Semagrams, Another Way to Capture Lexical Meaning in Dictionaries

Carole Tiberius and Tanneke Schoonheim

_Instituut voor Nederlandse Lexicologie (INL)_
_{carole.tiberius, tanneke.schoonheim}@inl.nl

The Dutch dictionary Algemeen Nederlands Woordenboek (ANW) offers a twofold meaning description: definitions are accompanied by a semagram, a frame-based representation of knowledge typically associated with a word. We describe in this paper the role of semagrams in the context of the ANW dictionary project and we provide some guidelines for the future.

**Keywords:** lexical meaning, frames, dictionary, Dutch

1. Algemeen Nederlands Woordenboek

The ANW is a comprehensive online scholarly dictionary of contemporary standard Dutch, which is being compiled at the Instituut voor Nederlandse Lexicologie (INL). It has been set up as an online dictionary right from the start and as such, it truly represents a new generation of electronic dictionaries in the sector of academic and scientific lexicography. The dictionary focuses on written Dutch and covers the period from 1970 to 2018. It is a corpus-based dictionary, based on the ANW corpus, a balanced corpus of just over 100 million words, which was compiled specifically for the project.

The ANW is a very informative dictionary, providing a detailed description of each lexical entry. Special attention is paid in the dictionary
to words in context (combinations, collocations, idioms, proverbs), to relations with other words (lexical relations like synonymy, antonymy, hyponymy, hypernymy)\textsuperscript{2}, to semantic relations within the entry (metaphor, metonymy, generalisation, specialisation)\textsuperscript{3} and to morphological patterns, the word structure of derivations and compounds. This is illustrated in Table 1 for the lemma *koe* ‘cow’ in the sense of a ‘female adult bovine’\textsuperscript{4}.

For the lemma *koe* ‘cow’, four main senses and three subsenses are distinguished in the ANW. For each one of them the relation to the core sense is specified (see information in italics below):

1.0 *volwassen vrouwelijk rund* ‘adult female bovine animal’
   1.1 *rund* ‘bovine animal’ (*generalisation*)
   1.2 *geslachte koe* ‘slaughtered cow’ (*metonymy*)
   1.3 *afbeelding van een koe* ‘image of a cow’ (*metonymy*)
2.0 *wijfje van een ander dier* ‘the female of another animal’ (*generalisation*)
3.0 *vrouw* ‘woman’ (*metaphor*)
4.0 *iets heel groot* ‘something big’ (*metaphor*)

One of the main innovations of the ANW is that it offers a twofold meaning description: definitions are accompanied by ‘semagrams’ (henceforth SG), a term coined by Fons Moerdijk (2007; 2008), the first chief-editor of the ANW. In this paper, we critically discuss the role of SGs

\textsuperscript{1} See for a general introduction to the ANW and its features Schoonheim and Tempelaars (2010).

\textsuperscript{2} This concerns for example the hypernymic/hyponymic relation between *transport* (hypernym) and *car* (hyponym), the synonymic relation between *bike* and *bicycle*, or the antonymic relation between *light* and *dark*.

\textsuperscript{3} This concerns for example the metaphoric relation between the two senses of the word *sea* in *they loved to look over the sea* where *sea* means ‘large amount of salt water’ and *they looked at a sea of possibilities*, where *sea* means ‘large amount’, the metonymic relation between the two senses of the word *China/china* in *China* ‘the Asian country bearing this name’ and *china* ‘porcelain or other types of ceramic, originally produced in China’, the generalisation in the two senses of the word paper, one of the general senses being ‘newspaper’, but in a sentence such as *the papers say they’ve lost their hope* in the more general sense of ‘all media, regardless whether they appear on paper or not’, and the specialisation in the two senses of the word *pill*, the general sense being ‘kind of medicine’, whereas the specialised sense is ‘birth control pill’.

\textsuperscript{4} http://anw.inl.nl/article/koe
in the practice of making a dictionary. The paper is organised as follows.

Table 1. Words in context: the lemma *koe* ‘cow’ in the ANW.

<table>
<thead>
<tr>
<th><strong>COMBINATIONS AND COLLOCATIONS</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>with preceding adjective</strong></td>
<td></td>
</tr>
<tr>
<td>een bonte koe ‘a spotted cow’</td>
<td>een gekke koe ‘a silly cow’</td>
</tr>
<tr>
<td>een dolle koe ‘a mad cow’</td>
<td>een heilige koe ‘a sacred cow’</td>
</tr>
<tr>
<td>een bronstige koe ‘a cow in heat’</td>
<td>een rode koe ‘a brown cow’</td>
</tr>
<tr>
<td>een drachtige koe ‘a pregnant cow’</td>
<td>een roodbonte koe ‘a spotted brown cow’</td>
</tr>
<tr>
<td><strong>with adjectival past participle</strong></td>
<td></td>
</tr>
<tr>
<td>een afgekalfde koe ‘a calved cow’</td>
<td>een nieuwmelkte koe ‘a newly milked cow’</td>
</tr>
<tr>
<td>een met BSE besmette koe ‘a BSE-infected cow’</td>
<td></td>
</tr>
<tr>
<td><strong>as subject of a verb</strong></td>
<td></td>
</tr>
<tr>
<td>grazen ‘graze’</td>
<td>kalven ‘calve’</td>
</tr>
<tr>
<td>herkauwen ‘ruminate’</td>
<td>loeien ‘moo’</td>
</tr>
<tr>
<td><strong>as object of a verb</strong></td>
<td></td>
</tr>
<tr>
<td>koeien evacueren ‘evacuate’</td>
<td>koeien insemineren ‘inseminate’</td>
</tr>
<tr>
<td>koeien fokken ‘breed’</td>
<td>koeien melken ‘milk’</td>
</tr>
<tr>
<td>koeien hoeden ‘herd’</td>
<td>koeien ruimen ‘slaughter’</td>
</tr>
<tr>
<td><strong>with a preceding noun</strong></td>
<td></td>
</tr>
<tr>
<td>een aantal koeien ‘a number of cows’</td>
<td>een kudde koeien ‘a herd of cows’</td>
</tr>
<tr>
<td><strong>IDIOMS</strong></td>
<td></td>
</tr>
<tr>
<td>de heilige koe ‘the sacred cow’</td>
<td>een waarheid als een koe ‘a truism’</td>
</tr>
<tr>
<td><strong>PROVERBS</strong></td>
<td></td>
</tr>
</tbody>
</table>
Er zijn wel meer koeien die Blaar/Bles heten ‘what’s in a name?’
Geen koe zo bont of er is wel een vlekje aan ‘there's no smoke without fire’
Je weet nooit hoe een koe een haas vangt ‘± you never know your luck; ± nothing is impossible (to a willing heart)’

**WORD FAMILY**

**As part of a derivation**
koetje ‘little cow’

**As right element in compounds**
adoptiekoe ‘adopted cow’
antiallergiekoe ‘anti-allergy cow’
BSE-koe ‘BSE cow’
kinderboerderijkoe ‘children’s farm cow’
kloonkoe ‘cloned cow’
melkkoe ‘milch cow’
vleeskoe ‘meat cow’
zoogkoe ‘suckler cow’

**As left element in compounds**
koehandel ‘horse trading’
koehoorn ‘cow’s horn’
koeienhoorn ‘cow’s horn’
koeienkeurmeester ‘cow inspector’
koeienkop ‘cow’s head’
koeienmelk ‘cow milk’
koeienoog ‘cow’s eye’
koeienstal ‘cowshed’
koeienvlees ‘cow meat’
koeienziekte ‘cow disease’
koeiemelk ‘cow’s milk’
koeiemest ‘cow manure’
koeistal ‘cowshed’

**LEXICAL RELATIONS**
hyperonym: rund ‘bovine animal’
antonym: stier ‘bull’

Section 2 introduces the notion of SG and discusses then the development of the type templates. In Section 3 we describe the role of the SG in the dictionary and we provide some guidelines for the future. Section 4 concludes the paper.

2. Semagrams

2.1 Origin and motivation

Moerdijk (2007; 2008) defines an SG as the representation of knowledge
associated with a word in a frame of slots and fillers. Slots are conceptual structure elements which characterise the properties and relations of the semantic class of a word (e.g. COLOUR, SMELL, TASTE, COMPOSITION, COMPONENTS, PREPARATION for the class of beverages). On the basis of these slots specific data is stored (fillers) for the word in question. The abstract structure is called a type template in ANW jargon. An SG is such a type template populated with concrete word data. Each semantic class has its own predefined type template with its own slots. For instance, the type template for the class of animals contains the slots PARTS, BEHAVIOUR, COLOUR, SOUND, BUILD, SIZE, PLACE, APPEARANCE, FUNCTION and GENDER, whereas the type template for beverages has slots for INGREDIENT, PREPARATION, TASTE, COLOUR, TRANSPARANCY, USE, SMELL, SOURCE, FUNCTION, TEMPERATURE and COMPOSITION. So far we have focused mainly on SGs for nouns, those for verbs and adjectives will be different. Here below is an example of an SG for a member of the animal class, i.e. koe ‘cow’, meaning ‘female adult bovine’.

5 SGs are not the only kind of frames used in a lexicographic project. When talking about frames, most linguists immediately think of Fillmore’s FrameNet. Understandable, but the ANW frames cannot be identified with his frames. The frames of FrameNet represent a conceptual structure of a stereotypical stitution. The actors, objects, properties and events, that can be distinguished in such a situation form the frame elements. These elements are found in sentences via the semantic role they play with regard to a predicate. Frame elements are ‘roles’ which can be used to semantically annotate the sentences of corpora.

The ANW frames, on the other hand, are more in line with the approaches of Wierzbicka (1985), Wiegand (1989, 1992), Konerding (1983) and Konerding and Wiegand (1994). They represent the conceptual structure underlying a word (lexical concept) rather than describing the situation underlying a sentence. They encode the knowledge typically associated with a lemma. The frame elements are not conceptual roles, but conceptual features. They are not the result of a semantic-syntactic analysis and their goal is not the semantic analysis of sentences. (Moerdijk 2008: 561)

6 Many features from the noun templates can be also used for verbs and adjectives. Slots like SPEED and MOTION can be filled with ‘fast’ and ‘moving forward’ in an SG of verbs like rennen ‘run’, and racen ‘race’. For an insight in the semantic classification of adjectives for Dutch, see Heyvaert (2010).
<table>
<thead>
<tr>
<th><strong>a cow</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>UPPER CATEGORY</td>
<td>is an animal</td>
</tr>
<tr>
<td>CATEGORY</td>
<td>is a bovine</td>
</tr>
<tr>
<td>SOUND</td>
<td>moos/lows, makes a sound that we imitate with a low, long-drawn ‘moo’</td>
</tr>
<tr>
<td>COLOUR</td>
<td>is often black and white spotted, but also brown and white spotted, black, brown or white</td>
</tr>
<tr>
<td>SIZE:</td>
<td>is big</td>
</tr>
<tr>
<td>BUILD</td>
<td>is big-boned, bony, large-limbed in build</td>
</tr>
<tr>
<td>PARTS</td>
<td>has an udder, horns and four stomachs: paunch, reticulum, third stomach, proper stomach</td>
</tr>
<tr>
<td>FUNCTION</td>
<td>produces milk and (being slaughtered) meat</td>
</tr>
<tr>
<td>PLACE</td>
<td>is kept on a farm; is in the field and in the winter in the byre</td>
</tr>
<tr>
<td>AGE</td>
<td>is adult, has calved</td>
</tr>
<tr>
<td>PROPERTY</td>
<td>is useful and tame; is considered as a friendly, lazy, slow, dumb, curious, social animal</td>
</tr>
<tr>
<td>GENDER</td>
<td>is female</td>
</tr>
<tr>
<td>BEHAVIOUR</td>
<td>grazes and ruminates</td>
</tr>
<tr>
<td>TREATMENT</td>
<td>is milked every day; is slaughtered</td>
</tr>
<tr>
<td>PRODUCT</td>
<td>produces milk and meat</td>
</tr>
<tr>
<td>VALUE</td>
<td>is useful</td>
</tr>
</tbody>
</table>

**Figure 1.** Semagram for koe (cow) (Moerdijk et al. 2008: 19)

At present the data in the slots is completed manually by the lexicographers based on information in the ANW corpus, reference works (such as dictionaries and encyclopaedia) as well as their language and world knowledge. Not all slots in the type template have to be completed in all cases. Only those for which there is a value are shown in the above example.

Moerdijk (2007; 2008) notes that using SGs for the representation of meaning in dictionaries is desirable for a number of reasons. First, inserting SGs in the semantic dimension of an electronic dictionary leads to a much richer semantic description, in which implicit knowledge of the definitions has been made explicit and more (also encyclopaedic) knowledge is
recorded than can be represented in the traditional definition formats. This follows the cognitive-semantic vision of meaning according to which the description of the meaning of a word cannot be captured solely via necessary and sufficient distinctive features. Features which are (proto) typical also need to be specified and these are not necessarily distinctive. Information such as that cows moo, ruminate, are often black and white spotted, have an udder, etc. are also relevant, capturing important aspects of our knowledge concerning cows. Hence, this kind of information should be part of the meaning description of the lemma *koe* ‘cow’ in the dictionary. It is precisely one of the functions of SGs to capture this kind of knowledge. They show how our knowledge and views of the world are reflected in language. Heyvaert (2014: 33) notes that SGs also capture the paradigmatic and syntagmatic relations of a word. As the SG for *koe* illustrates, the semantic relations between *koe* and its idiomatically related terms are made explicit. For instance, in the idiom *koeien van fouten* ‘huge mistakes’, *koeien* refers to the size of the mistakes, that is: huge (cf. the value of the SIZE slot in the SG for *koe*).

In addition, SGs can explain different aspects of the linguistic behaviour of a word. For example, the formation and interpretation of compounds, regular polysemy and the occurrence of some more or less lexicalised syntactic combinations. Nouns denoting animals, for instance, can be used to qualify a person by attributing one or more of the properties of the animal to that person. For example, the word *koe* can offensively be used to refer to a woman who is dumb (cf. the value of the PROPERTY slot), while *vos* ‘fox’ can refer to a sly person.

Using SGs should lead to more consistent and uniform definitions in dictionaries. Dictionary definitions often suffer from arbitrariness and inconsistency with respect to the choice of the keywords for the genus term as well as with respect to the nature and number of features included. Table 2 (taken from Moerdijk 2007: 65) shows that the same meaning is often expressed in different ways in different dictionaries. The different Dutch dictionaries use different genus terms and list different features in their

---

7 See also Rosch’s prototype theory (Rosch 1978).
8 [http://anw.inl.nl/article/koe#subject:verbindingen](http://anw.inl.nl/article/koe#subject:verbindingen)
definition of ‘cow’. Features such as ‘gives milk’, ‘moos’ etc. are given only in the Van Dale GWNT.

**Table 2.** Dictionary definitions for the lemma *koe* ‘cow’

<table>
<thead>
<tr>
<th>Dictionary</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verschueren:</td>
<td>vrouwelijk rund ‘female bovine animal’</td>
</tr>
<tr>
<td>Koenen:</td>
<td>volwassen wijfjesrund ‘adult female bovine animal’</td>
</tr>
<tr>
<td>Van Dale GWHN:</td>
<td>vrouwtje van runderen en van andere grote dieren, bv. olifanten, walvissen ‘the female of bovine animals and other large animals, e.g. elephants, whales’</td>
</tr>
<tr>
<td>Van Dale GWNT:</td>
<td>vrouwelijk huisrund, een tweevoetig herkauwend zoogdier dat loeit en melk geeft, in de regel een dat reeds gekalfd heeft (<em>Bos taurus</em>) ‘female domesticated bovine animal, a cloven-hoofed ruminant mammal that moos and gives milk, and usually has calved’</td>
</tr>
<tr>
<td>WNT:</td>
<td>vrouwelijk huisrund dat in de regel reeds gekalfd heeft ‘female domesticated bovine animal that usually has calved’</td>
</tr>
</tbody>
</table>

Within one dictionary the definitions of instances of the same semantic class can also differ significantly as shown in Table 3 with the lemmas *jak* ‘yak’ and *bizon* ‘bison’ as defined in the Van Dale GWNT.⁹

When editing is done in a modular way via semantic classes, decisions concerning genus terms and features to be included in the dictionary definitions as well as SGs, can be well controlled. All this leads to a more consistent and uniform semantic description in the dictionary.

Finally, SGs enhance search and query facilities. This is particularly the case for queries from content to form. For instance, a user who cannot think of the word *apiarist* can find this word via related elements (e.g. ‘bees’, ‘keep’) he knows. Any of these terms can be used for search (Zock et al. 2010). SGs play an important role in this endeavour, as they reflect the

---

⁹ Atkins and Rundell (2008: 124) give a similar example comparing the definitions for *lion* and *tiger* in an English learners’dictionary.
Table 3. Dictionary definitions for lemmas belonging to the same semantic class

<table>
<thead>
<tr>
<th>lemma</th>
<th>definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>jak</td>
<td>rund van de hooggebergten van Tibet, krachtig en sierlijk van bouw en met lange zijdeachtige haren, m.n. aan de staart, die de zgn. ‘paardenstaarten’ van de pasja's leverden; de Kirgiezen en Mongolen houden het als huisdier (Bos mutus) ‘a bovine animal from the highlands of Tibet, with a powerful and graceful build, with long silky hair, esp. on the tail, which provide the so-called ‘horsetails’ of the pashas; domesticated by the Kyrgyz and the Mongols’</td>
</tr>
<tr>
<td>bizon</td>
<td>geslacht van wilde runderen ‘a species of wild ox’</td>
</tr>
</tbody>
</table>

stereotypical knowledge speakers have about the concepts with which the words referring to them are associated (Moerdijk 2003).

SGs do not only allow to go from content to the corresponding form (word), but they also allow to retrieve sets of words on the basis of one or more shared content features. Thus a user of the ANW dictionary can retrieve all names for female animals in Dutch by combining in his query the values of the slots CATEGORY (‘animal’) and GENDER (‘female’). This is one of the reasons why the features in the ANW structure are more finely split than in the more global schemata described by Wierzbicka (1985) and Martsa (2000) where such queries cannot be made. In the next section we will give a short overview of the development of the type templates.

2.1 Type templates

The development of the type templates has been described in detail in Moerdijk (2007; 2008) and Moerdijk and Tiberius (2009). Here we summarise the characteristics of the resulting system.

- The ANW adopted a ‘bottom up’ approach to define the semantic hierarchy and the type templates starting from a manual analysis of dictionary definitions. For example, a simple definition ‘adult female bovine’ for cow would be split into the following segments:
  - KEYWORD: bovine
  - AGE: adult
  - GENDER: female
- The ‘semantic hierarchy’ consists of a (relatively small) number of
semantic classes and subclasses. The hierarchy is not a complete hierarchy, but a set of disjoint hierarchies (somehow like WordNet\textsuperscript{10}). The semantic hierarchy is refined as work progresses.

- An effort is made to keep the hierarchies of the different semantic classes clean (see the ontoClean portal\textsuperscript{11} on the cleanliness of ontologies). To this end:
  
  o one dimension for classification is adopted per semantic class. This dimension is predefined. For instance, the main dimension for classifying animal names is based on the taxonomic classification.

  o an element ‘feature class’ has been introduced to deal with alternative classifications, i.e. from a different perspective than the main dimension. For instance, in addition to the taxonomic classification, it is also possible to classify animal names on the basis of their FUNCTION (e.g. *trekdier* ‘draught animal’, *huisdier* ‘pet’), LOCATION (e.g. *zoetwatervis* ‘freshwater fish’), AGE (e.g. *kalf* ‘calf’) etc. Thus in the SG for *trekdier* the feature class will have the value FUNCTION. Note that feature classes can be combined.

  For instance, *merrieveulen* ‘filly’ will have the feature classes AGE and GENDER. Feature classes can be used to capture what Cruse (2002) calls nominal kind terms. These are non-taxonomic terms, which call for a classification into semantic properties rather than into taxonomic types. See also Lenci et al. (2000) for an illustration of the semantic dimension of humans as encoded in the SIMPLE project.

  o a clear distinction is made between type relations and other semantic relations such as roles and parts, which are dealt with

---

\textsuperscript{10} WordNet is a large lexical database of English (Fellbaum 1998). Words are grouped into ‘synsets’, sets of synonyms, each synset expressing a distinct concept. These synsets are interlinked by means of conceptual-semantic and lexical relations, especially hierarchical relations like hyponymy and hypernymy. The original English WordNet has inspired dozens of similar wordnets in other languages. Their aim is to organize words on the basis of their senses, but they do not provide a full definition of these senses (leaving the meaning of words implicit). Yet, this is precisely what we try to achieve via the SGs.
in terms of features in the type template of a semantic class. An SG link has been introduced to link SGs, such that an SG network can be formed. For instance, the SG for *renpaard* ‘race horse’ has a link to the (parent) SG of *paard* ‘horse’, and the SG of *actrice* ‘female actor’ has a link to the (related) SG of *acteur* ‘(male) actor’. It should be noted though that inheritance of information from one SG to another is not yet supported.

- Each main class has its own type template with class-specific features (which results from adding together all the analysed features from the definitions for all the words belonging to that particular class). Figure 2 shows the type template for nouns denoting animal names. The ANW type templates provide a semantic feature description far more fine-grained than most existing systems.

Perception

- Size: can be …. [long, tall];
- (Sound): makes a characteristic sound …. 
- (Smell): spreads a …. smell
- Colour: has a …. [colour, fur, feathers]
- Appearance: is….; has ……

Composition

- (Build) (maybe also under Appearance or Part)
- (Part): consists of ….; has…..

Whole

- (Breed): belongs to ….breed

Function or Goal

- (Function): used as [pack animal]

There are some similarities between the ANW SG approach and the Danish Wordnet (Pedersen and Braasch, 2009: 163). In DanNet, each ontological type evokes a specific template with a particular set of relations. For instance, for the ontological type Human+Part, the has_holo_member relation is obligatory. This is comparable to the ANW SGs, where each semantic class evokes a specific type template. The distinctions made in the resulting templates are, however, less fine-grained in DanNet than in the ANW SGs.

http://semanticweb.org/wiki/OntoClean_Community_Portal
The result is a principled way to represent the knowledge typically associated with a word. This formed the starting point for using SGs in the ANW.

3. The Semagram in the ANW

Since its introduction into the ANW dictionary, 25871 SGs have been compiled\(^\text{13}\). These SGs contain information, which to our knowledge, is not

\footnotesize
\(^{12}\) Content in between square brackets illustrates possible values; features in between round brackets are optional.
\(^{13}\) as of April 2015
available in a systematic way in any other resource, at least not for Dutch. This is especially true for the information in the feature slots. The feature classes form another asset. Using the information contained in the feature class field, allows the retrieval of sets of words adhering to one particular classification perspective within a semantic class. For instance, selecting the semantic class ‘vessel’ and the feature class FUNCTION yields words such as containerschip ‘container ship’, oorlogsschip ‘war ship’, woonboot ‘house boat’, etc. In all these words the functional aspect of the vessel is highlighted.

The SGs are thus a valuable resource. This being said, despite our efforts to prepare and test carefully (see section 2) the integration of SGs in the ANW dictionary, a number of methodological issues arose in practice. Let us see how well SGs perform at a practical level in the context of the ANW.

3.1 Number of features

As noted, the features in the ANW structure are finely split to allow for powerful queries. However, the number of features distinguished in the type templates turns out to be too big for a practical application. In total, 166 features are distinguished. Moreover, certain features are so similar that a particular value can be included in several places which can lead to inconsistencies.

For the SG of cow\textsuperscript{14} this seems to hold for the features FUNCTION and PRODUCT, FUNCTION and USE, PROPERTY and BEHAVIOUR, PROPERTY and VALUE, as well as for PLACE and HABITAT. Inconsistencies lurk, especially as the completion of certain feature slots for a particular semantic class is not enforced by the dictionary writing system. The lexicographers have to consult a manual to decide which features are obligatory for a specific semantic class.

Two possibilities have been considered to remedy this problem (Heyvaert personal communication). Either a complete revision of the 166 features is carried out or a more modest revision based on feature-clustering is performed. The first possibility is in the long run probably the better

\textsuperscript{14} http://anw.inl.nl/article/koe
solution. The second one is the more realistic in the short term, as it is computationally much easier to realise for an existing database.

3.2 Modular editing

At the moment editing is done per semantic class in the ANW. All entries in the same category are produced simultaneously. The semantic classification and the corresponding type template are refined before a new semantic class is started. As a consequence, polysemy becomes an issue, as different senses tend to belong to different categories (mouse: animal vs. device). This holds also for categories whose semantic classification and definition of type templates have not yet been revised. As Heyvaert (2014: 30) points out, there are two ways to go about this. One can describe only the sense that fits the semantic class which is being edited. This would yield an enormous amount of unfinished entries. As an alternative, one can provisionally describe the other senses and revise them once the class is being edited. We have adopted the second strategy, which means we will have to check and revise the information in the category slots at a later stage to remove inconsistencies. For instance, we have noted, that the completed SGs for nouns contain categories initially not foreseen by the semantic classification for nouns. An example of this is the category CIJFER (numeral) in the entry for *logaritme* ‘logarithm’ instead of the predefined category EENHEID (unit).

3.3 Encyclopaedic versus linguistic information

As pointed out by Moerdijk (2007; 2008), inserting SGs in the semantic dimension of an electronic dictionary leads to a much richer semantic description. SGs make explicit knowledge which otherwise remains implicit. Also, more (encyclopaedic) knowledge is encoded than in traditional definition formats. While the lexicographer does not need to worry too much about the split between semantics and encyclopaedia anymore, he still has to decide what is relevant and appropriate for inclusion and what is not and can be safely ignored. Peculiarities which are too detailed or too encyclopaedic and subjective do not belong in the
SG, as SGs should capture only our knowledge and our view of the world as it is reflected in language. The inclusion of scientific information about kingdom, phylum, class, order etc. for animal names (similar to Wikipedia), seems therefore debatable.

3.4 Prototypical and referential information

Moerdijk (2007; 2008) intended the SG in the first place as a vehicle for the description of the stereotyped knowledge speakers have about the concepts with which the words referring to them are associated. That is, in the description of the meaning of a word, not only necessary and sufficient distinctive features matter, but also those features which are (proto)typical and not necessarily distinctive. However, as pointed out by Heyvaert (2014: 31,32) SGs can also be used to describe the referential use of a word, as he illustrates with the SG for *hond* ‘dog’\(^{15}\). The feature slots for SIZE and APPEARANCE show the referential potential of the word, whereas the feature slot for PROPERTY lists the (proto)typical properties of the word.

\[\text{Afmeting: ‘size’}\] kan, naargelang van het ras, zeer klein zijn, zoals een chihuahua tot vrij groot, zoals een Duitse dog
‘can, depending on its race, be very small, such as a chihuahua or quite big, such as a Great Dane’

\[\text{Uiterlijk: ‘appearance’}\] heeft een vacht die verschillende kleuren kan hebben, en die effen of gevlekt kan zijn; kan langharig, kortharig of ruwharig zijn; heeft scherpe tanden met meestal uitstekende hoektanden; heeft een staart die bij sommige soorten kort en bij andere langer is, die soms wel en soms niet sterk behaard is, en die vaak opgericht is en kwispelt
‘has a fur which can have various colours, and which can be plain or spotted; can be long-haired, short-haired or shaggy; has sharp teeth with prominent canines; has a tail which in some varieties is short and in others longer, sometimes hairy, and which is often upright and wagging’

\[\text{Onstoffelijke eigenschap algemeen: ‘inmaterial general property’}\]
is aanhankelijk, onderdanig, trouw en gehoorzaam aan zijn menselijke

\(^{15}\) [http://anw.inl.nl/article/hond](http://anw.inl.nl/article/hond)
meester; is waakzaam; kan agressief, bedreigend en onbetrouwbaar zijn tegenover vreemden
‘is affectionate, subservient, loyal and obedient to its human owner; is vigilant; can be agressive, threatening and shady to strangers’

In order to distinguish both types of information, Heyvaert suggests to use ‘hedges’ such as in some cases or typically to indicate the status of the feature. This is comparable to the notion of weights used in computational linguistics. The possibility of adding weights to the different feature values has not been explored yet. We note though that the inclusion of referential information makes inheritance of information between SGs more complicated.

3.5 Multiple classifications

It is well-known that classification in natural language is not an exact science. There are fuzzy boundaries (e.g. are tomatoes fruit or vegetables?) and multiple classifications occur regularly. For instance, words referring to a person can sometimes also be used to refer to an organisation. There are also cases where they can refer to both at the same time, as illustrated by Heyvaert (2014: 33) with the following sentence All car dealers will be visited by a tax official, where the word car dealer refers to persons as well as companies. Heyvaert suggests that next to the categories person and organisation one should assume a third one: something like person or entity represented as such. We do not think that it is necessary to add a new category for each instance of double classification, but we feel that multiple classifications should be allowed, somehow akin to DanNet (Pedersen and Braasch 2009) where synsets referring to humans can belong to different ontological types, e.g. Human+Object, Human+Object+Group, etc.

3.6 Searching the dictionary

One of the reasons to include SGs in the dictionary, is that they enhance search and query facilities. This is particularly true for searches from content to form. Onomasiological searches in electronic dictionaries
have not been very successful so far, mostly because they are confined to definitions only. Yet, going from a definition to a word can only succeed if the words of the user coincide (more or less) with the words in the definition, which is seldom the case (Moerdijk 2002). As pointed out by Sierra (2000) the ideal onomasiological search must allow users to find the word they are looking for via any related word and in any order. Therefore, the approach described in Zock et al. (2010), in which a dictionary is enhanced with an index based on associations (which are automatically derived from corpora), is more promising for onomasiological queries. Similarly, the information in the SGs plays an essential role in the success of this type of queries in the ANW. As the SG contains more features of the target word than the definition, the chances of finding the target word increase.

Moerdijk et al. (2008) illustrate this for the word *barzoi* ‘borzoi’ where the user is looking for a particular dog and gives values for the most prominent features for the animal he is thinking of e.g. BEHAVIOUR: intelligent; CLASS: grey-hound; PLACE: Russia; BUILD: graceful; APPEARANCE: long-haired. In the latest release of the ANW dictionary, this type of search is called *Describe a word* (‘beschrijf een woord’) and allows the user to search for a word by giving a definition, a description, a paraphrase or by summing up synonyms or other words that he/she can associate with the word he is looking for.\(^{16}\)

The input from the user is then compared to the data in the dictionary database (SGs, definitions, lexical relations), and the results are presented in a list, ordered by relevance.

SGs also enrich the query options in searches from Words with certain features (‘woorden met bepaalde kenmerken’). This option enables users to retrieve words sharing one or several features within the main dimensions of the ANW, e.g. orthography, pronunciation, morphology, pragmatics,

\(^{16}\) In previous releases, this type of search was complemented by a guided search (See Moerdijk et al. 2008), where the user could choose a category from a menu and would then be asked to fill in the values for the most prominent feature slots of that category. Our log files showed that the guided search was not used very frequently, but that users tended to type in all search terms in the first search box that was offered, similar to searching with a search engine like Google.
combinatorics, idioms, etymology and semantics. This search option has been completely revised in the latest release to make it more user-friendly (cf. Tiberius et al. 2014). In the current release, the user is presented with a list of predefined queries rather than having to select queries from a complex feature tree. This list is a subset of all possible queries. It is meant to cover the most frequently occurring queries and to give the user an impression of the types of information included in the dictionary. It is not meant to cover all possible queries. For the advanced user, a link to the query language FunQy\textsuperscript{17} is provided. Using this relatively simple query language, users can create any query they like. They can search, for instance, for all the words belonging to a certain semantic class, for all the words sharing one or several features in the SG in combination with any other information category in the dictionary such as number of syllables, part of speech, etc.

3.8 Future improvements

As the ANW is a research project, as well as a regular dictionary of contemporary Dutch, the need to produce an acceptable amount of dictionary articles each year sometimes collides with the time for research. New ideas and elements can only be implemented if it is possible to update the already edited dictionary articles relatively easy. Based on our experience with the ANW, we would like to arrange some activities differently in future projects. Regarding SGs in particular, it is all right

\textsuperscript{17} http://anw.inl.nl/funqy
to preserve the modular approach of editing per semantic class, but it is advisable within these classes to focus on simplex keywords first. Once the SGs for these words are finished, it will be easier to systematically edit the derivations, compounds, synonyms and other related words of these simplex words and to use the linked simplex word SGs to them for showing the similarities and differences. Furthermore, to enforce consistency and to assist work of the lexicographers, it is necessary to integrate the semantic classification and the type templates in the dictionary writing system.

To further enrich the data in the SGs, it would also be interesting to explore the possibilities to link the SG data to other datasets, in particular to the Cornetto\textsuperscript{18} database for Dutch.

It would also be interesting to compare the type templates and the feature slots used in the ANW with the results of computational models for the extraction of property-based concept descriptions from corpora, see, for example, the Strudel algorithm (Baroni et al., 2010).

4. Summary

In this paper, we have discussed the benefits of SGs in general and for the ANW dictionary in particular. We have introduced the notion and sketched the theoretical background. The ANW is the first dictionary to include SGs on a large scale. We have shown that SGs provide an increase in search and query facilities. One the one hand, they lead to a much richer and more consistent semantic description in ‘semasiological’ queries. On the other hand, they are particularly well-suited to support ‘onomasiological’ queries by offering a structured way to find words through separate content elements. We have noted a number of methodological issues in the current approach and have suggested ways to remedy them in the future.

Acknowledgements

We would like to thank Fons Moerdijk, Frans Heyvaert and Egbert Beijk

\textsuperscript{18} http://wordpress.let.vupr.nl/cornetto/
for the thought-provoking discussions we have had about the semagrams over the years.

References


Moerdijk, F. 2002. *Het woord als doelwit. (The word as target).* Amsterdam, Amsterdam University Press.


Wiegand, H. 1989. Die lexikographische Definition im allgemeinen einsprachigen


Dictionaries


