Polysemy and questions of lumping or splitting in the construction of Swedish FrameNet

Karin Friberg Heppin
Språkbanken
Department of Swedish
University of Gothenburg
karin.friberg.heppin@svenska.gu.se

Dana Dannélls
Språkbanken
Department of Swedish
University of Gothenburg
dana.dannells@svenska.gu.se

Abstract

When working on a lexical resource, such as Swedish FrameNet (SweFN), assumptions based on linguistic theories are made, and methodological directions based upon them are taken. These directions often need to be revised when not beforehand foreseen problems arise. One assumption that was made already in the early development stages of SweFN was that each lexical entry from the reference lexicon, SALDO, would evoke only one semantic frame in SweFN. If a lexical entry evoked more than one frame, it entailed more than one sense and therefore required a new entry in the lexicon.

As work progressed, this inclination towards splitting, in the perpetual lumpers and splitters discussion (Kilgarriff, 1999), proved to be progressively untenable. This paper will give an account of the problems which were encountered and suggestions for solutions on polysemy issues forcing a discussion on lumping or splitting.

1 Introduction

Regular polysemy may be automatically recognized and disambiguated in a text if sufficient amount of data covering the word senses is provided (Alonso et al., 2013). For English, substantial computational work on automatic sense disambiguation has been done. Recent prominent work was carried out on **frame semantics** linguistic theory, more specifically Berkeley FrameNet (BFN) (Das et al., 2013). The vocabulary, comprising around 11,000 lexical units (LU), in BFN has been derived from annotated corpus sentences rather than from a lexicon. As a result, while less frequent words and word senses are represented, many frequently used word senses may be missing.

Furthermore, although BFN has a huge potential advantage for work on word sense disambiguation, it lacks formal definitions of polysemous behavior of words in frames. While there is, in many cases a straightforward relation between lexical units and semantic frames in BFN, there is no clear methodological approach for how to systematically deal with regular polysemy. Consequently, when building a new frame semantic resource, where BFN structure is taken as the interlingua, some theoretical and methodological approaches have to be considered.

In the construction of Swedish FrameNet, words with multiple semantically related meanings, i.e. polysemous Swedish lexical units, have forced more systematic approach to lumping or splitting of semantic frames and lexical entries.

In this paper we address problems of polysemy in FrameNet-like resources. We present polysemy problems we had to deal with during the construction of the frame semantics resource SweFN. We give an account of the reflections, and suggestions for solutions that have been taken on issues such as ambiguity, potential meaning, and vagueness, each forcing a discussion on lumping or splitting.

2 Swedish FrameNet (SweFN)

Swedish FrameNet has been developed as part of the SweFN++ project (Friberg Heppin and Toporowska Gronostaj, 2014; Borin et al., 2010) where the main objective is building a panchronic lexical macro-resource for use in Swedish language technology. This macro-resource consists of several separate resources with the SALDO lexicon (Borin et al., 2013) as the pivot resource to which all other resources are connected. One such resource is SweFN.

SweFN is a lexical semantic resource which has been constructed in line with Berkeley FrameNet

(Ruppenhofer et al., 2010). The theoretical approach taken is based on frame semantics (Fillmore, 1982) which assumes that all content words in a language are best explained by appealing to the conceptual backgrounds that underlie their meanings. Word senses are described in relation to semantic frames, including the semantic roles of the participants.

We have transferred the conceptual layer of BFN to SweFN and provided one-to-one (in a few cases many-to-one) links to BFN frames. These frames were populated with language specific lexical units (LUs) derived from the lexicon SALDO, which evoke the frame in question, and example sentences from corpora. SweFN differs from BFN in several respects including a number of new frames unique to SweFN, compound analysis and domain information. As far as the methodological approach is concerned, the top-down frame building approach was extended with a bottom-up procedure, having its starting point in the lexicon, taking polysemous words and finding or creating frames for all regular (or systematic) senses (Apresjan, 1974). Disambiguation decisions were based on explicit lexical criteria and corpus-related data, to assure homogeneity and usefulness of the resulting resource.

To demonstrate the patterns of semantic roles, example sentences are added from the KORP corpus collection (Ahlberg et al., 2013). The KORP infrastructure offers a functionality called Word Picture which provides statistical information on lexical collocational features. When we add LUs to SweFN frames Word picture is used to acquire an overview of possible senses of Swedish nouns, verbs, and adjectives.

SALDO, Swedish Associative Thesaurus version 2, (Borin et al., 2013) is a free electronic lexicon resource for modern Swedish written language, containing around 130,000 lexical entries. It has an hierarchal structure where lexical entries are associated to each other through two semantic descriptors: primary and secondary. The primary descriptor is obligatory while the secondary one is optional. The resource can be compared to Princeton WordNet (Fellbaum, 1998) from which it differs in several aspects (Borin and Forsberg, 2009). On the polysemy level, the average degree of highly ambiguous words in SALDO is 4.7%, comparing to 12.4% in WordNet 3.1 (Johansson and Nieto Piña, 2015).

3 Cases of polysemy

According to BFN (Fillmore et al., 2003; Fillmore and Baker, 2009), if a word evokes more than one frame it is represented as different LUs with different senses. This is the background to the original stance of SweFN that each entry of the SALDO lexicon would only evoke one frame. Evoking a new frame entails a different sense and thus constitutes a different LU. In the work on SweFN we have encountered three types of cases where it, at first glance, would seem that a lexicon entry could evoke more than one frame: (1) two frames stand in a hyponomy relation to each other; (2) there is a regular polysemy relation between two frames; (3) the concept categories behind the frames divide the world along different dimensions. In the following we elaborate on each of these cases.

3.1 Hyponymy relation

When there is a hyponomy relation between frames we see two possible solutions: (1a) If a lexicon entry evokes more than one frame which all have a common parent frame, the entry becomes an LU evoking this parent frame. An example of this is the verb *bila* (car.v) 'go by car'. It may evoke both the `Operate_vehicle` and the `Ride_vehicle` frame. However, both these frames are in a hyponym relation to the `Use_vehicle` frame, and thus the LU *bila* is listed in this parent frame thereby evoking also the child frames which, i this case, are related to the parent frame in a Perspectivized relation (see Figure 1). (1b) If instead, a lexical entry evokes only one of several child frames in a hyponym relation to one common parent frame, the entry is listed as an LU in the child frame. In this case the LU may still evoke the parent frame. An example of this situation is found in the child frames `Medical_professionals`, `Member_of_military`, `Performers`, and `Representative`, all inheriting from the parent frame `People_by_vocation`.

3.2 Regular polysemy relation

For regular polysemy relation between two frames, case (2), it is difficult to avoid a certain degree of arbitrariness in decisions of when to lump and when to split, regardless of whether these decisions concern entries in the lexicon or frames in the framenet. Take as an example the relation between the `Food` and the `Animals` frames, and like-
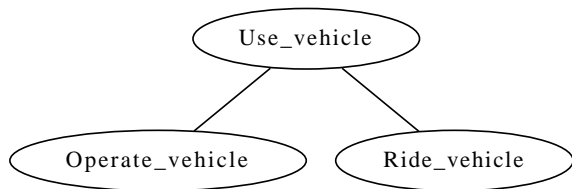
Figure 1: If a lexicon entry could evoke two frames which have a common frame in a hyperonym relation, as in the case with *bila* 'go by car', the entry is listed as an LU in the parent frame. Here `Use_vehicle` is perspectivized in both `Operate_vehicle` and `Ride_vehicle`.
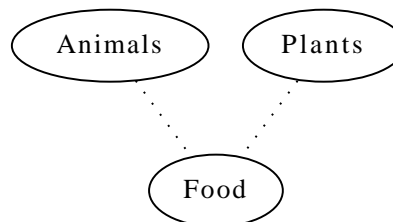

Figure 2: There is a regular polysemy relationship between `Animals` and `Plants` frames and the `Food` frame. All LUs in first frames could, with varying probability, evoke also the latter frame. In these cases they are Guest_LUs in this frame.

wise between the `Food` and the `Plants` frames. What constitutes food is matter from either animals or plants, the names of which become LUs evoking `Animals` or `Plants`. These words denoting animals or plants could also become LUs the `Food` frame, although with substantially varying probability, and in SweFN on the condition that they have separate entries in the SALDO lexicon. The probability that a certain word denoting animals or plants would have a food sense evoking the `Food` frame varies between cultures, circumstances of wellbeing, and what type of creature is doing the eating. In the SALDO lexicon there is only a small number of names of animals, e.g. *fisk* 'fish' and *lamm* 'lamb' with separate entries in the lexicon, for the animal and the food sense. Creating additional entries in the lexicon for additional animals and plants would not solve the problem as the decision on how probable in being consumed as food something would have to be in order to deserve a food sense in the lexicon would always be arbitrary. A solution to this situation, is to let LUs in the more basic frames, in this case `Animals` and `Plants`, appear as Guest_LUs in the other frame, as illustrated in Figure 2. A Guest_LU of a frame does not evoke this frame, and cannot be understood without the senses of the original frame, but may still, under certain circumstances evoke the frame in question. This means that example sentences may be given and annotated in the frame where the LU appears as Guest_LU (Ruppenhofer et al., 2010).

When there is a regular polysemy relation between frames it is not necessary to have more than one entry for a word in the lexicon or more than one LU evoking a frame in the framenet. However, from corpus evidence we learn that some species

of animals and plants are more commonly consumed as food since they are considerably more frequent in the food sense than in other senses. Such evidence could for practical purposes make it meaningful to have additional entries in the lexicon. The entries, in turn, could be listed as LUs in the corresponding frame. For example, a corpus search on the word lax 'salmon' in Korp's Word Picture gives implicit hints for the most frequent senses of the word, as shown in Figure 3. The search resulted in 19,217 instances of *lax* from modern Swedish corpora. Almost all collocates of *lax* belong to the food sense: *färsk* 'fresh', *benfri* (bone-free) 'without bones', *med potatis* 'with potatoes', *i ugn* 'in oven', *äta* 'eat', *innehålla* 'contain', *servera* 'serve', and *laga* 'cook' Some collocates could go with either sense, such as *vara* 'be', *bli* 'become' and *köpa* 'buy'. Only three of the collocates belong exclusively to the animal sense, namely *fiska* 'fish', *fånga* 'catch', *rädda* 'rescue'. Even though results like the one descibed above may motivate additional lexicon entries, decisions of when to do so will always be arbitrary.

Many Swedish verbs show a tendency of construction shift in the object position. As a result, they evoke pairs of frames, for example, `Emptying` and `Removing`, e.g., *tömma* 'empty', *evakuera* 'evacuate', and `Placing` and `Filling`, e.g., *lasta* 'load'. Under the original assumption of SweFN this would entail different senses and consequently different entries in SALDO and listing as different LUs in the two frames. Examples 1 and 2 show such a construction shift which causes a shift of focus from what is being moved (THEME) to the original location (SOURCE). A problem with creating distinct entries in the lexicon is that these verbs frequently are used without

Proceedings of the workshop on Semantic resources and semantic annotation for Natural Language Processing and the Digital Humanities at NODALIDA 2015

14

lax (noun)

Preposition		Pre-modifier		Post-modifier		Lax Verb		Verb lax	
1. med	929	1. färsk	303	1. med potatis	99	1. äta	54	1. äta	474
2. åt	74	2. norsk	173	2. i ugn	82	2. vara	530	2. laga	181
3. av	407	3. benfri	50	3. till middag	76	3. bli	123	3. fiska	112
4. enligt	26	4. god	173	4. i bit	48	4. säga	54	4. köpa²	163
5. på	520	5. rå	44	5. i bit²	48	5. innehålla	34	5. köpa	163
6. för	261	6. vild	48	6. i skiva	39	6. servera	22	6. grilla	62
7. till	227	7. grav	31	7. med sås	32	7. fånga	15	7. älska	84
8. ihop med	8	8. glad	57	8. med färskpotatis	22	8. fånga²	14	8. älska²	84
9. över	25	9. 100g	19	9. med grönsaker	28	9. skära	13	9. fånga²	46
10. efter	54	10. 50g	12	10. med grönsak	28	10. lägga	23	10. fånga	46
11. i form	5	11. ekologisk	15			11. göra	42	11. göra	135
12. utom	5	12. fin	32			12. laga	13	12. servera	45
13. runt	10	13. smarrig	11			13. köpa²	21	13. käka	42
14. efter smak	2	14. kall	18			14. köpa	21	14. kosta	42
15. ovanpå	4	15. 400g	7			15. smaka	11	15. steka	32

Figure 3: A search for the noun *lax* 'salmon' in KORP's Word Picture tool shows that almost all collocates belong to the sense of *lax* evoking the Food frame. Only three, *vild* 'wild', *fånga* 'catch.v', and *fiska* 'fish.v' exclusively collocates with the sense evoking the Animals frame.

object, in which case, a specific sense is not expressed. This is a form of polysemy and the problem may be solved similarly to case (1), described in Section 3.2, by having only one sense in the lexicon, making this an LU evoking the most pertinent frame and letting it be a Guest_LU in the related frame. Polysemy due to construction change applies to many LUs in the concerned pairs of frames, but far from all LUs. Which frame is more pertinent also varies between LUs. This requires a specification on LU level for when the polysemy relation holds, and in which direction.

(1) Olov Lindgren hade redan   evakuerat
     Olof Lindgren had   already evacuated
     [många hyresgäster]THEME när   [...]
     many   tenants   when [...]
     'Olof Lindgren had already evacuated
     many tenants when [...]' (Removing)

(2) [Byggnaden]SOURCE evakueras, [...]
     'building-DEF   evacuate-PASS
     'The building is being evacuated [...]'
     (Emptying)

Other Swedish verbs with tendency to such construction changes evoke, among others, the Removing-Emptying frames, e.g., *tömma* 'empty' and *torka* 'wipe', and the Placing-

Filling frames, e.g. *spreja* 'spray', *lasta* 'load'. A detailed description of corresponding construction changes for English may be found in Levin (2015) in the section on locative alternations.

3.3 Different dimensions

Finally, in the case of dividing the world into concepts along different dimensions, case (3), a solution may be to allow one lexical entry of the lexicon to evoke more than one frame. Consider the Swedish word for children who get one ear ache after the other: *öronbarn* (ear child) 'child that often gets ear aches'. As the sense is about persons being struck by disease, the LU evokes the frame People_by_disease. However, the word is used to denote children and therefore also evokes People_by_age, as in Example 3. This does not entail that there should be more than one entry in the lexicon, as both the age aspect and the disease aspect are evoked at the same time. What happens here is that the People frame is inherited by several frames dividing the concepts describing people along unrelated dimensions, e.g., People_by_age People_by_disease People_by_morality People_by_vocation etc. The consequense is that some lexical entries evoke more than one frame, especially in a language such as Swedish where compounding

Proceedings of the workshop on Semantic resources and semantic annotation for Natural Language Processing and the Digital Humanities at NODALIDA 2015

15

is a very productive linguistic process. The Danish WordNet has also dealt with this problem (Pedersen et al., 2010).

(3) I vår familj har vi öronbarn.
     in our family have we ear-children.
     'Our family's children often get ear aches'.

3.4 Complex relations

More than one of the situations, shown in cases (1)–(3) above, may be applicable for the some entries in the lexicon. A splitting approach, demanding one lexicon entry for each sense possibly evoking a frame, would in such cases result in a large number of lexicon entries, unmotivated from the perspective of how the words are used.

To illustrate this, consider the word *general* 'general'. The SALDO lexicon contains one entry for *general*, which now is listed in the SweFN `Member_of_military` frame. Other frames within the meaning potential would be `People_by_vocation`, `Leadership`, and `Appellations` (titles of individuals, often used together with the person's surname, e.g., *General Abas Khan*).

The current relations between frames in FrameNet show that `Member_of_military` inherits from `People_by_vocation`, a hyponomy relation described in case (1). The `Appellations` frame has a regular polysemy relation with `People_by_vocation` where all LUs in `People_by_vocation` could potentially evoke also the `Appellations` frame, regular polysemy relations described in case (2). At the same time the `Leadership` frame describes people along a different dimension than `People_by_vocation`, case (3). Being a leader may be inherent in being a general and a set of other vocations, but one does not need to have a profession or be in the military in order to be a leader. Neither is the case that all vocations or roles in the military involves being a leader. The sets of LUs evoking `Leadership` and `Member_of_military` or `People_by_vocation` are overlapping.

Summing it up, the SALDO entry *general* has several potential meanings which evoke the four frames `Member_of_military`, `People_by_vocation`, `Leadership`, and `Appellations`. Following the discussion above, the same lexical entry *general* would be listed in the `Member_of_military` and `Leadership`

frames.   As `Member_of_military` inherits from `People_by_vocation` *general* would also evoke `People_by_vocation`, and as there is a regular polysemy relation between this frame and `Appellations`, it would also evoke the latter frame as a Guest_LU. These relations are illustrated in Figure 4.
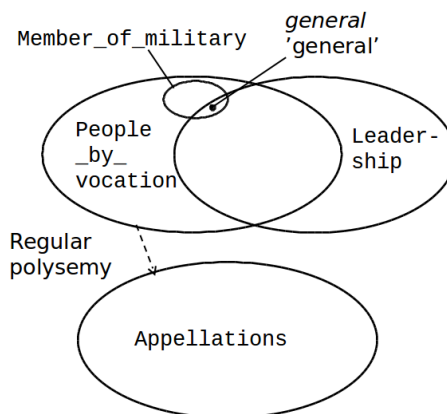

Figure 4: The lexical unit *general* evokes four frames without motivating as many entries in the lexicon. There is a hyponomy relation between the frames `People_by_vocation` and `Member_of_military`, an overlapping aspect of sense between these frames and the `Leadership` frame, while there is a regular polysemy relations between `People_by_vocation` (including `Member_of_military`) and the `Appellations` frames.

The cases described here in Section 3 show that the possibility of a lexicon entry evoking more than one frame does not always motivate adding a new sense to the lexicon or a regular LU to the framenet. In the current version of SweFN the lexical entries of SALDO are still only allowed to populate one frame. However, it has become obvious that solutions such as Guest_LU, additional parent frames, and allowing a lexicon entry to evoke more than one frame in restricted cases, must be considered.

4 Meaning potentials

The construction of a framenet tends to give bias to the splitting point of view. Work on a particular frame includes the phase of populating it with LUs. Encountering an entry in a lexicon, or a word/phrase in a corpus sentence, it is tempting to list it as an LU in the frame under construction if it in some sense evokes it. However,

the potential of an entity to evoke a frame does not necessarily mean that this is the only frame it may evoke, or that it primarily evokes this frame. Hanks (2013) describes words as having **meaning potentials** in that different senses are activated in different contexts, something which does not entail that the word in question has several distinctive senses. This fuzzyness is not a flaw in language, but a strength, as it makes language dynamic and flexible, useful for describing situations and contexts never encountered before. Neither is it always desirable to be specific.

Even though frames evoked by the word's different meaning potentials may have varying semantic types, without explicit internal relation in, for example, the FrameNet system, many words still need to keep their vagueness and should have the possibility to evoke more than one frame. As stated by Wierzbicka (1984) the aim must sometimes be to be vague:

> An adequate definition of a vague concept must aim at precision in vagueness – it must aim at PRECISELY that level of vagueness which characterises the concept itself.(Wierzbicka, 1984):210

4.1 Diverse meaning potentials

A group of words which is often used underspecified, having several meaning potentials of diverse semantic types, are words denoting institutions/businesses/organizations, including the activities and people within. To illustrate this we can look at how the noun *skola* 'school' (in the education sense) is represented in *Svensk ordbok*, a monolingual Swedish dictionary published by the Swedish Academy (Allén et al., 2009):

- Institution where education is performed

1. with focus on the activities performed within the educational institution
2. with focus on the building where the education is performed
3. with focus on the collective of persons working with/attending educational activities within a certain institution
4. other organization which teaches a particular skill or subject

The noun has one main sense with four subsenses. The different subsenses could be said to

evoke the frames in the list below. The list includes the initial part of the frame description in BFN:

- Main sense: `Institutions` "This frame concerns permanent organizations (the INSTITUTIONS) with a public character, meaning that they are intended to affect the lives of the public at large in a particular DOMAIN."

- Subsense 1: `Education_teaching` "This frame contains words referring to teaching and the participants in teaching."

- Subsense 2: `Buildings` "This frame contains words which name permanent fixed structures forming an enclosure and providing protection from the elements."

- Subsense 3: `Aggregate` "This frame contains nouns denoting AGGREGATES of INDIVIDUALS."

- Subsense 4: `Organization`. "This frame describes intentionally formed human social groups (here termed ORGANIZATIONS) with some definite structure and MEMBERS."

The various meaning potentials for a word are brought forward by the context, often put in focus by different collocates. Searching for collocates, with a tool such as Korp's Word Picture, may help detect senses, in a similar manner as for *lax* in Section 3.4. The collocational statistics for *school* in Word Picture shows that the main sense of the word together with subsenses 1 and 2 dominate.

Below is a list of frames followed by collocates to *skola* found by Word Picture. The frames are the ones which the potential meanings of *skola* evokes together with the collocates respectively:

- `Institutions`: *byta* 'change', *välja* 'choose', *driva* 'operate'

- `Education_teaching`: *kommunal* 'municipal', *vanlig* 'ordinary', *gå* 'attend'

- `Buildings`: *bygga* 'build', *ligga* 'be located', *brinna* 'be on fire'

The word *skola* shows several forms of regular polysemy in that is has several different meaning potentials, and is often used underspecified, including more than one sense. This is seen in Example 4 where the visitor, *Jag* 'I', may be seen as

visiting the persons, the activities, as well as the building of the school itself. Making one entry in the lexicon for each potential, each becoming an LU evoking a different frame, would not catch the possibility of vagueness and the relations between the senses would be lost.

(4) Jag ska besöka en skola i
     I   will visit a school in
     Köpenhamn.
     Copenhagen.
     'I am going to visit a school in Copen-hagen.'

Words with the potential of denoting institu-tions, organizations, businesses, and the people and activities within, often show this type of reg-ular polysemy, although with varying sets of po-tential meaning, and thus varying sets of frames evoked. In order to keep the possibility of vague-ness between the potential meanings of varying semantic types, a system allowing Guest_LUs in the frames evoked by subsenses should be de-veloped. However, as not all LUs in the basic frames, such as `Institutions`, `Businesses`, or `Organizations`, have the same set of subsenses, the Guest_LU relation must be established on the level of LUs, not frames.

The difficulty of choosing suitable frames for LUs denoting institutions, businesses, and organi-zations becomes apparent in the inconsistency in BFN for frames which are evoked by this group of nouns.

- *school* evokes `Locale_by_use`
- *theater* evokes `Buildings`, `Locale_by_use`, and `Fields`
- *bank* evokes `Businesses`
- *church* evokes `Buildings`
- *restaurant* evokes `Locale_by_use`
- *bar* evokes `Buildings`

Although there is a lack of consistency in how frames are split in BFN, BFN offers a possi-ble solution for some cases of underspecification, the non-perspectivalized frame (Ruppenhofer et al., 2010). A frame of this type contains a di-versity of LUs sharing a certain scene as back-ground, but which do not have consistent semantic types. Examples are the `Education_teaching` frame, which is evoked by LUs such as *study*.v, *teach*.v, *training*.n, and *educational*.a and the `Performers_and_roles` frame evoked by, for ex-ample, *act*.v, *star*.n and *part*.n. To obtain consis-tent perspective in each frame, the frames could be split further, but then the possibility to house polysemous words would be lost.

However, the purpose of non-perspectivalized frames in BFN was not to house polysemous words, but is described as being as a time-saving measure (Ruppenhofer et al., 2010). The solution of having non-perspectivalized frames is not opti-mal in that having the definition of frames deter-mined along the dimension of context instead of the dimension of participants and semantic roles, the frame definitions and division of the world are not consistent with each other.

4.2 Related meaning potentials

While some groups of words have diverse mean-ing potentials of a variety of semantic types, others have meaning potentials which are more closely related. Take the example of describing national-ity or residence. There are words such as Canadian and Londoner which may describe persons with origin in a certain place. However, the same word may also describe where a person lives or where they are citizens. The origin of a person may well be different from were he or she resides or is reg-istered. When stating a persons nationality or city it may be an advantage to be vague in this aspect.

In BFN and SweFN there are three frames which may be evoked by words for ori-gin/residence/citizenship: `People_by_origin`, `Residence`, and `People_by_jurisdiction`, which inherit from the `People` frame.[1] Parts of the frame descriptions, from the FrameNet website,[2] are given below.

- `People_by_origin` – This frame contains words for individuals, i.e. humans, with re-spect to their ORIGIN.

- `Residence` – This frame has to do with people (the RESIDENTS) residing in LOCA-TIONS, sometimes with a CO-RESIDENT.

- `People_by_jurisdiction` – This frame contains words for individuals, i.e. humans,

---

[1]The `Residence` frame does not inherit directly from `People`, but stands in a 'Used by' relationship to the `People_by_residence` frame which, inherits from the `People` frame and in BFN contains the three LUs *housemate*, *neighbor*, and *roommate*.

[2]`https://framenet.icsi.berkeley.edu/fndrupal/`

Proceedings of the workshop on Semantic resources and semantic annotation for Natural Language Processing and the Digital Humanities at NODALIDA 2015

18

who are governed by virtue of being registered in a certain JURISDICTION.

Most words denoting people in relation to geographic areas could evoke all of the frames above e.g., *stockholmare* (Stockholmer) 'person from Stockholm'. However, a few evoke only one e.g.: *malmöbo* (Malmö-liver) 'Malmö resident' evoking `Residence`, and *svenskfödd* (Swedish-born) 'born in Sweden' evoking (`People_by_origin`). For most of the words denoting people in relation to geographic areas it is desirable to maintain the possibility of vagueness, letting the context determine which meaning potentials should be realized. This may be solved by creating a new a frame on an intermediate level, inheriting from `People` and itself being inherited by the other three frames, with a name such as `People_by_locale`, for these LUs (see figure 5). The LUs which do not evoke all alternatives, such as *malmöbo* and *svenskfödd* should populate the frames that they do evoke. A solution such as this is a more elaborate example of case (1) described in Section 3.1.
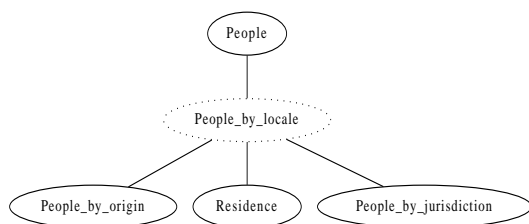

Figure 5: When meaning potentials evoke frames in close relation to each other, vagueness may be maintained by creating a new frame on an intermediate level, a parent frame to the more specific frames.

FrameNet has an intricate network of relations, such as inheritance or 'used by' relations between frames. For example, the frame `People` is inherited by several other frames, most of them with names on the format `People_by_`. A new frame, such as `People_by_locale`, would easily fit in this network having `People` as parent frame and the three frames described above as child frames. There are other cases where frames potentially evoked by an LU do not have connecting relations in the current FrameNet system, and are not as closely related. An example is the verb *bråka* 'fight', which may evoke both `Quarreling` inheriting from `Discussion` and `Hostile_encounter` inheriting

from `Intentionally_act` and itself is inherited by `Fighting_activity`. Solving this, and similar cases which do not lend themselves easily into any case category, could be done, by consulting corpus data to see if any use is more frequent, or by looking at derivational forms related to the words in question. A *fighter*, for example, would more likely be involved in physical fights than quarrels, suggesting that the `Hostile_encounter` frame would be main frame evoked by *fight*, leaving *fight* to be a Guest_LU in `Quarreling`.

5 Summary

There are a number of situations where a lexical entry of the lexicon, here SALDO, evokes more than one frame in the framenet, here SweFN, but where it is still not motivated to split the entry into several polysemous entries. As the relations between the word senses and between the evoked frames differ, different cases must be treated in different ways. This does not necessarily constitute a problem in a resource such as BFN which is not directly linked to a specific lexicon. However, in the case of SweFN, where the original assumption was, and as far as possible still is, that each lexical entry of the SALDO lexicon should only evoke one frame, special account must be taken for entries with several senses potentially evoking different frames. This is especially the case when there is a restriction in that the resource must be compatible with other resources such as SweFN being part of the macro-resource SweFN++.

In cases of hyponymy relations between frames, where all child frames are evoked, it is sufficient to list the LUs in the parent frame. If not all child frames are evoked, the LUs should be listed in the child frames they do evoke. When there is a regular polysemy relation between frames, the lexical entries are listed as LUs in the most basic frame, and as Guest_LUs in the less basic frame. For some pairs of frames, the regular polysemy relation holds for all LUs, while for other frame pairs the relation might only concern a subset of these. This calls for a system of relations in the framenet, not only between frames, but also between LUs in pairs of frames.

Other situations where an LU evokes more than one frame is due to the manner FrameNet resources are constructed: pairs of frames may be overlapping, `Leadership-People_by_vocation` or frames may be non-perspectivalized such as

Proceedings of the workshop on Semantic resources and semantic annotation for Natural Language Processing and the Digital Humanities at NODALIDA 2015

19

Education_teaching which is evoked by LUs of different semantic types within one domain. In these cases, the solution may be to allow, in a restricted manner, one lexicon sense become LU evoking more than one frame.

SweFN has had to let the assumption of one lexical entry – one frame be less restrictive. However, it is still the case that one SALDO entry cannot evoke more than one frame unless some type of relation is established. The exact forms of the relations are still to be decided.

Acknowledgments

References

Malin Ahlberg, Lars Borin, Markus Forsberg, Martin Hammarstedt, Leif-Jöran Olsson, Olof Olsson, Johan Roxendal, and Jonatan Uppström. 2013. Korp and karp a bestiary of language resources: the research infrastructure of språkbanken. In *Proceedings of the 19th Nordic Conference of Computational Linguistics*, pages 429–433. NEALT.

Sture Allén, Daniel Berg, Sture Berg, Martin Gellerstam, Louise Holmer, Ann-Kristin Hult, Susanne Lindstrand, Sven Lövfors, Sven-Göran Malmgren, Christian Sjögreen, Emma Sköldberg, Lennart Tegner, and Maria Toporowska Gronostaj, editors. 2009. *Svensk ordbok utgiven av Svenska Akademien. 1-2.*

Héctor Martínez Alonso, Bolette Sandford Pedersen, and Núria Bel. 2013. Annotation of regular polysemy and underspecification. In *Proceedings from the 51st annual meeting in Association for Computational Linguistics*, pages 725–730. ACL.

J. Apresjan. 1974. Regular polysemy. *Linguistics*, 142:5–32.

Lars Borin and Markus Forsberg. 2009. All in the family: A comparison of SALDO and WordNet. In *Proceedings of the 17th Nordic Conference of Computational Linguistics (NODALIDA 2009)*, NEALT Proceedings Series, Vol. 4 (2009), Odense, Denmark. Kristiina Jokinen and Eckhard Bick.

Lars Borin, Dana Danélls, Markus Forsberg, Dimitrios Kokkinakis, and Maria Toporowska Gronostaj. 2010. The past meets the present in Swedish FrameNet++. In *14th EURALEX International Congress*, pages 269–281, Leeuwarden. EURALEX.

Lars Borin, Markus Forsberg, and Lennart Lönngren. 2013. SALDO: a touch of yin to WordNet's yang. *Language Resources and Evaluation*, 47(4):1191–1211.

Dipanjan Das, Desai Chen, André F. T. Martins, Nathan Schneider, and Noah A. Smith. 2013. Frame-Semantic Parsing. *Computational Linguistics*, 40(1):9–56.

Christiane Fellbaum. 1998. *WordNet: An electronic lexical database*. MIT Press.

Charles J. Fillmore and Collin Baker. 2009. A frames approach to semantic analysis. *The Oxford Handbook of Linguistic Analysis*, pages 313–340.

Charles J. Fillmore, Christopher R. Johnson, and Miriam R.L. Petruck. 2003. Background to FrameNet. *International Journal of Lexicography*, 16(3):235–250.

Charles J. Fillmore, 1982. *Frame semantics*, pages 111–137. Hanshin Publishing Co., Seoul, South Korea.

Karin Friberg Heppin and Maria Toporowska Gronostaj. 2014. Exploiting FrameNet for Swedish: Mismatch? *Constructions and Frames*, 6(1):52–72.

Patrick Hanks. 2013. *Lexical Analysis: Norms and Exploitations*. MIT Press, Cambridge, Mass.

Richard Johansson and Luis Nieto Piña. 2015. Combining relational and distributional knowledge for word sense disambiguation. In *Proceedings of the 20th Nordic Conference of Computational Linguistics*, Vilnius, Lithuania.

Adam Kilgarriff. 1999. I don't believe in word senses. *Computers and the Humanities*, 31(2):91–113.

Beth Levin. 2015. Semantics and pragmatics of argument alternations. *The Annual Review of Linguistics*, 1:63–83.

Bolette S. Pedersen, Sanni Nimb, and Anna Braasch. 2010. Merging specialist taxonomies and folk taxonomies in wordnets – A case study of plants, animals and foods in the Danish WordNet. In *Proc. of the Seventh International Conference on Language Resources and Evaluation*, Valletta, Malta. ELRA.

Josef Ruppenhofer, Michael Ellsworth, R. L. Miriam Petruck, R. Christopher Johnson, and Jan Scheffczy. 2010. *FrameNet II: Extended Theory and Practice*. ICSI, Berkeley.

Anna Wierzbicka. 1984. Cups and mugs: Lexicography and conceptual analysis. *Australian Journal of Linguistics*, 4(2):205–255.

Proceedings of the workshop on Semantic resources and semantic annotation for Natural Language Processing and the Digital Humanities at NODALIDA 2015

20