Data-driven identification of fixed expressions and their modifiability
Villada Moirón, María Begoña

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2005

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Chapter 2

On fixed expressions

‘In investigating the properties of lexical items one is looking at the inmates of a prison (the lexicon) who have in common only the fact of their lawlessness (Di Sciullo and Williams, 1987).’

2.1 Introduction

Fixed expressions depart in their behavior from regular and productive phrases. Bearing in mind Di Sciullo and William’s (1987) analogy, this chapter aims at identifying the irregularities observed in fixed expressions. We mainly study the lexical, semantic and syntactic properties of these expressions that distinguish them from regular and productive phrases.

The goal of the characterization is mainly descriptive, namely to identify features of fixed expressions that ought to be captured or approximated by the data-driven methods explored in the chapters to follow, and indirectly, to find features of fixed expressions that need to be specified in a computational grammar. Some features apply to any type of fixed expressions and some others describe specific aspects of fixed expressions with a verbal phrase structure.

The characterization that follows is partly based on descriptive work on fixed and idiomatic expressions by Everaert (1993), Everaert and Kuiper (1996) and Moon (1998) as well as formal proposals of their treatment and representation in various grammar formalisms (Katz (1973); Gazdar et al. (1985); Erbach (1992); Krenn and Erbach (1994); Sailer (2000); Riehemann
Chapter 2. On fixed expressions

(2001); Sag et al. (2001) among others). Detailed descriptions of idioms have contributed a great deal of insight and structure to the contents in this chapter (Fernando and Flavell (1981), Verstraten (1989), Schenk (1994), Nunberg et al. (1994) and Abeillé (1995)). The works cited studied fixed expressions in languages such as English, French, Dutch and German. The evidence supporting the description is taken either from the literature consulted or from the output of the research presented in later chapters. The Dutch data is extracted from the Twente Nieuws Corpus (TWNC) (Ordelman, 2002), unless explicitly cited otherwise. The Twente Nieuws Corpus (TWNC) is made up of newspaper text and some television news reports.

2.2 Semantic properties of fixed expressions

When a language user encounters a fixed expression, (s)he may notice its non-literalness. The literal meaning of the component words when they are used independently plays no role in the meaning of the fixed expression. It is the defining context that allows speakers to identify the adequate interpretation (literal or idiomatic) of homonymous expressions such as *iets in één hand houden* ‘done under one’s person/institution’s responsibility’; it is the context that triggers the idiomatic or the literal meaning. The lexical context triggers only a literal interpretation in (1) but a figurative interpretation in (2):

(1) Nog in zijn schooljaren bleek het houten racket voor hem te zwaar om in één hand te houden. ‘When still in his school years, the wooden racket seemed to be very heavy for him to hold in one hand.’

(2) Voorlichting, onderzoek en controle moet je in één hand houden. ‘One must keep information, investigation and control together.’

A mismatch between the literal sense of the individual lexemes, the linguistic context and the discourse marks out the idiosyncratic expression and triggers the idiomatic interpretation. In some expressions, the lexical restrictions and syntactic constituency are sufficient to determine the intended interpretation. The syntactic dependents of *zitten* ‘sit’ (typically intransitive) as well as their lexical realization differ significantly in the context of the
figurative expression *iemand iets in het bloed zitten* ‘to be in s.o.’s blood’:

(3) Maar schaatsen zit ons *in het bloed*, zegt Karlstad.
    but skating sits us in the blood, says Karlstad
    ‘But skating is in our blood, says Karlstad.’

Nevertheless, some fixed expressions have a (partially) literal meaning. In the examples in (4), *zoen* ‘kiss’ refers to the kissing event, still denoting a literal meaning; *school* and *hospital* show the literal meaning they denote in other contexts, even while the syntax of the expressions in (4-b) unexpectedly requires no determiner for the singular count nouns.

(4) a. *een zoen geven* ‘to give a kiss’
    b. *by car, at school, in hospital*

Words that denote literal meaning in fixed expressions could in principle have the status of ordinary individual lexemes. However, their morpho-syntactic behavior in the context of the fixed expression may show irregularity.

### 2.2.1 Denoting and non-denoting words

Words in fixed expressions may be denoting or non-denoting. A ‘denoting’ word refers to a concrete entity or an abstract concept. A ‘non-denoting’ word bears no meaning assignment. Among denoting words some are polysemous, thus showing a literal and a non-literal meaning. The individual words *laatste* (‘last’), *hand* (‘hand’) and *leggen* (‘lay’, respectively) outside the expression *de laatste hand aan iets leggen* show a literal meaning. In the context of the fixed expression they show a non-literal meaning, thus they qualify as polysemous (5-a). Verbal lexemes sometimes show semantic bleaching or a defective meaning. So-called *light verbs* found in *support verb constructions* (SVC) have a literal meaning outside the SVC; that meaning is ‘bleached’ in its SVC use. For example, *geven* ‘to give’ in the expression *een zoen geven* ‘to kiss’. Denoting words may only have a literal meaning, the one present in the fixed expression also. The literal meaning of *sluiten* ‘close’ remains present in the context of *een vergadering sluiten* in (5-b).

(5) a. *de laatste hand aan iets leggen* (the last hand on something lay)
    ‘to put the finishing touches on something’
    b. *een vergadering sluiten* (a meeting close)
‘to adjourn a meeting’

Some words (in fixed expressions) qualify as non-denoting. Although the words have a literal meaning outside the expression, no meaning can be attributed to them in the context of the fixed expression (e.g. *plaat* ‘plate’ and *poetsen* ‘clean’ in *de plaat poetsen* (6-a)). In other cases, non-denoting words are *nonce words* i.e. words that exclusively occur in fixed expressions. *Petto* (6-b) is a nonce word in Dutch.

(6) a. *de plaat poetsen*
   the plate polish
   ‘to depart unnoticed’

b. *in petto hebben/houden*
   in store have/keep
   ‘to reserve’

Everaert (1993) also mentions *loer* in the expression *een loer draaien* ‘play a nasty trick on s.o.’. Native speakers maintain that the word *loer* has no existence on its own nowadays. In English, the word *dint* only occurs in *by dint of, amok* (a loanword from Malay) in *run amok* ‘in a violently raging manner’ and *lieu* (French loanword) in *in (the) lieu of* ‘instead of’; these expressions make up a subtype of *cranberry words* (Moon, 1998; Richter and Sailer, 2003).

Two tests help identify denoting phrases inside fixed expressions; the first one checks if the lexeme(s) exhibit the same meaning outside the fixed expression and the second, checks if a pronoun somewhere else in the discourse can refer to an NP part of a fixed expression. Sailer (2000) (among others) used these tests to find out irregularities in the semantics of verbal idiomatic expressions. Riehemann (2001) provides evidence of pronominal reference found in ‘decomposable’2 idioms in English. Experiments described in chapter 6 confirm that pronominalization also distinguishes certain arguments in support verb constructions from complements in regular verb phrases.

### 2.2.2 Opacity and conventionality

Non-literalness introduces unpredictability. The literal meaning denoted by the lexemes in a fixed expression does not always motivate the peculiar meaning of the whole expression. Whereas the idiomatic meaning of the expres-

---

1 An equivalent example in English is *shoot the breeze* ‘to talk’.
2 Section 2.2.3 gives a definition of this term.
2.2. Semantic properties of fixed expressions

Sions in (7) is totally unmotivated, one could guess the meaning of to put the cart before the horse in both English and Dutch (8) with a bit of effort.

(7) a. pull someone’s leg ‘to tease’, kick the bucket ‘to die’
    b. de plaat poetsen ‘to depart unnoticed’, een rib uit het lijf zijn ‘to cost an arm and a leg’ (Dutch)

(8) a. het paard achter de wagen spannen (Dutch)
    the horse after the cart set
    ‘to put/set the cart before the horse’ (to reverse the natural or proper order)

Other fixed expressions have rather transparent meaning:

(9) iets in bezit hebben (Dutch)
    something in possession have
    ‘to own something’

(10) iets in gedachten houden (Dutch)
    something in thoughts keep
    ‘to bear in mind’

Semantic unpredictability increases in parallel with semantic opacity, a more common concept in the literature (Fernando and Flavell, 1981; Nuneberg et al., 1994). A non-transparent meaning is the manifestation of opacity. Thus, the more opaque the meaning of a fixed expression is, the more unpredictable it is. Fernando and Flavell (1981) compared the semantic unpredictability observed in fixed expressions to a continuous meaning scale that ranges from fully transparent to fully opaque meaning. ‘Fully transparent’ applies to expressions that exclusively denote a literal meaning therefore, completely predictable (e.g. to watch a soccer match, een vergadering sluiten).

‘Fully opaque’ applies to unmotivated expressions e.g. the idioms shoot the breeze and de plaat poetsen. A large number of expressions exhibit either semi-opaque meaning or semi-transparent meaning, without there existing a clear-cut division between categories (e.g. in je maag zitten met ‘be troubled with’ and iemand van de straat houden ‘keep s.o. off the street’).

So far, we have introduced three interrelated properties of certain fixed expressions: non-literalness, unpredictability and opacity. Next, we study the semantic opacity a bit further, trying to determine what makes some expressions opaque.

By combining the meaning of a verb with the meaning of its syntactic dependents compositionally, language users can regularly derive the meaning of any literal expression. These facts are captured in the so-called principle of
compositionality of meaning attributed to Frege. This principle explains why hearers can understand strings or sentences that they have never encountered before. Janssen (1997) quotes ‘Frege’s principle of compositionality’ as follows:

‘The meaning of a compound expression is a function of the meanings of its parts and of the syntactic rule by which they are combined.’

The cumulative combination of the meaning of the individual lexemes renders the meaning of the expression *Pat bought these cherries*. However, there are exceptional cases of expressions whose meaning cannot be derived in a regular way as it is derived in productive phrases.

The meaning of some idiomatic expressions is semi- or fully opaque due to the fact that the expression consists of a word that does not occur or hardly occurs in other expressions in the language (*nonce word*). This is the case with the loanword *petto*. *In petto* occurs in the Dutch expressions *iets in petto hebben* ‘to have sth in reserve’ and the less frequent variant *iets in petto houden* ‘to keep sth in reserve’. A speaker either knows the meaning of the fixed expression or she does not given the presence of nonce words.

In other expressions, semantic opacity is present because the expression (if literally understood) violates the semantic truth conditions given our knowledge of the world (Fernando and Flavell, 1981). This is the case in the expressions in (11) and (12).

(11) a. *to rain cats and dogs*
b. *jump down someone’s throat* ‘to reprimand or contradict one fiercely’
c. *iemand naar de keel vliegen* ‘go for s.o.’s throat’

(12) a. *perder los papeles* (Spanish)
    lose the papers
    ‘to lose control of o.s.’
b. *hablar hasta por los codos* (Spanish)
    to talk even through the elbows
    ‘to talk one’s ear off’

In other cases, truth conditions violations are more subtle:

(13) a. *iemand van de straat houden*
    someone off the street keep
    ‘keep someone off the street (at work; busy)’
2.2. Semantic properties of fixed expressions

b. iets bij de hand hebben
   something by the hand have
   ‘to have something at hand’

Some fixed expressions are homonymous showing a literal and an idiomatic meaning. Sometimes, the literal and the idiomatic interpretation are not easily distinguished unless one has access to a salient context. Examples of homonymous expressions in English are kick the bucket ‘die’, saw logs ‘snore’ and spill the beans ‘reveal the secret’ and in Dutch, the expression iemand aan de tand voelen ‘to grill someone (by asking lots of questions)’.

The literal and the idiomatic meanings are in principle possible if no extra contextual information is supplied in (14); nevertheless, the idiomatic meaning is more plausible:

(14) De dokter voelde vader en zoon aan de tand.
   The doctor felt father and son on the tooth
   ‘The doctor grilled father and son (fig.)’

Due to the presence of nonce words, potential truth condition violations or homonymous expressions, the derivation of the meaning of some fixed expressions has to differ from that of expressions that only have a literal interpretation. The interesting data involves idiomatic expressions and other metaphorical expressions.

Semantic opacity is caused by a more fundamental characteristic of fixed expressions: conventionality. Assuming that the meaning of a fixed expression is unknown, such meaning cannot be derived from the meaning of its component lexemes. The idiomatic meaning of Pat spilled the beans is not the result of combining the meaning of the words in the sentence in a cumulative way. Provided spill and beans occur within the appropriate syntactic context, the idiomatic meaning ‘to reveal the secret’ is ‘triggered’, and not otherwise. The individual lexemes in idiomatic expressions exhibit another meaning that cannot be derived from the literal one using only linguistic rules. From a synchronic perspective, the expression receives a meaning that could have been arbitrarily assigned and that we cannot otherwise account for. Therefore, we can only accept that the idiomatic meaning of word combinations like to rain cats and dogs or spill the beans is conventionalized. To capture this arbitrariness, there is a wide-spread view that the meaning of an

\footnote{According to Nunberg et al. (1994), conventionality is an obligatory dimension of idiomaticity. We adopt Nunberg et al’s definition of conventionality: ‘a relation among a linguistic regularity, a situation of use, and a population that has implicitly agreed to conform to that regularity in that situation (…) (Nunberg et al., 1994, p. 492)’.}
idiomatic expression is assigned to the whole phrase (Katz, 1973; Fernando and Flavell, 1981; Schenk, 1994; Abeillé, 1995; Riehemann, 2001). Other approaches derive the meaning of less opaque idiomatic expressions compositionally (Nunberg et al., 1994; Sailer, 2000) or by introducing additional analytical devices (Gazdar et al., 1985).

The derivation of meaning in idiomatic expressions cannot be explained by the principle of compositionality. Instead, meaning is assigned arbitrarily to the expression as a whole. These claims imply that the idiom chunks do not carry meaning i.e. the constituents inside the idiomatic expression do not denote. However, empirical evidence of incomplete idioms, idiom variants and insertion of non-idiomatic lexemes seems to contradict this.

Counterevidence are occurrences of incomplete idioms (Fernando and Flavell, 1981; Moon, 1998; Riehemann, 2001). As an example, the verb scrape is left out but the meaning of the complete idiom scrape the bottom of the barrel (‘to obtain something (by necessity) from a poor source’) remains in (15):

(15) That suggestion came from the bottom of the barrel (from Riehemann (2001, p. 64))

Based on this, an idiom chunk receives the idiomatic interpretation without the need to appear in the same context as the rest of the idiom. Riehemann explains that this is possible for parts of idioms with strong metaphorical denotation and high frequency. In our view, this suggests that the bottom of the barrel has semantic significance on its own and is in the course of becoming institutionalized and acquiring the status of a (phrasal) lexeme. Another example is a red herring which in its idiomatic use used to be part of the idiom draw a red herring across the trail (Fernando and Flavell, 1981).

A Dutch example is the proverb Hoge bomen vangen veel wind ‘(fig.) a great tree attracts the wind’. The NP hoge bomen gave title to a television program that shows interviews with ‘intellectuals and authorities’. The NP may be in the process of becoming lexicalized given the existence of examples in which the NP has the same idiomatic meaning as in the proverb:

(16) Er kwamen ”hoge bomen” in het nieuws van deze prille regering in een door etnische verschillen nauwelijks regeerbaar land (…) (‘The news talked about the officials in this premature government in a (due to ethnic differences) hardly governable country (…)’).

---

The second piece of evidence is the existence of idiom variants such as *hit the hay/sack* ‘go to bed’, *blow one’s stack/cool/top* ‘lose one’s composure; go crazy’. Within each pair or triple the same lexemic sense is present even though there is variation of the form. Furthermore, cases of insertion of adjectival modifiers, relative clauses, quantifiers, possessive determiners, and other instances of variation pinpoint traces of meaning inside some idiom components (Nunberg et al., 1994).

This evidence suggests that some parts of idioms and opaque fixed expressions are assigned an interpretation, so that the expression must have some internal semantic structure. But can we reconcile these statements with the claims that the meaning of an idiomatic expression is assigned to the expression as a whole and that its meaning derivation cannot be explained by the principle of compositionality? We discuss these two issues in turn.

### 2.2.3 Decomposability

Once the meaning of the whole idiom is known, it is natural and intuitive to attempt to deconstruct the meaning seeking the motivation behind and seeking the analyzability of the expression. A process of meaning deconstruction begins by which the meaning of the whole expression is distributed across its parts (Nunberg et al., 1994; Abeillé, 1995). Sag et al. (2001) introduced the notion of *semantic decomposability* to refer to ‘the process of semantic deconstruction that starts off with the idiom and associates particular components of the overall meaning with its parts’.

According to the decomposability hypothesis, the overall idiomatic meaning of a semantically decomposable expression such as *spill the beans* can be split across its component lexemes; thus, the paraphrasis ‘to reveal secret information’ is mapped onto the component lexemes, such that *spill* is interpreted as the action *reveal* and *the beans* is interpreted as *secret information*:

\[
\text{(17) } \begin{align*}
\text{a. } \text{spill the beans} & \text{ reveal}(X, \text{ a secret}) \\
\text{b. } \text{shoot the breeze} & \text{ chat.idly}(X)
\end{align*}
\]

In contrast, the idiomatic meaning of *shoot the breeze* cannot be distributed across the component lexemes in the same way. The reason is that not all idiomatic expressions show isomorphic semantic structures in their literal and idiomatic meaning. English examples are *saw logs* ‘snore’, *kick the bucket* ‘die’; in Dutch *de plaat poetsen* ‘depart unnoticed’. In these expressions, the idiomatic meaning cannot be decomposed in such a way that all lexemes

---

5 Chapter 7 gives plenty of evidence; there, we investigate what evidence of modification is extractable with data-driven methods.
contribute semantic load; in these cases, the idiomatic *the breeze, the bucket* or *de plaat* lack semantic significance.

Speakers’ intuitions about meaning decomposability may differ (Riehemann, 2001). Some speakers have a richer knowledge of the language and more experience with figurative language than others, thus while an experienced speaker sees an expression as semantically decomposable because the speaker ‘sees’ meaning in the different chunks, another speaker may consider the same expression as an unanalyzable (or partially analyzable) unit.

*Semantic decomposability* does not explain how the idiomatic meaning is derived. Instead, this process explains why in some idiomatic expressions, their constituents seem to carry meaning, therefore, revealing internal semantic structure. That is, they explain why *the bottom of the barrel* or *red herring* may appear on their own with the meaning they have in the original idiomatic expression. Semantic decomposability accounts for the evidence above, evidence that does not rule out the hypothesis of *conventionalized meaning*.

The second issue is, does the above evidence falsify the claim that the meaning of idiomatic expressions cannot be derived according to the *principle of compositionality*? In our opinion, it does not. We provide next our motivation.

In earlier work, idioms were characterized as ‘non-compositional’ because their peculiar meaning cannot be derived by means of the semantic composition function applied to the individual meanings of their constituent lexemes. Some authors argued that non-compositionality is a defining property of idioms (Katz, 1973; Fernando and Flavell, 1981; Verstraten, 1989; Schenk, 1992, 1994; Everaert and Kuiper, 1996). Alternative views allow that both compositional and non-compositional idioms exist (Wasow et al., 1982; Nunberg et al., 1994; Abeillé, 1995), with non-compositional idioms being the archetypal form (Moon, 1998).

A few issues need to be mentioned here. First, these inconsistent views arise from the use of a narrow or a wide concept of *idiom*. The former view considers an ‘idiom’ to be a non-compositional expression with an idiomatic meaning, whereas the latter view makes use of a more inclusive concept of *idiom*. The latter observation is often encountered in the literature (Nunberg et al., 1994; Moon, 1998). Whether compositional or not, idioms are considered a subset of fixed expressions in this thesis.

Second, both views agree that the meaning of some idioms cannot be explained by the *principle of compositionality*. This claim is justified because the idiomatic meaning of an expression cannot be derived from the individual meanings of its component lexemes, that is, the meaning denoted by the lexemes when they appear in isolation. Compare the derivation of the literal
meaning of *de wind in de rug hebben* in (18-a) and its figurative meaning in (18-b).

(18)  

a. Het is de droom van elke fietser om altijd *de wind in de rug* te hebben.  
It is the dream of every cyclist to always the wind in the back to have  
‘To always have the wind at his back is the dream of every cyclist.’

b. Ze *hebben* vermoedelijk *de wind flink in de rug* hebben.  
they have most probably the wind robust in the back  
due-to the high dollar exchange rate  
‘They most probably benefit strongly from the high dollar exchange rate.’

Nevertheless, work on formal semantics by Zadrozny (1992) showed that there is always a function to derive the meaning of a whole expression from the meanings of its parts using a compositional semantics. Consequently, the notion of non-compositionality no longer explains the distinction between idiomatic expressions and non-idiomatic ones. In fact, Manaster Ramer and Zadrozny (1992) argue that compositionality plays no role in describing the semantic structure of an idiomatic expression given that such expression behaves like a construction to which meaning is assigned as a whole.\(^6\)

Zadrozny’s (1992) claims threaten the well established view that ‘idiomatic expressions show a non-compositional meaning’. Although, under Zadrozny’s approach compositionality would no longer distinguish idiomatic expressions from non-idiomatic expressions, we believe that the notion of compositionality is still useful from a descriptive point of view.

The notions introduced in the above discussion of the semantic properties of fixed expressions are summarized in Table 2.1. To conclude, by making use of the notions of compositionality and semantic decomposability, fixed expressions can be grouped in two classes: (1) compositional expressions, that is, (rather) transparent with predictable meaning (e.g. *to adjourn a meeting, make progress, iemand van de straat houden*) and (2), noncompositional expressions that can be further split into (i) semantic decomposable (semi-opaque) (e.g. *spill the beans, iets in de gaten houden*) and (ii) semantic non-decomposable (fully opaque) (e.g. *shoot the breeze*).

\(^6\)Manaster Ramer and Zadrozny (1992) view such expressions as constructions, a concept introduced by Leonard Bloomfield.
2.3 Lexical and morphological properties

Once a fixed expression is taken up by the speakers community, it becomes part of the language i.e. it becomes institutionalized\(^7\) (Fernando and Flavell, 1981). To qualify as an institutionalized expression, the expression does not need to have a non-literal meaning but irregularities affect at least the lexical level. As illustrations, the adjective sound in sound asleep cannot be replaced by other semantically plausible adjectives *intact/*perfect asleep; neither can one component particle be replaced in to and fro (*forward and fro, to and *backward). Let us study other idiosyncrasies of fixed expressions that affect the lexical and morphological levels.

2.3.1 Selectional restrictions and lexical fixedness

The ‘fixed’ nature of fixed expressions partly derives from the lexical co-occurrence of two or more lexemes. The predicate de laatste hand aan iets leggen is a combination of a rather transparent NP de laatste hand ‘the finishing touch’, a verb leggen ‘lay’ with its usual meaning and a PP headed by aan ‘on’:

(19) \(\text{de laatste hand aan iets leggen (Dutch)}\)

\(\text{the last hand on something lay}\)

\(\text{‘to give the finishing touches to something’}\)

The lexemes laatste, hand, aan, leggen cannot be replaced by other lexemes with a similar (idiomatic) meaning without destroying the idiomatic meaning, resulting in a nonsensical string (20-a). Leaving out the adjective causes meaning change (20-b).\(^8\) Clearly, the lexemes mutually select each other and the lexemes are fixed.

(20) a. \(\text{de laatste hand *op/aan iets *plaatsen/leggen}\)

\(\text{the last hand *on/on something *put/lay}\)

b. \(\text{*de hand aan iets leggen}\)

\(\text{the hand on something lay}\)

Word co-occurrence can be captured in the form of selectional restrictions. Selectional restrictions should not be understood here as a selection of a semantic nature observed in combinations like drink water/juice/wine/coffe. Instead, the selectional restrictions present in fixed expressions are (to some

\(^7\)Eventually the fixed expression appears in dictionaries.

\(^8\)In a 300-million-word corpus, the NP de laatste hand exclusively occurs in the context of the verb leggen, with a total of 205 occurrences.
2.3. Lexical and morphological properties

extent) arbitrary. Two types of selectional restrictions surface in fixed expressions: lexically driven and grammatically driven. Firth (1957) named the first type lexical collocations, and the second type syntactic colligations. Lexical collocations roughly consist of affinities of a lexico-syntactic type among open class words such as verbs, nouns, adjectives or adverbs. In contrast, grammatical collocations (or syntactic colligations) generally involve at least one closed class word among particles, prepositions, and auxiliary verbs (Smadja, 1993, p. 171). As illustration, zoek naar in (21-a) shows grammatically driven restrictions, where a deverbal noun zoek ‘search’ selects for a complement headed by a specific preposition. Instead, een vergadering sluiten shows lexically driven restrictions. ‘To adjourn a meeting’ can be expressed as een vergadering sluiten but not as een vergadering dichtmaken (21-b).

(21) a. We zijn op zoek naar projecten.
   we are in search to projects
   ‘We are looking for projects’

b. een vergadering sluiten/dichtmaken
   a meeting close/close
   ‘to adjourn a meeting’

Although sluiten and dichtmaken conceptually mean ‘close’, dichtmaken does not collocate with vergadering. Lexical collocations are more unpredictable than syntactic colligations in fixed expressions, partly because syntactically driven restrictions can also be found outside the fixed expressions. Only lexically driven restrictions are common to all fixed expressions.10

2.3.2 Idiom families

Idiom families or sets of semantically related fixed expressions exist in which a lexicalized phrase (constituent) may combine with a limited set of verbs. Such ‘families’ often involve a light verb (aka. support verb), a verb which is semantically defective. Among the possible patterns, there exist [PP verb] combinations with a common PP that show slight differences in meaning, depending on the light verb (22):

(22) a. aan de gang zijn/blijven
    on the go be/remain

---

9Benson et al. (1986) made this observation with respect to collocations.
10Chapter 4 presents many examples where colligations and collocations are both present in Dutch lexicalized prepositional phrases e.g. bij gebrek aan ‘in the absence of, for need of’.
Chapter 2. On fixed expressions

`get started/keep going'

b. iets aan de gang houden/brengen

something on the go keep/bring

`keep sth going/get sth started’

The light verb contributes tense, aspect and ‘Aktionsart’ information, denoting causation, change of state, result of an event, etc. The syntactic constituency of the expression varies according to the ‘Aktionsart’.

In other cases, the same verbal lexeme is used with different nominal complements, all of them denoting similar meaning e.g. hit the hay/sack ‘go to bed’ and blow one’s stack/cool/top. Sometimes, substitution is due to dialectal variation or difference in register. An example in standard Dutch is iemand voor de gek houden ‘to fool someone’ that corresponds to iemand voor de zot houden in Flemish Dutch.

2.3.3 Restricted morphological operations

At the morphological level, fixed expressions are non-productive, however some expressions may allow limited inflectional and derivational variation. In a literal use, certain nouns display singular/plural alternation or the diminutive morpheme; in fixed expressions such nouns show restrictions in the use of such morphemes. In some expressions, the noun within the lexicalized argument is always plural e.g. perken ‘limits’ in iets binnen de perken houden ‘to limit, contain’:

(23) Dat houdt de kosten binnen de *perk/perken.

that keeps the costs within the limit/limits

‘That limits the costs.’

A singular or a plural noun realization may correlate with meaning changes in some expressions:

(24) a. Zo houden ze de extra arbeidskosten in de hand.

this way keep they the extra work costs in the hand

‘In this way, they have control over the extra work costs’

b. Dat kan alleen goed als de overheid het heft in handen houdt

that can only well as the government the grip in hands holds

‘That can only succeed as long as the government remains in control.’
Concerning the use of diminutive, *oogje* the diminutive of *oog* ‘eye’ is the dominant surface form in the expression *een oogje in het zeil houden* ‘keep a good eye on’ (25). Nevertheless, rare instances are found with the noun *oog* also.\(^{11}\) The diminutive is sometimes the only possibility (*lijntje* in *iemand aan het lijntje houden* ‘keep s.o. dangling’ (26).

(25) Keurmeesters *houden* een extra *oog/oogje* in het zeil.
inspectors keep an extra eye/little eye in the sail
‘Inspectors keep a very careful look-out.’

(26) ...die ook Douglas tijdlang *aan het lijst-line* houdt.
...who also Douglas long on the *line/little line keep
‘...who also keeps Douglas dangling’

Some Dutch fixed expressions exhibit archaic forms in noun inflection; the -e ending is an archaic form, outcome of an inflectional process (case marking) that is no longer productive in contemporary Dutch ((27) and (28)).

(27) *ten* kost-*e* van, *ten* last-*e* van, *van* hart-*e*
at the expense of, at the expense of, of heart
‘at the expense of, at the expense of, honestly’

(28) *ter* faveur-*e* van
in favor of
‘in favor of’

In addition, forms resulting from a merger of a preposition and a non-productive case marker surface in some expressions e.g. *ten* and *ter* in (27) and (28). Marginally, a determiner marked for case co-exists with its bare form (e.g. *de* or *den* in (29)):

(29) *zich* *van* de/*den* *domme* houden
REFLEXIVE of the/the+CASE stupid keep
‘play ignorant/innocent’

Such archaic forms provide evidence of frozen forms in the lexicon. *Frozen* words found inside fixed expressions do not provide much evidence for morphological productivity ((27) and (28)).

Nouns within fixed expressions also show restrictions in derivational morphology. In some cases, typically in the more transparent expressions, instances of noun-noun compounding are observed. The noun *deur* ‘door’ within the fixed expression *iets buiten de deur houden* ‘keep sth away’ is

---

\(^{11}\)In the *tWNC* corpus used, 5 out of 254 instances of the expression show no diminutive.
sometimes modified by a prefix that may be realized by an acronym (30) or another noun:

(30) ... om genetisch gemanipuleerd voedsel buiten de EU-deur te houden.
...to genetic manipulated food out the EU-door to keep
‘in order to keep genetically manipulated food outside the EU’

Among productive morphological processes, tense inflection is the one that is observed most frequently in expressions that involve a verbal lexeme. All forms of the verb houden are possible in the context of the expression vinger aan de pols houden. Proverbs and sayings tend to show preference for a particular tense (e.g. Hoge bomen vangen veel wind typically occurs in the present).

In some fixed expressions, morphological processes are completely blocked but in other expressions, restricted morphological variation is possible. The semantically more transparent expressions (e.g. iets buiten de deur houden ‘keep away’) seem to participate in more productive processes than the opaque expressions (e.g. iemand aan het lijntje houden). Since we eventually would like to identify criteria to classify fixed expressions, it would be reasonable to ask if we could use the evidence of unproductive and semi-productive morphological processes as a determinant of semantic decomposability. Chapter 7 provides evidence and discusses this issue further.

2.4 Syntactic properties

Fixed expressions (whether metaphorical or not) show varied phrase structure types such as NPs (a red herring, an arm and a leg), PPs (op bezoek ‘visit’, naar huis ‘to the house’), coordinated phrases (bits and pieces, by and large, op en neer ‘up and down’ (Dutch)), verbal phrases that exhibit varied syntactic structures, (in-)complete sentences, etc. Such structures show a departure from regular phrases due to irregularities concerning their semantics, lexicon, morphology and/or syntax. Fernando and Flavell (1981) observed of English idioms that they do not exhibit a specific form or morpho-syntactic structure; this applies to fixed expressions in general.

The syntactic structure is sometimes ill-formed or non-productive, e.g. trip the light fantastic in English. Some are not easily categorizable such as numerous adverbial-like phrases in Dutch (e.g. bij lange na ‘far from (a)’, om en nabij ‘around’, al met al ‘all in all’, nooit en te nimmer ‘never and
never’, *op en top* ‘very much’, *uit en te na* ‘very thoroughly’. The latter are comparable to *by and large* in English which is also syntactically ill-formed. Among PPs with rather transparent meaning, there are many syntactically marked prepositional phrases that include a determinerless noun phrase e.g. *at school, in hospital, etc.*; English and Dutch *determinerless* PPs have been studied recently by Baldwin et al. (to appear). In addition, so-called *syntactic constructions* abound where certain lexemes are fixed: *the X-er, the Y-er* in English, *van A tot B* ‘from A to B’ or phrases exhibiting the pattern *[NOUN POSTPOSITION]1, NOUN POSTPOSITION]2* such as *trap op, trap af* ‘up and downs the stairs’, *dag in, dag uit* ‘day in, day out’.

A common characteristic of fixed expressions is that they fail to conform to the traditional concept of lexeme i.e. a one to one mapping between form and meaning. The lexicalized phrases allow spaces in between; in addition, *some complex prepositional phrases* and *verbal phrases* may allow intervening adjuncts or other elements but at some level of description they behave as a unit. Evidence of intervening adverbs within ‘suspected fixed phrases’ is found in corpora (31):

(31) a. De teneur is dat Camus ‘niets aan actualiteit heeft ingeboet’, *in tegenstelling misschien tot* andere ‘existentialisten’ in het naoorlogse Parijs … ‘The drift of the argument is that Camus has not contributed anything to current affairs, as opposed to other existentialists in post-war Paris’

b. Economisch rekenen kunnen de ambtenaren al 47 jaar, *in tegenstelling dus tot* de dames en heren … *(lit.) …* economic figures can the officials already 47 years, in contrast thus with the ladies and gentleman …

c. *met dank* natuurlijk *aan* Shakespeare … ‘…thanks to Shakespeare, obviously’

Chapter 4 provides evidence of Dutch lexicalized *collocational prepositional phrases* that in the context of certain verbs allow extraposition (32):
Chapter 2. On fixed expressions

(32) Het orkest zal onder leiding staan van een Duitse dirigent.
    The orchestra shall under guidance stand of a German director.
    ‘The orchestra will be directed by a German director.’

Fixed arguments in support verb constructions and other idiomatic expressions are often separated from the verbs by adjunct modifiers (33), auxiliary verbs (34), etc.

(33) Hij ging onmiddellijk aan de slag.
    He went immediately on the turn
    ‘He got to work immediately.’

(34) Grunberg heeft ons dus niet voor de gek willen houden.
    Grunberg has us thus not in front of the crazy will hold
    ‘Grunberg would not have wanted to fool us.’

In addition to spaces and intervening elements, fixed expressions may allow open slots. In (35), the accusative NP complement harde overtredingen is an open slot because the NP may be realized by any semantically plausible NP. A few subject idioms exhibit an open slot in their complementation. Er schot zitten in ‘there exists progress in’ shows a partially lexicalized subject schot and a verbal PP in complement with an open NP slot (36). Open slots may need to satisfy syncategorematic constraints, e.g. only phrasal type (NP) in (35), and/or more specific morphosyntactic restrictions, e.g. (empty) specifier or adjective modifier in (37).

(35) [Harde overtredingen]np werden niet uit de weg gegaan.
    Strong offences were not out the way gone
    ‘Strong offences were not avoided.’

(36) a. In deze zaak zit geen schot
    in this case is no progress
    ‘There is no progress in this case.’

b. Er zit eindelijk schot in
    there is finally progress in
    ‘There is finally some progress.’

(37) Hij kon er echter geen gevolg aan geven.
    he could there really no consequence on give
    ‘He was not able to carry it out.’
Word order constraints may need to be specified in certain clausal constructions, especially in subordinate clauses. According to Hoeksema (p.c.), fixed arguments (whether NP, AP, PP) in Dutch idiomatic expressions (and also support verb constructions) tend to precede the verbal cluster inside subordinate clauses. Thus, word order constraints will have an affect on the fixed argument as well as the open slots.

Internal morphosyntactic flexibility is observed in some expressions like *het geld niet op iemand’s rug groeien* ‘not be made of money’ which may surface with three related syntactic patterns:

\[(38)\]
\[\text{a. } \left[\text{NP}_{\text{nom. het geld}} \text{niet PP}_{\text{op POSS rug}} \text{ groeien}\right]\]
\[\text{b. } \left[\text{NP}_{\text{nom. het geld}} \text{NP}_{\text{dat pron}} \text{niet PP}_{\text{op de rug}} \text{ groeien}\right]\]
\[\text{c. } \left[\text{NP}_{\text{nom. het geld}} \text{NP}_{\text{dat pron}} \text{niet PP}_{\text{op POSS rug}} \text{ groeien}\right]\]

The head noun *rug* within the complement PP may be preceded by a definite or a possessive determiner agreeing with the dative NP (if present) (39):

\[(39)\]  
\[\text{Natuurlijk groeit het geld me, niet op de/mijn, rug}\]
\[\text{Obviously grows the money me-DAT not on the/my back}\]
\[\text{‘I am not made of money, obviously.’}\]

If the dative NP is not present, a possessive determiner is required (40). More fixed expressions exhibit agreement relations between two of their constituents, often between the subject or the object and another fixed argument.

\[(40)\]  
\[\text{Natuurlijk groeit het geld niet op mijn rug}\]
\[\text{Obviously grows the money not on my back}\]
\[\text{‘I am not made of money, obviously.’}\]

Quantifiers, demonstrative determiners, intensifiers and various sorts of modifiers may be allowed in some fixed expressions. Whereas the Dutch expression *vinger aan de pols hebben* ‘follow the latest developments’ allows adjectives (41) and post-nominal PP modifiers (42), *een oog(je) in het zeil houden* ‘keep an eye on’ blocks variation inside its PP complement; by contrast, variation is allowed inside the NP headed by oog (43).

\[(41)\]  
\[\text{Parmentier toont de vinger aan de literaire pols te hebben}\]
\[\text{Parmentier shows the finger on the literary pulse to have}\]
\[\text{‘Parmentier shows that he’s followed the latest developments in literature.’}\]

\[(42)\]  
\[\text{een oog(je) in het zeil}\]
\[\text{houden}\]
\[\text{‘keep an eye on’ blocks variation inside its PP complement; by contrast, variation is allowed inside the NP headed by oog (43).}\]

\[(43)\]  
\[\text{een oog(je) in het zeil}\]
\[\text{houden}\]
\[\text{‘keep an eye on’ blocks variation inside its PP complement; by contrast, variation is allowed inside the NP headed by oog (43).}\]

\[12\text{See Chapter 6 for empirical evidence in support of this claim.}\]
Chapter 2. On fixed expressions

(42) Hebben ze nog een vinger aan de pols van de tijdgeest?
Have they still a finger on the pulse of the spirit of the age
‘Do they still follow the latest developments on the feel of the times?’

(43) ...dat hij op deze manier ook een wetenschappelijk oog in het zeil kan houden
...that he in this way also a scientific eye
in the sail can keep
‘(...) that in this way, he can scientifically keep an eye on it.’

Fellbaum (1993) tried to identify denoting NPs in idioms by exploring the range of determiners accepted and the determiner’s use inside the NPs. Fellbaum investigated the definite, indefinite and zero determiners (the latter found in bare plural NPs). The nature of the determiner points at the semantic status of the noun phrases in many idioms, and this in turn predicts the range of possible determiner variations (Fellbaum, 1993).

Let us examine some examples by Fellbaum (1993). The noun beans in the context of spill accepts several determiners and the idiom still maintains its figurative interpretation (44). Beans also allows the insertion of adjectives as the examples given by Riehemann (2001) show in (45).

(44) John spilled the/no beans
(45) royal beans, some very complimentary beans, the politically charged beans, the oboistic beans

On the contrary, if the determiner the in front of bullet is replaced in to bite the bullet, the figurative meaning disappears (46):

(46) a. You will have to bite the bullet.
    b. You will have to bite a/no/every bullet.

This different behavior may be explained by the fact that the NP object in the expression spill the beans has semantic significance (the figurative meaning ‘the secret’) whereas the bullet in the context of bite the bullet does not.

Fellbaum (1993) observed that some determiners allow limited substitution by other quantifiers, negative determiners, demonstratives or insertion of specificity-inducing adjectives is possible in some compositional VP idioms.13 The substitution by other determiner brings about changes in the truth values of the expressions, just like in literal language (e.g. in (44)). For us, one relevant observation by Fellbaum is that determiner’s flexibility is only

---

13Fellbaum’s use of the term ‘compositional idioms’ corresponds to ‘decomposable idioms’.
found in compositional idioms where it is ‘largely an indicator of the noun’s referential status’. In non-compositional idioms the determiner is invariable.

In order to license intervening elements as well as open slots in the relevant fixed expressions, their lexical representation requires internal syntactic structure. Open questions include: what constraints are needed and how finely grained must they be.

### 2.4.1 Sense syntax asymmetry

A predicate expresses a semantic relation among participants; these participants are called the arguments of the predicate. The argument structure of e.g. a verbal lexeme consists of the semantic relation expressed by the verb and the semantic roles expressed by the participants. Here, the argument structure of a predicate is represented as \( \text{relation} (\text{participant\_role}_i, \ldots) \). Adjuncts are not included in the argument structure.

At the syntax-semantics interface, verbal lexemes inside fixed expressions may show **non-homomorphism** between the argument structure and the syntactic frame specifying the syntactic dependents. Non-homomorphism obtains when syntax and semantics are constructed differently, e.g. when the literal verb is different from the verb within the lexicalized fixed expression.

In its most usual sense, the lexeme *gaan* denotes a two-place relation \( \text{go} (\text{agent}, (\text{goal})) \); the argument structure exhibits two participant roles. In syntax, the corresponding verbal lexeme shows the subcat frame: \([\text{NP}_{\text{nom}} \text{NP}_{\text{acc}} \text{XP}]\). At the syntax-semantics interface, the agent is realized by the \(\text{NP}_{\text{nom}}\) and the goal is typically realized by a directional XP (PP or NP), reflecting a homomorphism between the semantic and the syntactic structures.

The phrase *uit de weg* (lit. ‘out of the path’) combines with *gaan* in an idiomatic context (47). In this case, the PP *uit de weg* is not a directional adjunct since, removing the phrase *uit de weg* brings up ungrammaticality as well as meaning change (48).

(47)  
\[
\text{John ging de problemen uit de weg.} \\
\text{John went the problems out of the way} \\
\text{‘John avoided the problems.’}
\]

(48)  
\[
\text{* John ging de problemen} \\
\text{John went the problems}
\]

In the idiom *iets uit de weg gaan* ‘avoid something’, *gaan* no longer shows such a homomorphism. The verbal lexeme exhibits an ‘unexpected’ syntactic constituency. Whereas one could say that *gaan* selects for a \([\text{NP}_{\text{nom}} \text{NP}_{\text{acc}} \text{PP}_{\text{uit de weg}}] \) frame in syntax, in semantics, *iets uit de weg gaan* denotes a two-
Chapter 2. On fixed expressions

place relation \texttt{avoid(agent,theme)} expecting an agent and a theme. \textit{Uit de weg} contributes no meaning on its own, rather it is the combination \textit{uit de weg gaan} that denotes. (49) specifies the syntactic dependents of the verb \textit{gaan} as well as the participant roles in the two uses: literal (49-a) and idiomatic (49-b).

(49) a. \textit{gaan} \{ \texttt{NP} \texttt{nom}, \texttt{XP} \} : \texttt{go(agent,(goal))}
b. \textit{gaan} \{ \texttt{NP} \texttt{nom}, \texttt{NP} \texttt{acc}, \texttt{PP} \texttt{uit de weg} \} : \texttt{avoid(agent,theme)}

At the level of description, non-homomorphism between the argument structure of the main predicate in a fixed expression and its syntactic valence is claimed to be a consequence of the fact that the fixed expressions exhibit non-compositional semantics. The phenomenon does not affect all fixed expressions and it is by no means restricted to Dutch.

2.4.2 Syntactic versatility

Diagnostic tests such as topicalization, extraction, passivization, raising, control, coordination, etc. are often applied to determine whether certain constituents in fixed expressions behave like regular constituents in non-fixed expressions, i.e. those constituents to which general grammar principles and other lexical rules may apply.

Topicalization refers to the realization of a constituent (other than the subject in SVO languages) in initial position in a sentence. Concerning the Dutch expression \textit{iets uit de weg gaan} ‘avoid something’, no examples of the topicalized fixed chunk \textit{uit de weg} were found in the TwNC corpus. However, native speakers accept (50) as possible, in which the fixed chunk occurs in topic position, once the context has been enriched.

(50) Colijn ondernam weinig actie; maar \textit{uit de weg ging} Colijn een rechtstreekse aanval op de vorstin echter wel

‘Colijn undertook little action; but he, however, avoided a direct attack on the queen.’

Some authors argue that topicalization only applies to meaningful expressions (Schenk, 1994). In the example \textit{Marie’s hart brak Piet} ‘Marie’s heart broke Piet’ taken from Schenk (1994, ex. 67), part of the topicalized constituent obligatorily carries meaning. Marie needs a referent. Topicalization changes the topic comment structure (the information structure) of a sentence. Since a meaningless chunk is not part of a topic comment structure,
2.4. Syntactic properties

this structure cannot be changed by topicalizing such an expression (Schenk, 1992).

Schenk’s (1992) argument can be questioned. Though topicalized constituents typically denote meaning, it is not necessarily always the case (e.g. existential ‘there’ and weather and expletive ‘it’ in English (Schenk, 1992)). A few instances of the topicalized fixed chunk in the expressions *iets ter sprake komen* ‘to come up (during discussion)’ and *door de mand vallen* ‘fail as; make a poor showing’ were found in the same corpus (51):\(^{14}\)

(51)  a. *Ter sprake kwamen* onder meer de veiligheid en softdrugs.
      to-the talk came under more the safety and soft drugs
      ‘Safety and soft drugs, among others, came up in the discussion.’
    b. *Door de mand viel* prof. T. Schalken, de bekendste
      through the basket fell prof. T. Schalken, the most-famous
      rechtsgeleerde Nederlander . . .
      jurist Dutch . . .
      ‘Prof. T. Schalken, the best known Dutch lawyer, made a poor
      showing ( . . . )’

If a fixed chunk of a fixed expression appears in topicalized constructions, the only conclusion we can make is that the fixed chunk is an ‘independent’ syntactic constituent. Investigating whether the chunk is meaningful or not falls outside the scope of this work.\(^{15}\)

A fixed complement (*gevolg* in *gevolg aan geven*) may be subject of a passive sentence (52-a) but not all expressions allow passive. *De plaat poeßen* ‘to depart unnoticed’ does not allow passive (52-b):

(52)  a. Maar er werd nooit *gevolg aan gegeven.*
      But there was never consequence on given
      ‘But there has never been any follow-up.’
    b. *De plaat wordt gepoëst* door Karel.
      the plate was cleaned by Karel

Similarly, some fixed constituents can realize the subject of a raising construction (53) or the subject of a control *vp* (54) (cf. (Nunberg et al.,

---

\(^{14}\)We sought topicalization instances of fixed chunks for a few fixed expressions and they seem rather rare in the corpus we used. Perhaps topicalization is more common in spoken language.

\(^{15}\)Deciding whether a topicalized fixed constituent carries meaning or not, is crucial in a syntax-semantics analysis since one needs to assign the (formal) semantic representation to the topicalized constituent.
This also shows that fixed expressions may occur in embedded contexts.

(53) ... omdat *er schot* lijkt te zitten in de regeling
    ... because it progress seems to sit in the regulation
    ‘... because it seems that there has been progress in the regulation.’

(54) The lawsuit over contested seats on the Auburn Board of Trustees may be over, but now *the piper wants to be paid*.

Though not commonly found, a head noun part of a fixed expression may be located in a different clause than the rest of the expression. Riehemann (2001) provides substantial evidence of this phenomenon in English semantic decomposable idioms. *Gijseling* part of the Dutch expression *iemand in gijseling houden* ‘hold s.o hostage, imprisoned’ introduces a relative clause where the rest of the expression is realized:

(55) ... niet de ruisloze *gijseling* waarin zijn
    ... not the noiseless hostageship where his
    communicatie-apparaten hem *houden*
    communication-apparats him hold
    ‘... not the noiseless hostageship in which his communication appar-
    ats keep him’

A difficult part of every study of fixed expressions is to find generalizations on their syntactic distribution. At an observational level, there is no uniform presence or absence of syntactic constraints in all fixed expressions since not all fixed expressions exhibit the same syntactic versatility.

It has been claimed in the literature that certain constraints observed in productive phrases also hold for idioms (Katz, 1973; Abeille, 1995; Nunberg et al., 1994). For example, passivization applies to transitive verbs (but also impersonal verbs); control is sometimes possible if the fixed constituent refers to an animate entity, etc. This would explain that in English *spill the beans* ‘reveal a secret’ can be passivized but *kick the bucket* ‘die’ cannot. *Kick the bucket* in its idiomatic use denotes a one-place semantic relation, thus lacking a topic argument (direct object) that can be promoted to subject position. In addition, irregularity pervades in some idiomatic expressions that only occur in particular syntactic contexts (e.g. *het pleit is/wordt beslecht* ‘the argument was settled’ only occurs in the passive).

To be certain, further systematic research on the syntactic versatility of

---

verbal fixed expressions is necessary, research that needs to include expressions with various morphosyntactic structure, not just V NP combinations.

Other factors observed by Nunberg et al. (1994) are proverbiability, informality and affect. These factors are not relevant in the presentation that follows.

We summarized the semantic, lexical and morphosyntactic properties in Table 2.1. These features need to be taken into account when deciding the necessary representation of a fixed expression in a computational lexicon and the description of the grammatical operations it may undergo.

In deciding whether an expression in the language qualifies as a fixed expression, a lexicographer or a computational linguist does not need to check whether each single item listed in Table 2.1 applies or not. However, semantic and morphosyntactic properties need to be tested, since some fixed expressions are semantically compositional but show morphosyntactic deficiencies, or alternatively, they might show traces of regular morphosyntax, but have lexical or semantic irregularities (Everaert, 1993).

2.5 Summary

An exhaustive characterization of fixed expressions is a complex undertaking, partly due to the idiosyncrasies of the data but also due to the different judgements and intuitions that native speakers and linguists have about the data.

The previous characterization aimed at collecting evidence of features that need to be captured by data-driven models to be described in Chapters 4 and 5. Furthermore, such features (at least some of them) need to be specified in a computational lexicon and allowed by the corresponding grammar.
### Lexeme level

| Ordinary lexemes | high co-occurrence frequency  
|                 | *vergadering sluiten*  
| Nonce words     | peculiar meaning only  
|                 | present in fixed expression  
|                 | *de laatste hand aan iets leggen*  
|                 | only exist in fixed expression  
|                 | *in petto hebben*  

### Morphology

| Inflectional    | singular/plural morpheme in nouns  
|                 | diminutive  
|                 | if adjective gradable  
|                 | archaic forms (-e ending)  
|                 | case marking (determiners)  
|                 | tense inflection  
| Derivational    | prefix, compounding  

### Semantic

| Denoting vs. non-denoting lexemes | only literal meaning  
|                                  | polysemous  
|                                  | non-denoting  
| Opacity                         | transparent  
|                                  | semi-transparent  
|                                  | semi-opaque  
|                                  | fully opaque  

### Syntactic

| Morpho-syntactic structure     | regular structure  
|                                 | synt. marked  
| Internal variation              | ill-formed  
| Intervening adjuncts            | modification, quantification, determiners  
| Agreement relations             | between required constituents  
| Syntactic versatility           | with subject/object  
| Open slots                      | topicalization, passive, ...  
| Non-homomorphism (syntax-semantics interface) | words/phrases  

Table 2.1: Features to check in a (potential) fixed expression.