysis appear in this approach: Recipient–Goal/Location and Opponent–Goal. The first connection is due to the preposition *aan* ‘to’, the latter to *tegen* ‘against’. These pairs can be explained by an animacy-driven connection in the spirit of Aristar (1996, 1997). Recipients and Opponents could be seen as inanimate Goals or Locations. The difference between the cluster analysis and the semantic map approach is that in the cluster analysis semantic role pairs are only identified if they behave similarly in general, for all prepositions, whereas in the semantic map approach the connection is already established if two roles are expressed by a single preposition. This latter approach is more like that of Aristar (1997), as it is sensitive to the behavior of individual prepositions.

Although the semantic map is based on 10 frequent prepositions of Dutch only, most of its connections correspond to relations established in other grammaticalization or semantic map studies (Heine and Kuteva 2002; Haspelmath 2003; Malchukov and Narrog 2009). Due to

its limited scope however, many functions that are discerned elsewhere do not appear on our map.

## 5 Discussion

How should we understand the findings of the previous section? First, typing for animacy was shown to occur only with low-frequency Ps. This makes sense from the perspective that high-frequency elements in general are semantically less specialized than low-frequency ones (cf. for example Zipf 1965, see also Lestrade 2008 for a discussion directed at the spatial domain). More general meanings apply more often and therefore are used more often. Because of their frequency these items bleach semantically, take up more meanings and thereby become even more frequent and less specific. The less specific some construction is, the less typing it will exhibit; because of its general meaning, it can combine with any complement. This is precisely what we find in the different typing behavior of frequent and less frequent Dutch prepositions.

Secondly, we considered the question whether semantic role interpretation can be seen as a function of animacy. On the one hand, the meaning extension for type incongruent combinations that Aristar proposes for the case domain is not found in the adpositional domain in general. With the exception of the pair Benefactive–Purpose, semantic roles do not seem to cluster in animacy dependent pairs. On the other hand, when we examine the prepositions individually, at least some of the semantic roles each preposition expresses can be

analyzed as animacy-driven pairs. Therefore, meaning extensions for type incongruent pairs such as described by Aristar (1996, 1997) can be found in Dutch PPs, it is just that these meaning pairs are part of a broader distribution of semantic roles that a P may express. In sum, it may be possible in many languages to come up with nice examples of such pairs, but this does not mean that all meanings of some construction can be explained in this light.

The main difference between our findings and those of Aristar's can probably be attributed to a difference in methodology. Whereas Aristar's study is exclusively based on (constructed) data from grammatical sources we relied exclusively on usage data. We believe both approaches have their merits and come with their own problems. The first approach allows one to easily discern patterns and identify cross-linguistic generalizations at the risk of becoming anecdotal and obscuring variation. Our own quantitative approach, on the other hand, is rooted in a firm empirical basis which may make it difficult to extract clear patterns. In order to validate our findings and tentative analyses we probably have to enlarge our data set and include data from other languages as well. The kind of approach advocated here also brings along the question of what is the right way to analyze the data. We have presented here two different ways of data analysis (cluster analysis and semantic maps) which, although not completely incompatible, gave us slightly different results (see also the discussion in Croft and Poole 2008 and Narrog and Ito 2007). This shows that there are many remaining interesting issues which we have to leave for future research.

## 6 Conclusion

In this chapter, we considered the role of typing in the Dutch prepositional domain. We showed that it is only the less frequent Ps that show restrictions with respect to the animacy of their complements.

On the basis of the semantic role profiles of 10 frequent prepositions in Dutch we were able to analyze the relation between the semantic roles expressed by prepositions in the form of the semantic map in Figure 7 and the cluster analysis in Figure 6. Clearly, looking at only 10 prepositions in one language is not an empirically sound basis for far-reaching conclusions on conceptual relations between semantic roles. What it does show, however, is that semantic roles do not come in neatly ordered pairs. Meaning extensions for type incongruent pairs are part of a broader distribution of semantic roles that a P may express.

## References

- Aissen, J. (2003). Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory* 21(3), 435–483.
- Aristar, A. R. (1996). The relationship between dative and locative: Kuryłowicz’s argument from a typological perspective. *Diachronica* XIII(2), 207–224.
- Aristar, A. R. (1997). Marking and hierarchy: Types and the grammaticalization of case-markers. *Studies in Language* 21(2), 313–368.

- Baayen, H. (2008). *Analyzing Linguistic Data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Bossong, G. (1985). *Differentielle Objektmarkierung in den neuiranischen Sprachen*. Tübingen: Narr.
- Carletta, J. (1996). Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics* 22, 249–254.
- Croft, W. and K. T. Poole (2008). Inferring universals from grammatical variation: Multidimensional scaling for typological analysis. *Theoretical Linguistics* 34(1), 1–37.
- de Hoop, S. and A. Malchukov (2008). Case-marking strategies. *Linguistic Inquiry* 39 (4), 565–587.
- Haspelmath, M. (2003). The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In M. Tomasello (Ed.), *The New Psychology of Language*, Volume 2, pp. 211–243. Mahwah NJ: Lawrence Erlbaum.
- Heine, B. and T. Kuteva (2002). *World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press.
- Hopper, P. J. and S. A. Thompson (1980). Transitivity in grammar and discourse. *Language* 56(2), 251–299.
- Kulikov, L. (2006). Case systems in a diachronic perspective: A typological sketch. In L. Kulikov, A. Malchukov, and P. de Swart (Eds.), *Case, Valency and Transitivity*, pp. 23–47. Amsterdam/Philadelphia: John Benjamins Publishing Company.

- Lestrade, S. (2008). On the division of labor between spatial constructions. submitted.
- Levinson, S. C. and D. P. Wilkins (2006). Background to the study of the language of space. In S. C. Levinson and D. Wilkins (Eds.), *Grammars of Space. Explorations in Cognitive Diversity*. Cambridge University Press.
- Malchukov, A. (2008). Animacy and asymmetries in differential case marking. *Lingua* 118(2), 203–221.
- Malchukov, A. and H. Narrog (2009). Case polysemy. In A. Malchukov and A. Spencer (Eds.), *Handbook of Case*, pp. 518–534. Oxford: Oxford University Press.
- Narrog, H. and S. Ito (2007). Re-constructing semantic maps: the comitative-instrumental area. *Sprachtypologie und Universalienforschung* 60(4), 273–292.
- Nunberg, G. (1995). Transfers of meaning. *Journal of Semantics* 12(2), 109–132.
- Primus, B. (1999). *Cases and Thematic Roles: Ergative, Accusative and Active*. Tübingen: Niemeyer.
- Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge, MA: The MIT Press.
- Silverstein, M. (1976). Hierarchy of features and ergativity. In R. M. W. Dixon (Ed.), *Grammatical Categories in Australian Languages*, pp. 112–171. Atlantic Highlands: Humanities Press.



- de Swart, P. (2007). *Cross-linguistic Variation in Object Marking*. Ph.D. thesis, Radboud University Nijmegen. LOT Publications.
- Stolz, T., C. Stroh, and A. Urdze (2006). *On comitatives and related categories: a typological study with special focus on the languages of Europe*. Berlin: Mouton de Gruyter.
- Tsunoda, T. (1981). Split case-marking patterns in verb types and tense/aspect/mood. *Linguistics* 19(5-6), 389–438.
- Tsunoda, T. (1985). Remarks on transitivity. *Linguistics* 21(2), 385–396.
- van der Wouden, S., H. Hoekstra, M. Moortgat, B. Renmans, and I. Schuurman (2002). Syntactische annotatie voor het Corpus Gesproken Nederlands (cgn). *Nederlandse Taalkunde* 7(4), 335–352.
- Zaenen, A., J. Carletta, G. Garretson, J. Bresnan, A. Koontz-Garboden, T. Nikitina, M. C. O’Connor, and T. Wasow (2004). Animacy encoding in English: Why and how. In D. Byron and B. Webber (Eds.), *Proceedings of the ACL Workshop on Discourse Annotation*, pp. 118–125.
- Zipf, G. K. (1965). *Human behavior and the principle of least effort. An introduction to human ecology*. New York and London: Hafner publishing company.

## Appendix

Table 1: Semantic role definitions

---

AGENT (AGT)	animate instigator of an action
BENEFACTIVE (BEN)	person in favor of whom/to whose benefits an action is performed
CAUSE (CSE)	object/situation causing an action to occur
CAUSER (CSR)	person causing an action to occur
COMITATIVE (COM)	person together with whom an action is performed (+anim)
EXPERIENCER (EXP)	recipient of a cognitive stimulus
GOAL (GOAL)	entity to which an action is directed (focus on motion, not on arrival/possession) (+/-anim)
INSTRUMENT (INS)	object with which an action is performed
LOCATION (LOC)	place (point in space) at which an action is performed, or prominently ends
MALEFACTIVE (MAL)	person who is negatively affected by the action performed
MANNER (MAN)	way in which an action is performed
MEASURE (MSR)	extent to which an action is performed (has progressed)
OPPONENT (OPN)	actor with opposite direction, intention, force
PATIENT (PAT)	undergoer of an action, change of state
PERFORMER (PRFM)	person starring in a performance
POSSESSOR (PSR)	person to whom something belongs
POSSESSEE (PSE)	possessed object
RECIPIENT (REC)	entity receiving an object
SOURCE (SRC)	entity from which an action originates (not deliberately causal)
STIMULUS (STM)	cognitive impulse
VEHICLE (VEH)	object through which an action is performed, like instrument but more prominently moving
TIME (TIME)	point in time at which an action is performed
THEME (THM)	undergoes the action but does not change its state
TOPIC (TPC)	Restriction on discourse or reference domain, discourse topic

---

Table 2: Semantic role profile of 10 Dutch prepositions

P	agt	ben	com	cse	csr	exp	goal	ins	loc	mal	man	msr	opn	pat	path	pps	prfm	psm	psr	rec	src	stm	thm	time	tpc	veh	
aan	0	1	0	1	1	5	18	0	19	0	0	2	0	2	0	1	0	4	5	24	1	0	7	0	0	0	
bij	0	0	2	0	0	1	0	0	73	0	0	1	0	0	0	0	0	0	3	1	0	1	1	1	5	0	
door	52	0	0	15	1	0	0	3	14	0	0	0	0	0	1	0	0	0	0	0	0	1	1	4	0	0	
in	0	0	0	0	0	0	1	0	55	0	2	2	0	0	0	0	0	0	0	1	0	0	1	5	3	2	
met	0	0	48	0	0	0	1	10	1	0	5	4	1	0	0	0	4	6	0	0	0	1	9	0	3	5	
op	0	0	0	0	0	0	13	0	23	0	5	2	0	1	0	0	0	0	0	0	0	1	5	13	22	0	
over	0	0	0	0	0	0	0	0	6	0	0	6	0	0	0	0	0	0	0	3	0	0	74	8	1	0	
per	0	0	1	0	0	0	0	0	0	0	3	63	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
tegen	0	0	1	0	0	0	50	0	13	4	1	0	16	0	0	0	0	0	0	0	0	0	11	5	1	0	
voor	0	6	0	0	0	2	4	0	6	0	0	7	0	0	0	14	0	0	3	10	0	0	6	4	25	0	