Chapter 7

Variation within Support Verb Constructions

7.1 Introduction

In the course of study, theoretical and descriptive linguists have conceded that fixed expressions – especially the prototypical type i.e. idioms – allow limited variation and adjective modification. This concession requires that the linguistic description of fixed expressions in lexicalist grammars allow internal syntactic structure. An appropriate linguistic description of e.g. idiomatic support verb constructions (svcs) that show modification inside one or more arguments needs to have internal syntactic structure. De aap uit de mouw komen ‘the truth is revealed’ and veeg uit de pan krijgen ‘get a lashing/licking’ allow adjectival modifiers ((1), (2)) or a post-nominal PP modifier (2) without loss of figurative meaning; presence of modification suggests that the constituents (aap, een veeg, uit de pan) are only partially lexicalized.

(1) Na de val van die andere grote ideologie (van het communisme) zijn er in Oost-Europa weer gevaarlijke apen uit de mouw gekomen.
‘After the fall of the other big ideology (the communism), dangerous truths were revealed in Eastern Europe.’
Chapter 7. Variation within Support Verb Constructions

(2) Vorige week kreeg de politie nog een flinke veeg uit de pan

Last week took the police yet a big blow out the pan

van de Universiteit Leiden en de Anne Frank Stichting,

of the University Leiden and the Anne Frank Foundation.

‘Last week, the police got a sharp recrimination from the University of Leiden and the Anne Frank Foundation.’

Besides adjectival modification, svcs exhibit variation that affects determiners and, also noun morphology. Uit het/zijn dak gaan ‘go completely wild’ allows a definite or a possessive determiner with the noun without loss of the figurative meaning:

(3) Als het aan hem lag, kreeg Nederland nog een dag ‘om uit

if it on him lies, gets Nederland yet a day ‘in order to out

z’n/het dak te gaan’.

his/the roof to go

‘If he could decide, the Netherlands still gets another day to go wild.’

Onder de tafel houden ‘to hide, keep secret’ is found with a definite determiner or without any determiner:

(4) Ook de suggestie dat de WHO het onderzoek onder (de) tafel

also the suggestion that the WHO the research under (the) table

wilde houden, was bezijden de waarheid.

would like to keep, was far from the truth

‘In addition, the suggestion that the WHO would like to keep secret the research was obscuring the truth.’

The noun aap typically occurs in the singular in the context of the idiom
de aap uit de mouw komen, however, a plural noun is present in (1) above.

In fact, corpus evidence shows that determiners inside idiomatic and figurative expressions are not always fixed (frozen). In the context of the idiom

iemand het hemd van het lijf vragen ‘want to know everything about s.o.’,

the determiner de ‘the’ accompanies the noun hemd ‘shirt’ in 50% of the idiom instances (5). Outside the idiom, hemd ‘shirt’ shows the determiner

het ‘the’; note that nouns used in a definite context take either de or het,

both realizations of the definite determiner.

(5) Hij vraagt mij de hemd van het lijf.

he asks me the shirt of the body

‘He wants to know everything about me.’
The concession mentioned above opened a long (ongoing) debate about the effect of the insertion of modification in the semantics of *fixed expressions*. The object of the debate is whether internal modification affects only the meaning of the constituent or the meaning of the whole predicate. Two views are portrayed in the literature on idiomatic expressions. The first view maintains that internal modification is only syntactic and it affects the meaning of the whole predicate (Fernando and Flavell, 1981; Verstraten, 1989; Schenk, 1994; Abeillé, 1995; Everaert and Kuiper, 1996). Other instances of adjective modification (in which the meaning of the adjective cannot modify the whole predicate) are 'word play', not deserving any treatment in the grammar since they are considered as marginal. The second view maintains that with a few exceptions, idiom constituents that allow internal modification, have meaning and the modifier alters such meaning (Ernst, 1982; Gazdar et al., 1985; Nunberg et al., 1994; Riehemann, 2001). Corpus-based studies established the existence of various types of fixed expressions differing in their semantic structure and also in their internal syntactic flexibility (Moon, 1998; Sailer, 2000; Riehemann, 2001; Sag et al., 2001). More specifically, Riehemann’s findings support Nunberg et al.’s claims (1994) that syntactic flexibility (modification, passivization, topicalization, etc.) is more likely to be found among semantically decomposable idioms.\(^1\)\(^2\)

The first view is mainly built on human judgements. It is well known that human judgements on linguistic data often disagree; further, the disagreements are more noticeable when assessing metaphorical and idiomatic phenomena. Therefore, we are inclined to accept those views resulting from the systematic exploration of the behavior of idiomatic expressions and other types of fixed expressions in a large number of corpus examples.

### 7.1.1 Mining linguistic descriptions from corpora

Motivated by the need to describe various types of fixed expressions in a lexicalist grammar, we need to know if lexemes and constituent phrases in support verb constructions (svcs) always exhibit the same form. The investigation described in this chapter attempts to establish what variation and modification is observed in support verb constructions found in large corpora. By variation we mean, changes affecting determiners or the noun’s morphology inside a partially lexicalized constituent and, by modification, the insertion of adjectives or other noun modifiers.

(Non-)Productive morphology, specier variation and adjectival modific-\(^1\) Decomposable idioms have been defined and described in Chapter 2, section 2.2.3.  
\(^2\) The empirical basis of this observation was VERB NP English idioms.
variation vary from expression to expression. Our goal here is to develop a method that extracts evidence of variation and modification at a large scale, preferably with little manual intervention. The main goal of this chapter is to assess to what extent a data-driven method allows us establish in which contexts variation and modification are possible in Dutch support verb constructions (svcs).

The goal of the corpus-based method is to find evidence of specifier variation and adjectival modification. No evidence of modification in the corpus does not mean that internal modification is banned, especially concerning infrequent expressions. Our second goal is to determine which predictions can be made about internal modification; more precisely, we explore if there exists some correlation between specifier variation and adjectival modification. As a derivative issue, we also examine whether specifier variation and internal modification pinpoint internal semantic structure within the support verb constructions we studied.

7.1.2 Overview

We designed a corpus-based method to extract relevant information from corpora and investigated whether the method is useful to determine modifiability and variation within Dutch support verb constructions that require a pp argument. Section 7.2 describes the types of variation and modification found inside Dutch prepositional support verb constructions (svcs). Section 7.3 describes the design, application and evaluation of the corpus-based method.

On the basis of the extracted evidence we pursue predictions about which type of expressions are likely to allow modification or some variation. Inspired by Fellbaum’s (1993) work in English, we explore in section 7.4 to what extent determiner changes observed inside the NPs (in svcs) pinpoint NPs likely to allow some modification. The evidence extracted from the corpus is the empirical basis for assessing to what extent determiner changes (morphosyntactic cue) correlate with a potential for modification, thus suggesting the existence of internal semantic structure (decomposability).

7.2 Modification and variation types

First, an assumption must be introduced, namely, what we understand as a base form and what we understand as a variant of a support verb construction. A fixed expression is represented as a combination of the minimum required lexemes that make up the expression. In prepositional svcs, the
minimum required lexemes include the argument PP and the verb; each expression is represented as a triple consisting of the sequence [PREPOSITION NOUN VERB].

(Thus, determiners are left out). The base form includes the triple and it may include other lexicalized arguments; the base form of the expression *voet bij stuk houden* includes the noun *voet*, the PP *bij stuk* and the verb *houden* (6-a). Any instance of an SVC whose NP within the PP argument differs from the NOUN lexeme is considered a variant.

(6) a. **VOET BIJ STUK HOUDEN**
   b. De verzekeraars *hielden* echter *voet bij stuk*.
      the insurers kept really foot by piece
      ‘The insurance companies really stuck to their guns.’

In expressions that allow no morphosyntactic variation or modification within the required arguments, tense inflection is still possible. The base form is the infinitive verb form. The base form given in (6-a) corresponds to *voet bij stuk houden* ‘to stick to one’s guns’, expression that only allows tense inflection (6-b). The expression *iemand aan zijn woord houden* ‘hold s.o. to his word/promise’ has as base form (7-a), but *woord* allows various possessive determiners; (7-b) shows one of its variants.

(7) a. **IEMAND, ACC AAN WOORD HOUDEN**
   b. We *houden* Melkert aan zijn *woord*.
      we keep Melkert on his word
      ‘We keep Melkert to his word.’

Next, we enumerate the types of variation and internal modification that the method ought to retrieve. At the lexeme level, we seek evidence of productive inflectional (plural and diminutive morphemes) and derivational morphology ((noun noun) compounding, other prefixes). At the phrase level, we seek variability in specifiers (determiners, quantifiers, etc.), adjectives and participials.

### 7.2.1 Evidence we seek

Nominal lexemes may show use of diminutive (in some cases), for example, both forms *oog* ‘eye’ and *oogje* ‘little eye’ co-exist in the expression *een oog(je)* *in het zeil houden* ‘keep a look-out’ (8).

---

3In chapters 5 and 6 abstractions of support verb constructions are represented by the triple VERB PREP NOUN. Here, we use a slightly different ordering inside the triple – PREP NOUN VERB – echoing (to some extent) the convention followed in Dutch lexicography.
Chapter 7. Variation within Support Verb Constructions

(8) a. We zouden al snel te weinig politie hebben om we might already fast too little police have to overal een oog in het zeil te houden. everywhere an eye in the sail to keep

‘We might already have too little police to keep a look-out everywhere.’

b. Politieagenten houden een oogje in het zeil.

police agents keep a little eye in the sail

‘Police agents keep a look-out.’

We also explore singular and plural morpheme alternation in nouns. Deur ‘door’ in iemand/iets buiten de deur houden ‘keep sth/s.o. away’ allows this alternation:

(9) a. ...om genetisch gemanipuleerd voedsel buiten de EU-deur te ...to genetic modified food uit the EU-door to houden keep

‘...in order to keep genetically modified food outside the EU.’

b. ...om deze ongewenste gast buiten de deuren van het ...to this unwanted guest out the doors of the ziekenhuis te houden hospital to keep

‘...in order to keep this unwanted guest away from the hospital’

Some nominals exhibit evidence of derivational morphology, for example, instances of an acronym accompanied with the head noun deur (9-a) or a genitive noun modifier of bezit ‘possession’ in iets in bezit houden ‘to own sth’:

(10) ... hoopte Boeve na de rust de bal langer in Waalwijks ... hoped Boeve after the break the ball longer in Waalwijks bezit te kunnen houden possession to can keep

‘...After the break, Boeve hoped to be able to keep the ball in possession of the Waalwijk’s team a longer time.’

With regard to phrase internal modification, we study specifiers, adjectives and participial phrases (present or past participle). Among specifiers: zero determiner (11-a), definite (9-b), indefinite (11-b), reciprocals (12-a), possessives (7-b), demonstratives (12-b) and quantifiers (12-c).
7.2. Modification and variation types

(11) a. Evengoed moest Van Eeden zelf regelmatig financieel
    Nevertheless must Van Eeden self regularly financial
    bijspringen om de kolonie op poten te houden.
    step in in-order-to the colony on paws to hold
    ‘Nevertheless, Van Eeden had to often step in financially in order
    to keep the colony running.’

b. Blessures houden het team in een wurgende greep
    Lesions hold the team in a strangling grasp
    ‘Lesions hold the team in a strangling grasp’

(12) a. In die kringen houden ze allemaal de hand boven
    In these circles hold they all the hand above
    elkaar hoofd en …
    each other’s head and …
    ‘In these circles they all protect each other and (continue with
    their daily routine).’

b. Er is geen IQ van bovenmenselijke proporties voor nodig
    there is no IQ of superhuman proportions for needed
    om tot deze slotsom te komen.
    to until this conclusion to come
    ‘No superhuman dimension is needed to arrive at this conclu-
    sion.’

c. We willen alles in één hand houden.
    we want everything in one hand hold
    ‘We want to have control over everything.’

A last type of variation falls outside modification because it involves whole
NPs that are realized by reflexives. Reflexives may instantiate either open
slots, e.g. in zich op de vlakte houden ‘not commit oneself’ (13-a), or an NP
within complement PPs (e.g. iets voor zich houden ‘keep sth to oneself.’)
(13-b).

(13) a. De branche-organisaties van de tussenpersonen, de NVA en
    the branch-organizations of the intermediaries de NVA and
    NBvA, houden zich wat op de vlakte.
    NBvA, hold themselves a bit on the plain
    ‘The agents organizations, de NVA and NBvA, do not commit
    themselves (a little).’

b. Boeren zwijgen, houden hun problemen voor zich.
    farmers are silent, hold their problems for themselves
    ‘Farmers are silent, keeping their problems to themselves.’
Among adjectival modification, we explore presence of pre-nominal adjectives (14-a) (*iemand uit de wind houden* ‘give shelter’), past participles (14-b) (*in opstand komen* ‘revolt’) and other intervening material (14-c) (*iets binnen de perken houden* ‘limit; contain’).

(14) a. Zij moet *uit de partijpolitieke* wind worden gehouden  
She must out the party political wind was held  
‘She must be kept away from the political polemic.’

b. toen *de Indiaanse bevolking in de provincie in gewapende* opstand kwam . . .  
‘when the Indian people in the province in armed uprising came . . .’

c. Meadow Wingfield *held in Groningen Smokin Pay Roll nog*  
Meadow Wingfield held in Groningen Smokin Pay Roll yet  
redelijk makkelijk *binnen de voor hem plezierige* reasonably possible within the for him pleasant  
limits  
‘In Groningen, Meadow Wingfield kept Smokin Pay Roll within a (for him) pleasant bounds.’

We focus on the enumerated modification and variation types. In addition, some expressions allow relative clauses and PP post-nominal modifiers. Whereas relative clauses are uncommon, PP post-nominal modifiers are not.  
(15-b) shows an example of PP modification in the expression *aan het hoofd staan* ‘be in charge of’. We ignore these two types of modification because we attempt to extract the evidence from an automatically annotated corpus and with automatic means. As already discussed in section 6.5.2, disambiguating a syntactic attachment site, e.g. a PP–attachment site, is one of the hardest problems for present-day parsing technology. Needless to say, Alpino also has difficulties with this problem.

(15) a. Minister Alders kwam tot de bittere *slotsom* dat we het al twee jaar niet goed hebben gedaan.  
Minister Alders came to the bitter conclusion that we it already two years not well have done  
‘Minister Alders reached the bitter conclusion that we have not done very well in the last two years.’

b. Piëch *stond jarenlang aan het hoofd van Audi.*  
Piëch stood years long on the head of Audi
‘Piëch was in charge of Audi for several years.’

In this work, we will not investigate syntactic flexibility at the sentence level, that is, we do not assess whether the expressions which are the object of study (svc) allow passive, topicalization, clefting, coordination, etc. Chapter 6 partly explored the syntactic flexibility of these expressions.

7.3 A corpus-based method

The method used to retrieve evidence of morphological variation and internal modification in support verb constructions resembles the method for applying linguistic diagnostics, described earlier in section 6.3.2.

7.3.1 Settings

For each support verb construction (svc), all sentences containing the three component lexemes (automatically extracted triple) found in the Twente Nieuws Corpus (TWNc) were collected in a subcorpus. For example, for the expression *uit zijn dak gaan* ‘go crazy’, all sentences that include the preposition *uit* ‘out’, the noun *dak* ‘roof’ and the verb *gaan* ‘go’ or one of its inflectional variants are collected in a subcorpus.

The Alpino parser was used to automatically annotate all subcorpora. *dt_search*, a treebank query tool, is used to extract evidence from the fully parsed subcorpora. The reader is referred to the detailed description of the resources and tools provided earlier in 6.3.

Given a list of support verb constructions, we assume that the subcategorization frame information for each svc was automatically inferred or retrieved from a dictionary. As illustration, the verb in *uit zijn dak gaan* requires a subject NP and a PP. The PP has *uit* as its head, and its object NP is headed by *dak*.\(^4\)

With access to automatically parsed data, subcategorization frames and a standard search query language such as *dt_search*, we can extract all instances of an svc that satisfy rather specific morphosyntactic features and head-complement dependencies; these requirements – expressed as *dt_search* queries – are applied to XML-encoded syntactic dependency trees. The dependency structure tree shown in figure 7.1 is the parsed output of the example in (16).

---

\(^4\)The grammar informing the parser treats svcs as regular phrases built out of individual lexemes or phrases selected by the verb, thus with an underlying lexicalist representation of svcs.
Het publiek ging uit zijn dak.

Figure 7.1: This syntactic dependency tree corresponds to the parsed sentence in (16).
7.3. A corpus-based method

(16) Het publiek ging uit zijn dak.
the audience went out his roof
‘The audience went crazy.’

Among the information contained in the parsed trees, we use: (i) categorical information (phrasal (e.g. **np, pp**) and lexical category (e.g. **det, noun**)), (ii) syntactic information (grammatical function or dependency relation (direct object **obj1**, head **hd**, modifier **mod**, determiner **det**)) and (iii) lexical information (lexemes and word forms). Dependency nodes are crucial in stating daughter–ancestor relations between constituents and sub-constituents in an SVC. In addition, the XML encoded tree provides positional information, clausal type, sentence type, etc. that we neglect.

### 7.3.2 The extraction process

A **dt_search** query applied on the corresponding parsed subcorpus searches for all SVC instances. Two types of queries are needed: **narrow search** and **wide search** queries. Narrow search queries seek instances of a head-dependent relation between a **verb** and a sister **pp**, given necessary lexical restrictions as input. Sometimes narrow searches in addition to the **pp** and its head **verb**, specify the head of an **np**; for example, **[vinger np aan pols pp houden]** corresponding to the expression **een vinger aan de pols houden** ‘(fig) keep a finger on the pulse’. Wide searches state that the **pp** is located under a clausal node whose head is **verb**. These are needed because the parser may make a mistake and attach the sought **pp** to a previous noun, thus in the annotated data the **pp** and **verb** do not satisfy a head-dependent relation. Finally, the vaguest search states that a given **pp** needs to occur within the same sentence as the **verb**. This type of search is used in case the other two types fail to retrieve any evidence. Next, we study local constraints expressed in the search queries.

The query shown in figure 7.2 extracts all possible realizations of the determiner inside the **pp** argument. The query in figure 7.3 seeks internal adjectival modification.

We examine the query in 7.2 closely; the query applies to the example **iets in het bezit houden** ‘to own sth.’ **dt_search** retrieves original corpus sentences and frequency counts about the matched hits. An **np** node that has a determiner phrase daughter (**./node[@cat="detp"]**) and no adjectival phrase daughter (**not(./node[@cat="ap"])**) is sought; the matched node instantiates an object (**../node[@rel="obj1"]**) of a specific preposition (**in**) and the resulting **pp** needs to occur under a clause headed by a specific **verb**. Lexical restrictions are specified by giving values to the **word** and/or **root**
Chapter 7. Variation within Support Verb Constructions

\[
\text{dt\_search}\]

'//node[@cat="np"] and 
   ./node[@cat="dtp"] and not(./node[@cat="ap"]) and 
   ./node[@rel="hd" and @root="bezit"] and 
   ../node[@rel="obj1"] and 
   ./node[@rel="hd" and @word="in"] and 
   (../../../node[@rel="hd" and @root="houden"] or 
   ../../node[@rel="hd" and @root="houden"] or 
   ../node[@rel="hd" and @root="houden"]))}'

houd.inbezit/*.xml

Figure 7.2: Query to extract possible specifiers inside the NP node under the PP in the expression \textit{iets in bezit houden}.

\[
\text{dt\_search}\]

'//node[@cat="np"] and 
   ./node[@cat="ap"] and 
   ./node[@rel="hd" and (@word="gedachten" or @root="gedachte")] 
   and ../node[@rel="obj1"] and ../node[@rel="hd" and @word="op"] 
   and (../../../node[@rel="hd" and @root="breng"] or 
   ../../node[@rel="hd" and @root="breng"] or 
   ../node[@rel="hd" and @root="breng"]) ] }'

breng.opgedachten/*.xml

Figure 7.3: \texttt{dt\_search} query to extract possible adjectives in the NP under the PP in the expression \textit{op gedachten brengen}.
head attributes; in the example, we provide the preposition (word="in"), the
noun’s base form (root="bezit") and the verb’s base form (root="houd").

Queries need to capture deeply embedded SVCs. A verbal complement
embedded under several modal or auxiliary verbs is rather common. To
allow uncertainty about the location of the PP argument node with respect
to its head verb, disjunctive constraints are introduced in the queries (figures
7.2 and 7.3).

7.3.3 The evidence retrieved

A search query retrieves each SVC realization that satisfies the query re-
quirements, as well as its frequency in the subcorpora. Figure 7.4 shows the
observed specifiers in iets in bezit houden ‘to own sth’:

41 zero
1 haar
2 het
1 hun
1 je
3 mijn
1 ons
1 privé-bezit
1 Waalwijks
11 zijn

Figure 7.4: Observed specifiers (zero and possessive determiner) and com-
ounding inside the PP argument in the SVC iets in bezit houden ‘to own sth’.

Figure 7.5 gives an excerpt from the observed adjectival modification
in iemand op gedachten brengen ‘give s.o. the idea’. Op andere gedachten
brengen ‘change s.o.’s idea’ is the most frequent realization with 634 out
of a total of 682 occurrences. Some parsing errors surface in the evidence
extracted; see line marked with ‘*’ in figure 7.5.

Table 7.1 summarizes the queries output for a few SVCs.\textsuperscript{5} Abbreviations
refer to the following: SPECifiers, MODification and total FREQUENCY of the
SVC. Modification may be limited to one adjective (e.g. eigen ‘own’) or less
restrictive in the form of various adjectives. If only one adjective is found,
we give the adjective in italics, else we use the adj label.

\textsuperscript{5}Morphological productivity is not included because no evidence of diminutive, sin-
gular/plural alternation or compounding was found for these expressions.
Chapter 7. Variation within Support Verb Constructions

1 aangename gedachten
1 amoureuze gedachten

634 andere gedachten
1 andere, meer pragmatische gedachten
1 andere politieke gedachten
1 andere, redelijke gedachten
1 beeldende gedachten
1 bepaalde gedachten
2 betere gedachten
* 1 diplomatieke wijze tot andere gedachten
1 duivelse gedachten
1 heel andere gedachten over de doodstraf
1 hitsige gedachten
1 hogere gedachten
1 kritische gedachten
1 meer poëtische gedachten

Figure 7.5: Excerpt from the observed adjectival modification in the svc op gedachten brengen.

<table>
<thead>
<tr>
<th>Expression</th>
<th>SPEC</th>
<th>MOD</th>
<th>FREQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>ter sprake komen</td>
<td>z(1202)</td>
<td></td>
<td>1202</td>
</tr>
<tr>
<td>in kaart brengen</td>
<td>z(2838)</td>
<td></td>
<td>2838</td>
</tr>
<tr>
<td>in bewaring stellen</td>
<td>z(13)</td>
<td>verzekerde</td>
<td>(3)</td>
</tr>
<tr>
<td>in spanning houden</td>
<td>z(51), d(2)</td>
<td>adj(4)</td>
<td>53</td>
</tr>
<tr>
<td>op gedachten brengen</td>
<td>z(681), d(1)</td>
<td>adj(671)</td>
<td>682</td>
</tr>
<tr>
<td>uit de slaap houden</td>
<td>d(55), p(84)</td>
<td></td>
<td>139</td>
</tr>
<tr>
<td>tegen het licht houden</td>
<td>d(648), i(2)</td>
<td>adj(3)</td>
<td>650</td>
</tr>
<tr>
<td>van stuk brengen</td>
<td>p(247), d(2)</td>
<td>adj(2)</td>
<td>249</td>
</tr>
</tbody>
</table>

Table 7.1: Observed realizations in specifiers (SPEC) and adjectival modification (MOD). Numbers provide the frequency counts. Possible determiner types are: zero, definite, indefinite, possessive, demonstrative as well as quantifiers.
7.3. A corpus-based method

The method extracts evidence of morphological productivity, variation of specifiers and adjectival modification, i.e. positive and negative evidence. A description of the positive evidence follows. We investigated 107 Dutch SVCs: 94 expressions that require a PP argument among which some show an NP\textsubscript{acc} open slot; lexical restrictions affect the verb and the PP argument; in addition, 13 other expressions are made up of a (partially) lexicalized NP and a PP argument.

Most SVCs fall in one of three groups: (a) totally fixed, (b) semi-fixed and (c) flexible. **Fixed** SVCs show no variation and no modification in the lexicalized NP (if present) and PP constituent(s). 42% of the SVCs studied are fixed. E.g. *voet bij stuk houden*. **Semi-fixed** SVCs show partially lexicalized constituent(s) (20.5% of the studied SVCs). Rarely, a singular noun appears in plural e.g. in *puntje bij paaltje komen* ‘to come to the crunch (fig.)’. Variation affects the lexeme’s morphology and/or the specifiers slot. Expressions whose ‘fixed’ argument requires a reflexive also belong to this group such as *iets voor zich houden* ‘keep sth. to o.s.’. **Flexible** SVCs allow adjectival modification (37.5% of the studied SVCs). The data is rather varied. There are SVCs that show: (i) non-productive morphology and no specifier variation but they show a limited number of adjectives e.g. *door het (politieke) leven gaan* and, (ii) specifier variation (some show compounding) and limited adjectival variation (e.g. *in (aanmerkelijk rustiger) vaarwater zitten*). Border-line cases exhibit no morphological productivity and either definite/possessive determiner alternation or no specifier variation; modification involves a unique adjective (e.g. *in (verzekerde) bewaring stellen* ‘put into care/custody’).

Negative evidence (noise) typically includes sentences where the **verb** and the **PP** occur within the same clause but not in the SVC context (in its literal use). Often, the PP is an adjunct or a complement of another verb. The reason for this noise can be attributed to the uncertainty in the search queries or errors in the annotated data. This introduces a shortcoming since human inspection needs to ensure that the evidence corresponds to the SVC use.

In order to assess the usability of the method, next we discuss its main strengths and limitations.

7.3.4 Evaluation

Examining the evidence retrieved by the corpus-based method just described, a researcher quickly forms an impression about which expressions are totally fixed and which expressions allow some variation and/or modification. One also has direct access to the realizations of the variable slots, the SVC fre-
No additional resources were needed since the svc list (automatically extracted), the automatically parsed subcorpora and the treebank query tool (dt_search) had already been used to apply the linguistic diagnostics (described in section 6.3.2). The fact that we explore occurrences of an svc in any syntactic structure (main or subordinate sentence, imperative, questions, etc.) means that no additional constraints about linear precedence or clause type need to be stated. This facilitates formulating the dt_search queries.

The corpus-based method lacked information about the subcategorization requirements of the svcs. This information was manually entered for each expression. Once we have a list of PREPOSITION NOUN VERB triples, methods described in the literature on automatic acquisition of subcategorization information might be successful in finding out the remaining svc syntactic requirements. This is an open issue for future research, but a starting point would be the approach by Briscoe and Carroll (1997).

Fully parsed data is needed, as well as a tool that identifies clauses in which a certain pp and a given verb co-occur. Furthermore, knowledge of the grammar and the lexicon used by the parser is helpful. In particular, one needs to know whether some svcs or idiosyncratic phrases are already annotated in the lexicon as lexicalized phrases. Briefly, the search queries need to anticipate the peculiarities and decisions of the parser.

The success of the search queries is dependent on parsing accuracy. Sometimes extracted evidence shows the specific pp we seek but as argument of another verb. Parsing accuracy introduces another shortcoming: evidence of relative clauses and pp post-nominal modifiers cannot be automatically retrieved. Because of structural ambiguity, attachment decisions are still a hard parsing problem. This led us to ignore these two types of modification in our research. Thus, further research is needed and we cannot provide a full answer to our central question with this method.

The method encounters a more general problem. Specifier changes or insertion of modification may destroy the svc reading. The queries could extract evidence that looks like a variant of the svc base form; in practice, the svc interpretation does not apply. For example, in most of the instances of the expression de hand boven het hoofd houden ‘to protect s.o.’ (lit. the
hand above the head hold), *hoofd* is preceded by the definite determiner; there are also a few instances with a reciprocal *elkaar* ‘each other’s’ and some instances with possessive determiners. The query results suggest that all three specifiers are possible; however, the instances with possessive determiners are literal uses. Occasionally, a PREPOSITION NOUN VERB triple clusters homonymous expressions. A search that specifies the triple base form in hand houden could match any of the following expressions *iets in één hand houden* ‘to be the boss’, *het heft in handen houden* ‘remain in control’, *de touwtjes in handen houden*, *iets in handen houden* ‘have control over sth’ or *iets in de handen houden* ‘to hold sth in one’s hands (lit.)’. In some cases, the subcategorization requirements of the SVC differ from those of the regular literal phrase (e.g. *iemand van de straat houden* ‘keep s.o. off the street’ vs. *van de straat houden* ‘to love the street’). One needs to assess which use of the triple is present in the extracted sentences before variation and modification evidence can be attested.

The corpus-based method cannot be fully automated; that is, extraction of variation and modification evidence can be done fully automatically. Instead, the evidence retrieved needs to be manually inspected. This brings up a last limitation of the method. At least one instance of each variation and modification type requires manual inspection. The researcher needs to establish whether the SVC interpretation is present or only a literal reading applies. Yet, all the tools we used facilitated this process and they provide plenty of relevant linguistic empirical evidence.

A last limitation affecting most corpus-based research is that having found no evidence of variation and modification does not mean that it is not possible in SVCs. A lot more data is needed. Yet, this is (as far as our knowledge reaches) the first time one applies full parsing for corpus-based linguistic exploration in Dutch.

### 7.3.5 Summary

The corpus-based method extracts evidence of variation and modification within Dutch support verb constructions. The method is sufficiently efficient in extracting proof of morphological productivity, specifier variation and adjectival modification inside SVCs, but at least one instance of each type of variation needs to be manually assessed to determine whether the SVC interpretation is present. The evidence retrieved allows us to establish required syntactic structure and furthermore, a preliminary classification of SVCs has been proposed. The method described a first attempt to use automatically parsed corpora for corpus linguistic purposes in Dutch.
Chapter 7. Variation within Support Verb Constructions

7.4 Assessing modifiability

The positive evidence retrieved with the corpus-based method shows us which svc's allow modification. Evidence corresponds to instances of modification that are present in our corpus and that are captured by the method. However, it may happen that svc's (of which we lack evidence) in principle allow modification but either that the method failed to capture this or, that no evidence exists in the corpus. Our goal is to attempt to deduce in which contexts modification ought to be possible. The reasons behind this are the lack of evidence just mentioned and the need to establish the internal syntactic structure and to establish whether syntactic flexibility correlates with internal semantic structure in the svc's explored.

Fellbaum (1993) investigated whether idioms could be shown to have meaningful internal structure. Fellbaum observed that the nature of the determiner (form and use) points at the semantic status of the noun phrases in many idioms. Among the English v np idioms studied by Fellbaum, determiner substitution is found only in compositional idioms where it is 'an indicator of the noun's referential status'. Instead, the determiner is invariable in non-compositional idioms. Fellbaum also claimed that the particular form of the determiner in some non-compositional idioms cannot be explained according to the rules for literal language that appear to operate in compositional idioms. Fellbaum used grammaticality judgments and introspection to determine which idioms allow determiner alteration and the insertion of specificity-inducing adjectives in order to determine the sort of np referent.

Because native speakers differ a lot in their grammaticality judgements, especially when the data involves idiomatic and figurative phenomena, we take a different approach from Fellbaum's and turn to corpora exploration to infer what the actual use of determiners and modification is.

The attested evidence of determiner alternation and adjectival modification in 107 Dutch support verb constructions suggests that determiner alteration may be a useful morphosyntactic cue to predict adjectival modification. Table 7.2 depicts the data distribution.

The chance of observing modification when no determiner changes are attested is much smaller than the chance of having no modification when no determiner changes have been attested (ratio 1:5.3); alternatively, absence of modification is more likely when no determiner changes are observed than when determiner changes have been attested (ratio 2.5:1).

To establish whether determiner changes correlate with modification, the relative risk compares the probability of observing determiner changes and modification with the probability of observing determiner changes and no
7.4. Assessing modifiability

<table>
<thead>
<tr>
<th></th>
<th>modification</th>
<th>no modification</th>
<th>marginals</th>
</tr>
</thead>
<tbody>
<tr>
<td>determiner changes</td>
<td>31</td>
<td>19</td>
<td>50</td>
</tr>
<tr>
<td>no determiner changes</td>
<td>9</td>
<td>48</td>
<td>57</td>
</tr>
<tr>
<td>marginals</td>
<td>40</td>
<td>67</td>
<td>107 (total)</td>
</tr>
</tbody>
</table>

Table 7.2: Comparing the proportion of svcs that allow modification when taking into account determiner changes.

modifiability. For our sample, the relative risk\(^7\) is 1.89 and it indicates that the chance of observing modification is higher when determiner changes are observed. If our svc sample is representative, 1.89 instances of modification are expected for every instance of no modification in cases where determiner changes are exemplified.

Nevertheless there are some problems. Our conclusions on the small sample cannot be guaranteed to be representative enough. Further, determiner changes are also observed among svcs for which we found no evidence of adjectival modification. In these svcs, it may be possible that modification is ruled out but it may also be possible that our corpus does not include pertinent evidence. Bearing these two problems in mind, our hypothesis is that ‘if certain determiner changes are observed in an np inside a support verb construction it is possible that the np shows modification’. If this hypothesis is true, certain determiner changes can be said to predict modification in svcs. Alternatively, if we do not observe certain determiner changes, the np shows no internal modification.

Given the attested evidence, we study patterns in the determiner changes. We assess to what extent determiner changes correlate with the presence of modification. We also identify which changes are informative (thus, possibly good predictors) and which changes are uninformative.

Eventually, we aim at establishing whether specifier variation and adjectival modification mark out ‘denoting noun phrases’ inside required arguments. The intuition is that in decomposable expressions constituent chunks carry meaning, thus they are more likely to show determiner substitution and internal adjectival modification. By contrast, constituents in non-decomposable expressions are non-denoting so that one expects to find little or no determiner alternation and modification.

---

\(^7\)The relative risk is a descriptive measure that gives us the ratio of the ‘success’ probabilities for two groups in a 2X2 contingency table. Applied to Table 7.2, relative risk is computed as the ratio between (i) the probability of jointly observing determiner changes and modification \((p_1 = 31/107)\) and (ii) the probability of observing determiner changes but no modification \((p_2 = 19/107)\).
7.4.1 Trends in extracted evidence

Table 7.2 showed the distribution of determiner changes and adjectival modification in the SVCs. A characterization of the observed patterns within the four subgroups (corresponding to the four cells in table 7.2) follows. Three tables (7.3, 7.4 and 7.5) summarize the patterns of morphological productivity, specifier alternation and adjectival modification found in attested evidence.

- **Totally fixed expressions** SVCs that show no morphological productivity, specifier variation and adjectival modification total 42%. Table 7.3 depicts the observed patterns. The specifier is either a zero or a definite determiner. One expression shows the quantifier ‘én’ ‘one’.

- **Semi-fixed with variable determiner** expressions show determiner changes but no modification. 20.5% of the SVCs exhibit determiner changes with a clear preference for one determiner (zero, definite or possessive) and no attested adjectival modification. A few show singular/plural alternation. Table 7.4 lists the patterns found. Possible determiner alternations are (b1) zero or definite determiner, e.g. *in* (het) *leven houden* ‘keep in existence’ and (b2) definite or possessive determiner, e.g. *aan het/poss hart gaan* ‘to deeply affect s.o.’. In spite of the determiner alternation, the meaning remains the same.

- **Semi-fixed with variable adjective** expressions show no determiner changes but they do show modification e.g. *iemand van zijn stuk brengen* ‘upset s.o.’. Table 7.5 depicts the patterns found.

- **Flexible expressions** show determiner changes and modification. A first subgroup shows a clear preference for one determiner type but rare instances with another determiner were observed (e.g. *in gijzel-ing houden* ‘to take as hostage’). A second subgroup allows multiple specifier types and some expressions show morphological productivity (e.g. *iets in bezit houden* ‘to own’).

The first three patterns (in table 7.5) correspond to fuzzy cases; in two cases, a unique optional adjective has been attested (de touwtjes in (eigen) handen houden, orde op (eigen) zaken stellen); in the other case, specifiers may vary and a specific adjective is required (in de/zijn kouwe kleren gaan zitten).
Table 7.3: Morphological productivity, specifier variation (SPEC) and adjectival modification (MOD). Patterns found among totally fixed SVCs. Abbreviations used are: compounding, number morpheme, diminutive), specifiers alternation (namely, dets determiners and quantifiers) and adjectival modification. Among determiner types we assessed zero, definite, indefinite, possessive and demonstrative determiners. Adjectival modification is either limited to 1 adjective or various adjectives. Columns record which realizations of a specific type have been attested (`+') or not (`-'). Rows depict a pattern in the attested data and provide an example SVC. `+*' marks a rare determiner realization found in the corpus. AGR means that a possessive determiner fills the specifier slot.
### Table 7.4: Morphological productivity, specifier variation (SPEC) and adjectival modification (MOD). Patterns found among semi-fixed svcS. See caption in Table 7.3 for explanation of abbreviations.

<table>
<thead>
<tr>
<th>Morphology</th>
<th>SPEC</th>
<th>MOD</th>
<th>example</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>comp num dim</td>
<td>dets</td>
<td>quan</td>
</tr>
<tr>
<td>-</td>
<td>sg    -</td>
<td>+</td>
<td>+*</td>
</tr>
<tr>
<td>-</td>
<td>sg/pl -</td>
<td>+</td>
<td>+*</td>
</tr>
<tr>
<td>-</td>
<td>sg    -</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>-</td>
<td>sg    -</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>-</td>
<td>sg/pl -</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>-</td>
<td>sg    -</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>-</td>
<td>sg    -</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Morphology</td>
<td>SPEC</td>
<td>MOD</td>
<td>example</td>
</tr>
<tr>
<td>------------</td>
<td>------</td>
<td>-----</td>
<td>---------</td>
</tr>
<tr>
<td></td>
<td>comp</td>
<td>num</td>
<td>dim</td>
</tr>
<tr>
<td></td>
<td>sg</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>pl</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>pl</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>pl</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>sg</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>sg</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>sg</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>sg</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>sg</td>
<td>-</td>
<td>+*</td>
</tr>
<tr>
<td></td>
<td>pl</td>
<td>-</td>
<td>+*</td>
</tr>
<tr>
<td></td>
<td>sg</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>sg</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>sg</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>sg</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>sg</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>sg</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>sg</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>sg</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>sg</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>sg</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>sg</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>sg</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>sg</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>sg</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>sg</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>sg</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 7.5: Morphological productivity, specifier variation (SPEC) and adjectival modification (MOD). Patterns found among flexible svcs. See caption in Table 7.3 for explanation of abbreviations.
7.4.2 Informative determiner changes

*Semi-fixed with variable determiner* expressions question the scope of our hypothesis given that determiner changes are attested but, we found no presence of adjectival modification. *Semi-fixed with variable adjective* expressions do not contradict our hypothesis, instead they confirm the wisdom of the hypothesis not be formulated bi-directional. If an SVC allows modification, this does not imply that certain determiner changes are possible. *Flexible expressions* do not contradict our hypothesis.

However, there are some determiner changes that are common to *semi-fixed with variable determiner* expressions and *flexible* expressions. Such changes may not be construed as a morphosyntactic cue that predicts modification. Determiner alternations exclusively found in SVCS that show adjectival modification are: (d1) zero, definite and possessive determiner and (d2), either indefinite or demonstrative or a quantifier and one or more other determiner types.

Among *semi-fixed with variable determiner* expressions showing alternation (b1) there are syntactically regular and also asyntactic NPs. In syntactically regular NPs (e.g. *in (de) gedachte(n) houden*) the noun is either a plural countable or an abstract noun and therefore, such nouns may occur with a zero or a definite determiner. Asyntactic determinerless NPs are part of a PP denoting location (e.g. *onder (de) tafel houden, iemand van de straat houden*); there are rare instances with a definite determiner. Though we speak of ‘asyntactic’ NPs, a descriptive grammar observes that in locative expressions (*plaatsaanduidende verbindingen*) such as *boven water* ‘above the water’, *in huis* ‘in the house’, *onder water* ‘under water’, *onder tafel* ‘under the table’, the determiner is left out (Haeseryn et al., 1997, p. 198). A few expressions that typically exhibit a definite determiner sometimes show no determiner (*aan de slag gaan* ‘get going’). NPs that show definite–possessive determiner alternation (change (b2)) include nouns that in their literal meaning denote body parts and institutionalized entities (*aan het hart gaan* ‘deeply affect s.o.’, *uit zijn dak gaan* ‘go crazy’). These expressions seem to be active metaphors. In idiomatic expressions, the definite–possessive determiner alternation in NPs denoting limbs, body parts or clothing has been explained as being present in some (regional) dialects of Dutch by Haeseryn et al. (1997). The scope of this determiner alternation is puzzling given that some NPs denoting body parts inside the SVCS do not allow it (e.g. *de hand boven het hoofd houden* ‘to protect s.o.’). The zero and definite determiner alternation suggests a tendency to leave out the determiner in the SVCS, thus asyntactic NPs are not only locative expressions but also other expressions denoting a metaphorical location like *onder de duim houden* ‘keep under s.o.’s thumb.
7.4. Assessing modifiability

7.4.3 Underlying semantic structure

While we investigated a method to study some morphosyntactic characteristics of support verb constructions that may play a role in variation potential, we also seek to contribute with our findings to the theory of modification in fixed expressions and idiomatic phenomena.

Our findings on SVC data do not provide further support for the view defended by Nunberg et al. (1994) that syntactic flexibility (internal at least) is a sign of semantic decomposability. The fact that determiner alternation and insertion of modification is possible in some SVCs indeed suggests that the underlying semantic structure of such SVCs must be different.

Two informants considered possible paraphrases of the flexible SVCs in Table 7.5. If a paraphrase exhibited a semantic structure that is homomorphic with the semantic structure of the SVC, then, the informant labelled the expression as semantic decomposable (e.g., *iets tegen het licht houden* ‘inspect closely’). If no paraphrase with a homomorphic semantic structure was found the expression is semantic non-decomposable (e.g., *iets in de gaten houden* ‘keep an eye on sth.’). Most expressions are compositional predicates or metaphorical ones. A few expressions that show no variation and no adjectival modification qualify as semantic decomposable (*in één hand houden*); expressions that show limited determiner changes and no modification (semi-fixed with variable determiner expressions) are either fully compositional or semantic decomposable; in the latter case the PP has a metaphorical interpretation (*iemand van de straat houden, iets in leven houden* ‘keep in existence’). Finally, a few expressions that allow specifier variation and adjectival modification qualify as semantic non-decomposable, e.g., *iets tegen het licht houden and iemand aan het lijntje houden* ‘keep s.o. dangling’.

The attested evidence of determiners variation and internal modification is not sufficient however to establish a parallelism between the presence of variation and modification and semantic compositionality or semantic decomposability. On the basis of our data, no clear-cut distribution that considers morphosyntactic and semantic behavior emerges. At a higher level, this implies that specifier variation and adjectival modification in SVCs are not sufficient cues for automatically classifying expressions into semantic classes. Furthermore, if with these morphosyntactic cues we cannot predict semantic structure, we also cannot predict syntactic flexibility at the sentence level; that is, we cannot predict which SVCs allow passivization, clefting, topicalization, etc. With improvements in parsing accuracy, an open alternative is to estimate syntactic flexibility directly on the basis of large syntactically
annotated corpora.

With this work we aimed at identifying morphosyntactic cues inside SVCs that could indicate semantic structure. In recent years, there have been various attempts to infer semantic decomposability or semantic compositionality using machine learning techniques.

Baldwin et al. (2003) aimed at demarcating simple decomposable multi-word expressions (MWES) from idiosyncratically decomposable and non-decomposable ones. Their method looks at the semantic similarity between an MWE and its constituent words. Their hypothesis is that where the similarity between the MWE and its constituent words is high, the MWE is probably simply decomposable. Baldwin et al. (2003) concentrate on noun-noun compounds and particle verbs (verb-particle) bigrams. They adopt latent semantic analysis, a method that measures the similarity between the MWE and its composite words. As advantages of this method they mention that LSA makes no assumptions about the lexical or syntactic composition of the inputs, therefore it constitutes a ‘fully construction and language insensitive method of looking at decomposability’. To evaluate this approach, the LSA method is compared to the judgments extracted from measuring the similarity between a MWE and its constituent words in WordNet. In this way, they compare a corpus-based approach to decomposability and a dictionary-based approach. The correlation between the LSA similarity method and WordNet semantic distances is very low for noun-noun compounds and a bit higher for verb-particle bigrams; in any case, the correlation between LSA and WordNet similarities is very low. Two reasons for the disappointing results are that LSA similarities are more erratic among high-frequency MWES and classificational inconsistencies in WordNet.

McCarthy et al. (2003) describe a model to identify the degree of semantic opacity or non-compositionality in English phrasal verbs. Three methods are compared: (i) measuring compositionality with an automatically acquired thesaurus, (ii) with statistics commonly used in multi-word lexeme extraction from corpora and (iii) human compositionality judgements. The first method builds on Lin’s (1998) research on constructing a thesaurus automatically in order to find compositional and non-compositional multi-word expressions. The intuition behind it is that a semantically opaque verb-particle combination shows little similarity with the simplex verb; alternatively, a transparent or compositional verb-particle combination shows a significant semantic similarity with the simplex verb. Neither automatic method achieves good results.

Detecting semantic decomposability in idioms (and lexicalized phrases) requires two steps: first, to come up with a meaning or paraphrase with the same semantic structure as the idiom, and second, decide whether the
idiomatic words (noun) can be thought of as carrying part of that meaning (Riehemann, 2001, p. 72). Although this exercise might be easy for some idiomatic expressions, in most lexicalized phrases it is difficult to find a paraphrase with similar semantic structure. Speakers often disagree and do not see the internal semantic structure in the same way. Bearing this in mind, it is not surprising that computational models designed to detect semantic decomposability of multi-word lexemes produced disappointing results so far, as the task seems to be very difficult.

7.5 Summary

We described a semi-automatic method to extract evidence of morphosyntactic variation and modification found within svcs in a large corpus. The method is effective in retrieving phrase internal variation and modification, however the method needs to be improved before it can exploit the presence or absence of relative clauses and PP postmodifiers. The extracted evidence can be used to improve an existing formalization and implementation of svcs. Such improvements ought to contribute to increasing parsing accuracy. We also established that specifier variation, in particular, certain determiner changes correlate, albeit less than perfectly, with adjectival modification. Finally, we observed that specifier variation and adjectival modification in the svcs we studied do not delimit semantically decomposable expressions. On the basis of our data, no clear-cut distribution that considers morphosyntactic and semantic behavior emerges. At a higher level, this implies that specifier variation and adjectival modification in SVCS are not sufficient cues to automatically classify expressions into semantic classes.
Chapter 7. Variation within Support Verb Constructions