The Rocky Road towards a Swedish FrameNet – Creating SweFN

Karin Friberg Heppin, Maria Toporowska Gronostaj

Språkbanken, Department of Swedish, University of Gothenburg, Gothenburg, Sweden
karin.friberg.heppin, maria.gronostaj@svenska.gu.se

Abstract

The Swedish FrameNet project, SweFN, is a lexical resource under development, designed to support both humans and different applications within language technology, such as text generation, text understanding and information extraction. SweFN is constructed in line with the Berkeley FrameNet and the project is aiming to make it a free, full-scale, multi-functional lexical resource covering morphological, syntactic, and semantic descriptions of 50,000 entries. Frames populated by lexical units belonging to the general vocabulary dominate in SweFN, but there are also frames from the medical and the art domain. As Swedish is a language with very productive compounding, special attention is paid to semantic relations within the one word compounds which populate the frames. This is of relevance for understanding the meaning of the compounds and for capturing the semantic and syntactic alternations which are brought about in the course of compounding. SweFN is a component within a complex of modern and historical lexicon resources named SweFN++, available at <http://spraakbanken.gu.se/eng/swefn>.

Keywords: FrameNet, Swedish, frame semantics, semantic roles, linguistic resources

1. Introduction

SweFN[1] is an acronym for Swedish FrameNet (the Swedish name of the project is Svenskt FrasNät), a lexical resource under development, designed to support different applications within language technology (Borin et al., 2010). The theoretical approach is based on frame semantics, the brainchild of Charles J. Fillmore (Fillmore et al., 2003; Ruppenhofer et al., 2010). The English version of FrameNet[2], elaborated by the Berkeley research group, provides the guidelines and FN data. It contains more than 10,000 lexical units and more than 1,000 related frames, exemplified in more than 170,000 sentences.

According to FrameNet, a lexical unit (LU) is a pairing of a word or multiword expression with its meaning. Each sense of a polysemous word or multiword expression evokes a different semantic frame, a script-like conceptual structure which describes a particular type of situation, object, or event along with its typical participants. The participants of a frame are described in terms of semantic roles or frame elements.

The SweFN project started with a pilot project in 2009 and turned into a full scale project in 2011. The project aim is to construct a lexical resource with 50,000 entries able to support text generation, text understanding and information extraction. The Swedish FrameNet is found on the SweFN website, and is available as a free resource (CC-BY-SA 3.0, LGPL 3.0). It is part of a larger complex of modern and historical lexicon resources named SweFN++ (Borin et al., 2009).

2. The SweFN

By March 2012, SweFN covered 561 frames comprising more than 18,700 lexical units. Each lexical unit must be gathered from SALDO, a free Swedish electronic association lexicon (Borin, 2010). If a desired lexical unit does not yet exist in SALDO it is proposed as a new entry. Over 1,400 new lexical units have so far been proposed.

The SweFN frames and frame names correspond to the English ones, with some exceptions. We have followed the selection of core frame elements from the Berkeley FrameNet, including their definitions, internal relations, and also non-core elements and their definitions. Thus the meta-information of the frames, such as semantic relations between frames also applies to the Swedish FrameNet.

Compared to the Berkeley FrameNet, SweFN is expanded with information about the domain of the frames, at present: general language, the medical and the art domain. The frames also contain notation about the semantic types from the SIMPLE lexicon (Lenci et al., 2000).

The SweFN frames are presented in tables with following content fields:

Name of the frame: in most cases identical to a corresponding one in the English FN. In these cases the name functions as a link to the English frame where a complete description of the frame is found.

Domain: inclusion of domain information opens for creation of sub-framenets for special vocabularies, e.g. art and medicine.

Semantic type: referring to ontological classification taken from the SIMPLE ontology.

List of core frame elements: in most cases identical to the corresponding ones in the English FN. The name of a core FE is matched with a colour visualising its type.

List of peripheral frame elements: in most cases identical to the corresponding ones in the English FN. The name of a peripheral FE is matched with a colour visualising its type.

Examples: a set of semantically annotated examples from corpus texts. The LU evoking the frame is put in red. FEs are in colours matching the corresponding FE name.

List of instantiated compound patterns: defined by the type of frame element preceding the compound head. The compound heads are lexical units which evoke the frame.

Examples of compounds corresponding to the patterns

---

[1] <http://spraakbanken.gu.se/eng/swefn>
[2] <https://framenet.icsi.berkeley.edu/fndrupal/>

described above: the listed compounds are marked with the border between the constituents, a frame element and an LU.

List of lexical units, from SALDO, populating the frame: the LUs function as links to the SALDO equivalents where information on semantic associative relations and on morphology is found.

List of lexical units populating the frame but without equivalents in SALDO: these LUs are sent to a list of proposals for expansion of SALDO with new lexical units.

Note field: reserved for comments. New or modified frames are provided with explanations.

3. Ensuring integration and information flow in lexical resources

The entries in SALDO do not always match entries of other lexicons, nor the frames in FrameNet. Each lexical unit from the SALDO lexicon can populate no more than one frame. This is often problematic as different aspects of one and the same lexical unit may fit into different frames. The solution is either to propose new SALDO entries, or simply to decide on which of the possible frames should be populated by the existing units.

We will illustrate the difficulties of integrating lexical resources with the Swedish polysemous noun 'brott' which has the following main senses: (1) crime, (2) misdemeanour, (3) fracture, (4) fracture of bone, (5) pause, (6) quarry, and (7) collapsing wave. The current version of SALDO has only three entries for 'brott'. The number of SweFN frames evoked by 'brott' should be at least seven. The senses of 'brott' are listed below, together with corresponding SALDO entries, existing and proposed, and the possible SweFN frames evoked. At present there are only two Swedish frames that are populated with an instance of 'brott', senses (1) and (5). In the other cases there is a missing unit in at least one of the resources.

(1): *Crime*. SALDO: Consistent with SALDO brott..1, associated to *break the law*. SweFN: Evokes the frame Offences and possibly Committing_crime.

(2): *Misdemeanour*. SALDO: No entry, but we propose brott..4, associated to *go against norm*. SweFN: Evokes the frame Compliance.

(3): *Fracture*. SALDO: Consistent with brott..2, associated to *break*. SweFN: There is no existing frame which could be evoked by (3). One solution is to create a new frame under the Fragmentation_scenario related to Breaking_apart and Becoming_separated.

(4): *Fracture of bone*. SALDO: No entry, we propose brott..6, associated to *fracture of bone*. SweFN: Evokes the frame Medical_disorders.

(5): *Pause*. SALDO: Consistent to brott..3, associated to *interrupt*. SweFN: Evokes the frame Process_stop.

(6): *Quarry*. The simplex word 'brott' is rare. It mostly occurs as head of compounds. SALDO: The simplex word has no entry, but brott..5 is proposed, associated to *mining*. There are compounds in SALDO with 'brott' as head, such as dagbrott..1 *opencast*, stenbrott..1 *stone pit*. SweFN: Evokes the frame Mining.

(7): *Collapsing wave*. SALDO: No entry. We propose brott..7, associated to *ocean*. SweFN: Evokes the frame

Natural_features.

4. Interpretation of compounds in terms of frame elements

As Swedish is a strongly compounding language, special attention is paid to the analysis of compound lexical units. Nominal, adjectival and verbal compounds are described in terms of semantic relations where the compound head is a lexical unit evoking the frame in question and the modifier a frame element of that frame.

Below are selected parts of the frame Assistance which illustrate the annotation of frame elements within compounds and correspondingly of the constituents of these compounds used as separate lexical units.

The core and the non-core frame elements are listed with their corresponding tags. The example sentences is a pair where the first sentence contains a frame evoking LU which is a compound with the modifier being a core frame element. The second sentence contains the compound constituents, now as separate units, the modifier a frame element and the head the frame evoking LU.

Frame: Assistance
Core elements: Benefited_party [B], Focal_entity [F], Goal [G], Helper [H]
Non-core elements: Degree [DG], Domain [DO], Duration [DU], Explanation [E], Frequency [F], Instrument [I], Manner [MA], Means [ME], Place [PL], Purpose [PU], Time [T]

Benefited_party+LU fattig|hjälp 'help for the poor'
Focal_entity+LU minnes|stöd 'help for memory'
Goal+LU flyt|hjälp 'help to float'
Helper+LU advokat|hjälp 'help by lawyer'
Instrument+LU dator|stöd 'help with computer'
Manner+LU akut|hjälp 'emergency help'
Means+LU bär|hjälp 'help to carry'

Example sentences:
Här i Trouville skulle du kanske kunna organisera [LU [B fattig]hjälp] [ME med kol och soppa och bra begagnade kläder].
Here in Trouville you maybe might organize [LU [B poor] help] [ME with coal and soup and good used clothing].
Det blir en [DG stor] [LU hjälp] [B till fattiga människor], säger Bachar Ghanoum.
It will be a [DG great] [LU help] [B to poor people]*, says Bachar Ghanoum.*[3]

The aim of SweFN is to collect a sufficient number of annotated sentences of this kind to use in machine learning for automatic compound disambiguation, by determining the frame element of the compound modifier. See further the frame Assistance at the end of the article.

The examples illustrating Swedish frames attempt to follow the crieria put forward for good examples by Kilgarriff et al. (Kilgarriff et al., 2008), being further discussed by

---

[3]These and other example sentences are literal translations demonstrating the Swedish way of expression.

Atkins and Rundell (Atkins and Rundell, 2008). According to Kilgariff et al.

"A good example must be:

- typical, exhibiting frequent and well-dispersed patterns of usage

- informative, helping to elucidate the definition

- intelligible to learners, avoiding gratuitously difficult lexis and structures, puzzling or distracting names, anaphoric references or other deictics which cannot be understood without access to the wider context. We call this its 'readability'." (Kilgarriff et al., 2008)

The above requirements have been translated into practical and measurable features to be used by the GDEX tool and defined in terms of sentence length, word frequenceies, key word position in a sentence and preferences for selecting whole sentences.

It is also known that good examples have to be tuned to the type of the lexical resource under construction as well as the needs and expectations of its users. Thus, the two resources we have been working with, namely SweFN and SALDO, might show somewhat different preferences for example selection. In the case of SweFN, we have attempt to find examples which capture all core elements in an exhaustive way, in case of SALDO the compatibily with the evoked sense was in focus. Thus the examples of hte SALDO lexicon can be thought as a complement to the set of examples chosen for SweFN.

5. Content expansion

There are two levels being directly involved in the content expansion, namely LU level, where the aim is 50,000 lexical units, and frame level where the aim is to cover all frames of the Berkeley FrameNet populated by lexical units, as well as frames specific to Swedish. The expansion of LU level can be partially done by filling the existing frames with adequate lexical units. Concerning nouns specific frames listing catecories of artifacts, people, plants and anmials have been and will be crated. The frames focusing on verbs are more fine-grained and less populated. Also for verbs a number of new frames need to be created.

At present there are 31 frames in SweFN which do not have a match in the Berkeley FrameNet. They are described and listed below. There are eight completely new frames:

Animals
Entity_specific_modes_of_being
Falling_ill
Furniture
Inner_parts_of_body
Languages
People_by_disease
Plants

A frame populated by lexical units for medicines and similar substances has been separated from the frame Active_substance for an additional, more specific frame. In a similar way a frame populated by lexical units for contraction has been separated from Expansion and a frame for causing contraction from Cause_expansion. The names of the modified frames are expanded with the _mod suffix. The changes described resulted in the following frames:

Active_substance_mod
Active_substance_medical
Expansion_mod
Contraction
Cause_expansion_mod
Cause_contraction

The frame Observable_body_parts has had a slight change of spelling:

Observable_bodyparts

There are a number of frames that have been divided in two or more in order to get more specific frames, Cause_change_position_on_a_scale, Change_position_on_a_scale, Expertise, Medical_conditions, Noise_makers, People_by_morality, and Stimulus_focus.

Cause_change_position_on_a_scale_decrease
Cause_change_position_on_a_scale_fluctuation
Cause_change_position_on_a_scale_increase
Change_position_on_a_scale_decrease
Change_position_on_a_scale_fluctuation
Change_position_on_a_scale_increase
Expertise_negative
Expertise_positive
Health_status
Medical_disorders
Musical_instruments
Sound_makers
People_by_morality_negative
People_by_morality_positive
Stimulus_focus_negative
Stimulus_focus_positive

There are also cultural differences that need to be considered. As there are no juries in Sweden we decided on a broader name, Deliberation, for the frame Jury_deliberation. This makes it possible to describe different court systems, with or without a jury.

6. Concluding remarks

Using the Berkley FrameNet as a prototype for constructing a FrameNet describing another language speeds up the process. Having basically the same frames and names of frames also opens possibilities for cross-language applications, such as machine translation or language learning.

However, it is necessary to be observant of situations where the languages or language cultures are not compatible.

As described above, in the case of Jury_deliberation, cultural differences can make certain frame names unsuitable. There are also cases when concepts are not expressed in the same manner in different languages. The English frame Indicating contains a single lexical unit: *name* (verb). There are Swedish equivalents corresponding to *list*, but these can only be used to name more than one entity. Another near equivalent is *namnge*, but this can only be used when asking specifically for a name, as in the following situation[4]:

> *No doctor can name the disease that's killing him.*
>
> *The accused, who can not be named for legal reasons, [...]*

In other situations, such as the one below:

> *You have to name our destination.*
>
> *Name the novel in which this festive illustration can be found.*

Swedish uses different types of expressions, for example:

> *You must say what is our destination.*
>
> *In which novel is this festive illustration found?* or *Say the name of the novel in which this festive illustration can be found.*

In such cases it must be decided if a new entry, here for the verb 'säga' *say*, should be created in the SALDO lexicon. Would this be motivated? Or would such an entry only be the result of trying to squeeze Swedish into the same form as English?

Finally, it should be noted that all of the lexical resources used for constructing SweFN are freely available for downloading. Furthermore, the reuse of lexical data elaborated in within EU projects, like SIMPLE and Parole, not only enriches the final resource and makes the process of creating the resource more efficient, but at the same time it poses a challenge for integration.

Acknowledgments

7. References

B.T. Sue Atkins and Michael Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford University Press.

Lars Borin, Dana Dannélls, Markus Forsberg, Maria Toporowska Gronostaj, and Dimitrios Kokkinakis. 2009. Thinking green: Toward Swedish FrameNet++.

Lars Borin, Dana Dannélls, Markus Forsberg, Maria Toporowska Gronostaj, and Dimitrios Kokkinakis. 2010. The past meets the present in Swedish FrameNet++.

Lars Borin. 2010. Med Zipf mot framtiden - en integrerad lexikonresurs för svensk språkteknologi. *LexicoNordica*, 17:35–54.

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

Adam Kilgarriff, Miloš Husák, Katy McAdam, Michael Rundell, and Pavel Rychlý. 2008. GDEX: Automatically finding good dictionary examples in a corpus. In *Proceedings of the XIII EURALEX International Congress*.

A. Lenci, N. Bel, F. Busa, N. Calzolari, E. Gola, and M. Monachini. 2000. SIMPLE: A general framework for the development of multilingual lexicons. *Lexicography*, 13(3).

J. Ruppenhofer, M. Ellsworth, M.R.L.Petruck, C.R. Johnson, and J. Scheffczyk. 2010. *FrameNet II: Extended theory and practice*. Morgan Kaufman Publishers, <https://framenet2.icsi.berkeley.edu/docs/r1.5/book.pdf>.

---

[4]The examples are from the Berkley FrameNet.

Assistance

ram	Assistance
domän	Gen
semantisk typ	Purpose_act
kärnelement	Benefited_party, Focal_entity, Goal, Helper
periferielement	Degree, Domain, Duration, Explanation, Frequency, Instrument, Manner, Means, Place, Purpose, Time
exempel	Majblommans verksamhet bygger på att barn hjälper barn genom att sälja majblommor. Föreningen Forum för mor- och farföräldrar hjälper dem som drabbats. Kamraterna bistår med handräckning av verktyg. Rygg- och nackkliniken bistår också med ergonomisk information. Du ska stödja mig i allt jag gör. Intensiv medietäckning hjälpte till att sprida budskapen och ge en ökad medvetenhet om problemet. Då hjälper vi till med information och profylax på förskolan. Han hade gett sitt liv för att tjäna sin husbonde och för att nå sina egna drömmars fulländning. Jag är för ett generöst bistånd till u-länder. Med hjälp av snabbguiden här får du bättre koll på livet. Föräldrarna har advokater till sin hjälp, men ingen för huvudpersonens, barnets, talan. En advokat kan också hjälpa till att dela upp ett arv. Ett målsägandebiträde är en advokat som hjälper brottsoffret med juridiska frågor. Två officerare har med advokathjälp försökt få armén att häva avskedandena. Jag är tvungen att med advokathjälp försöka få pengar från honom, så här långt har jag inte fått någonting. Och man kunde enkelt utöka brottsoffrens möjlighet att förses med advokathjälp. När vi hade betalat allt fick vi bärhjälp av de snälla killarna som var anställda på IKEA för att få ner allt till parkeringen en våning under. Peter Centerwall skriver att han inte tilllåtits handla på systembolaget när hans 16-åriga dotter varit med som bärhjälp. En kvinna erbjuds bärhjälp av en okänd man efter det att hon handlat på Lidl i Frölunda. Träffade på ett par kändisar, varav en behövde bärhjälp med sina två flak med öl. En mamma frågar mig om jag kan hjälpa henne bära stubbar. Jag fick hjälpa S bära upp plattorna från lagret till bilen, från bilen upp till lägenheten - sju stycken, sju gånger, två trappor upp. Tur att vi har fått hjälp att bära! Här i Trouville skulle du kanske kunna organisera fattighjälp med kol och soppa och bra begagnade kläder. Så vi måste hjälpa de fattiga att få det bättre - därför att det är vad Anna skulle ha velat. Det blir en stor hjälp till fattiga människor, säger Bachar Ghanoum.

Figure 1: The frame Assistance. The fields shown in this figure are: Name of the frame, Domain, Semantic type, Core elements, Periferal elements, and Example sentences. The frame is continued in figure 2.

	Det finns inget **datorstöd för utredare**. Inom båda nätverken används **datorstöd i läs- och skrivträningen**. Pedagogiska institutionen på Göteborgs universitet skall senare göra en utvärdering som skall klargöra ifall inlärning **med stöd** av dator förändrar elevens sätt att lära, den kognitiva processen och begreppsbildningen. En dagbok fungerar som **minnesstöd**. Även om han har sjökortet med som **stöd för minnet** när han seglar sin egen båt. Hon plockade fram klasslistan som **stöd för minnet**. **Hajar** får dock **lite flythjälp** av stora mängder olja som de förvarar i levern. Fast det är svårare att drunkna i kvicksand än i vanligt vatten eftersom sanden **hjälper dig** att flyta. Men ofta är **uppsökande psykiatrisk akuthjälp** alldeles nödvändig. På så sätt kan fler **kvinnor snabbare** få **akuthjälp på Kvinnohuset då överslussning kan ske till mellanboende**. Vid en kärlekskris kan **en eller bägge parter** behöva **hjälp akut**.
sms	Benefited_party+LU, Focal_entity+LU, Helper+LU, Goal+LU, Instrument+LU, Manner+LU, Means+LU
sms-exempel	Benefited_party+LU_EX_fattig|hjälp, armé|understöd Focal_entity+LU_EX_minnes|stöd, eld|understöd Helper+LU_EX_advokat|hjälp, läkar|hjälp, psykolog|hjälp Goal+LU_EX_flyt|hjälp, lots|hjälp Instrument+LU_EX_dator|stöd Manner+LU_EX_akut|hjälp, med|hjälp, själv|hjälp Means+LU_EX_bär|hjälp, drag|hjälp
saldo	**vb:** assistera..1 avlasta..1 befrämja..1 bispringa..1 bisträcka..1 bistå..1 biträda..1 båta..1 handleda..1 hjälpa..1 rycka_in..1 stödja..2 tjäna..1 undsätta..1 vårda..3 **vbm:** gå_kjortelvägen..1 hjälpa_på_traven..1 hjälpas_åt..1 ställa_upp..3 **nn:** advokathjälp..1 akuthjälp..1 arméunderstöd..1 assistans..1 assisterande..1 assistering..1 bistånd..1 bärhjälp..1 datorstöd..1 draghjälp..1 eldunderstöd..1 fattighjälp..1 flythjälp..1 handräckning..2 handtag..2 hjälp..1 hjälpande..1 hjälpsamhet..1 krishjälp..1 lotshjälp..1 läkarhjälp..1 medhjälp..1 minnesstöd..1 nödhjälp..1 psykologhjälp..1 service..1 självhjälp..1 stöd..2 tjänande..1 tjänst..3 understöd..1 undsättande..1 undsättning..1 undsättningsförsök..1 **nnm:** hjälp_på_traven..1 **av:** hjälpsam..1
saldo (nya)	**vb:** stötta..3 **vbm:** hjälpa_till..1
kommentar	[T Då] [LU hjälper [H vi] till] [F [F med information] och [F profylax]] [PL på förskolan]. ;;In the compound 'undsättningsförsök' (rescue attempt) the grammatical head is more or less empty. It is the modifier which carries most of the meaning.
skapad av	KFH
skapad	2011-08-31
modifierad	2011-10-12

Figure 2: The continuation of frame Assistance (from figure 1). The fields shown in this figure are: Example sentences (continuation), Instantiated compound patterns, Compound examples, Lexical units with links to the SALDO lexicon, Suggestions for new entries in SALDO, Notes, Creator, Date of creation, and Date of modification.