



UNIVERSITY OF GOTHENBURG
DEPT OF SWEDISH

CLT

Språk
BANKEN



Swedish FrameNet++ (SweFN++)

DIMITRIOS KOKKINAKIS,

...on behalf of the Swe-FN++ team

Department of Swedish, Språkbanken
University of Gothenburg, Sweden
dimitrios.kokkinakis@svenska.gu.se





Overview of SweFN

SweFN++ full-scale lexical resource designed to support Swedish language technology applications. Its goal has been a lexical macro-resource for use as an infrastructural component in Swedish language technology research and in the development of NLP applications and annotated corpora for Swedish.

Objectives:

1. link (reuse, enhance, harmonize) a number of existing free lexical resources into an integrated lexical macro-resource
2. create a full-scale Swedish FN integrated into macro-resource
3. develop methodologies making maximal use of language technology tools and text corpora to minimize human effort required for accomplishing (1) and (2)
4. make all resources and tools developed freely available under open-content/open-source licenses





Transferring FN to SweFN

SweFN follows the Berkeley FrameNet concerning:

- the names of the frames
- definitions of the core frame elements
- definitions of non-core frame elements
- the semantic relations between the frames.

But:

- we have developed our own software, *Karp*
- we have our own interface
- example sentences are picked from the corpus infrastructure *Korp*
- we only annotate for semantic roles/frame elements
- syntactic annotation is available through the Korp corpus infrastructure
- we analyze compounds internally



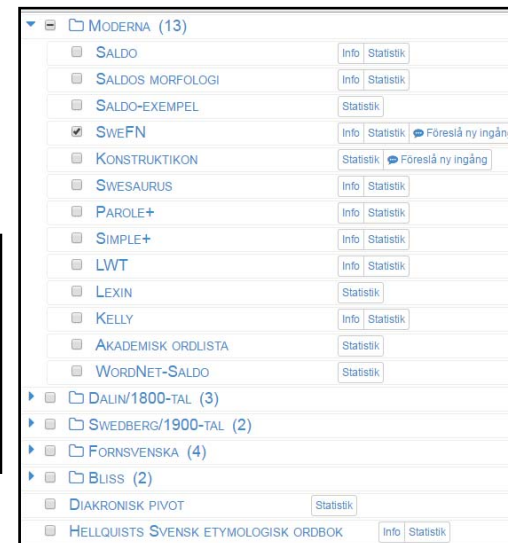
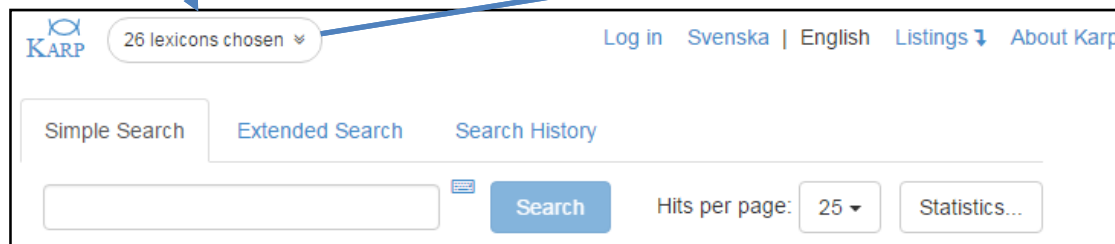


Swedish FrameNet

- Swedish FrameNet (SweFN) was developed in a project that the Swedish Research Council funded, 2010–2014 (SweFN++)
- SweFN lexicon is available under a CC-BY license
 - to download (in LMF): <<http://spraakbanken.gu.se/eng/resource/swefn>>
 - to search the lexicon use KARP (open lexical infrastructure):

<<http://spraakbanken.gu.se/karp>>

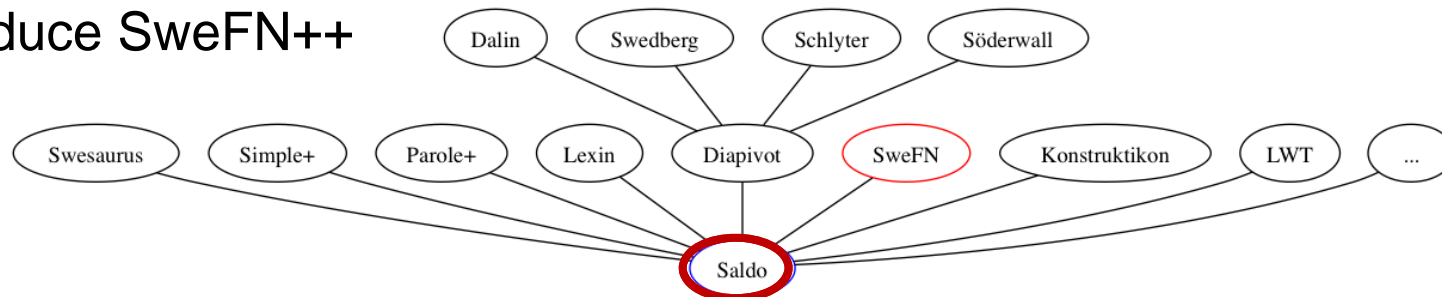
...and then select SweFN





Swedish FrameNet

- SweFN lives in a wider ecosystem of lexical resources, whose backbone is the SALDO lexicon (builds on Swedish Associative Thesaurus – hierarchical structure), to which all lexical resources are connected; several freely available resources were used to produce SweFN++



- all resources are integrated into the KARP system (26 lexical resources; >700000 entries) <<http://spraakbanken.gu.se/karp>> for managing and searching – open lexical infrastructure
- Swedish FN has tried to reuse as many of ICSI frames as possible; greater effort on LUs





Swedish FrameNet

Lexical frames 1,194 (*BFN v1.5: 1,033*)

LUs 38,700 (*BFN v1.5: 12,714*)

Example sentences 9,006 (*BFN v1.5: 195,590*)

New LU suggestions 2,818

Multilinguality and FrameNet duality: language independent frames with language dependent content; concepts are mostly language independent and contain SRs, or FEs, which are also mostly language independent. LUs and annotated sentences are language-specific

Arriving	
domän	Gen
kärnelement	Goal Theme
periferielement	Circumstances Cotheme Degree Depictive Event_description Frequency Goal_conditions Manner Means Mode_of_transportation Path Period_of_itations Place Purpose Re_encoding Source Time
arv	Eventive_affecting
exempel	<ul style="list-style-type: none"> - [V]Theme [kom]LU [till Sverige]Goal [som flyktingar efter trakasserier och förföljelse]Depictive [1970]Time . Det [kommer]LU [mer kläder]Theme [från AA]Source [inom kort]Time . [I måndags]Time [anlände]LU [Kris Beech]Theme [till Jönköping]Goal igen. [Strejkbrytarnas]Theme [ankomst]LU [den 13 maj]Time väckte en oerhörd ilska. [Tåget mot Norsborg]Theme [ankommer]LU [om en minut]Time . Bussar till Victoria och Valetta ansluter vid varje [färjeankomst]LU . [Vi]Theme [landade]LU [i LA]Goal [vid lunchtid på tisdagen]Time . Du siktar dock högt, medan [bollen]Theme [smiter in]LU [längs golvet]Path . Snön knorrar under sulorna och [vi]Theme [smiter in]LU [på en krog där vi blir ensamma]Goal . [Dottern]Theme har [landat]LU [i New York]Goal , snart är hon framme i Phoenix. Dottern har landat i New York, [snart]Time [är]COP [hon]Theme [framme]LU [i Phoenix]Goal .
sms	Theme+LU Goal+LU Depictive+LU
sms-exempel	Theme+LU_EX_buss ankomst, båt ankomst, flyg ankomst, färje ankomst, pendeltåg sankomst, tåg ankomst Goal+LU_EX_fram komst, hem kommen Depictive+LU_EX_ensam kommande vb komma ¹ ankomma ¹ anlända ¹ hamna ¹ inlöpa ¹ tillskvynda ¹ anlöpa ¹ nå ¹ ab framme ³ fram ³ nn hemkomst ¹ ankomst ¹ bussankomst ¹ båtankomst ¹ flygankomst ¹ framkomst ¹ färieankomst ¹ pendeltågsankomst ¹ tågankomst ¹ inresa ¹ vbm komma fram ² infinna sig ¹ inställa sig ¹ komma in ¹ komma till ² komma hem ¹ dyka upp ¹ av ensamkommande ¹ hemkommen ¹
lu-förslag	landa ⁴ smita in ¹ i kapp ²



Basic Infrastructure

Backbone of SweFN++ Lexical Macro-structure

- SALDO: associative lexicon
- Karp: open lexical infrastructure
- Korp: corpus search infrastructure

<http://spraakbanken.gu.se/karp/>

MODERNA (13)

- SALDO Info Statistik
- SALDOS MORFOLOGI Info Statistik
- SALDO-EXEMPEL Statistik
- SWEFN Info Statistik Föreslå ny ingång
- KONSTRUKTIKON Statistik Föreslå ny ingång
- SWESAURUS Info Statistik
- PAROLE+ Info Statistik
- SIMPLE+ Info Statistik
- LWT Info Statistik
- LEXIN Statistik
- KELLY Info Statistik
- AKADEMISK ORDLISTA Statistik

KARP 1 lexicon chosen

Log in Svenska | English

Simple Search Extended Search Search History

medical Search Hits per page: 25 Statistics...

Hits 12

Page: 1 / 1

SWEFN + 12 HITS (DISPLAYING 12)

ID	Medical_interaction_scenario
DOMAIN	Med
SWEXCN	
SEMANTIC TYPE	
INHERITANCE	
CORE ELEMENTS	Affliction, Medic, Patient
PERIPHERAL ELEMENTS	Body_system, Medical_centre
EXAMPLES	Politiker - inte patienter - ska bestämma vem som får [bedriva] ^{SUPP} [sjukvård] ^{LU} . Tusentals barn brännskadas varje år så de måste uppsöka [akutvård] ^{LU} . Nadja jobbar i dag i [ett specialteam] ^{Medic} : [på Angereds närsjukhus som] ^{Medical_centre} [ger] ^{SUPP} [vård] ^{LU} [till barn med flyktningbakgrund] ^{patient} .
COMPOUNDS	Affliction+LU sinnessjukvård Medical_centre+ LU Patient+LU mödrarvård, narkomanjvård
COMMENT	I BFN listas bl a. doctor, nurse som LUs både här och i ramen Medical_professionals. I SweFN listas de enbart i ramen [Medical] professionals. Man skulle kunna utöka listan på perifera FE med FE Typ för att analysera sammansättningar som: Intensivvård, akutvård, långvård etc...Tusentals barn brännskadas varje år så de måste uppsöka [LU akutvård]. I den satsen uppsöka kunde taggas som GOV.
LEXICAL UNITS	noun akutvård, folktandvård, företagshälsövård, hemsjukvård, hälsovård, intensivvård, långvård, mödravård, narkomanjvård, nykternetsvård, sinnessjukvård, sjukhusvård, sjukvård, slutenvård, vård ² , öppenvård

KARP 26 lexicons chosen

Log in Svenska | English Listings About

Simple Search Extended Search Search History

Search Hits per page: 25 Statistics...

KORP GP 1994 selected - 21,338 of 9,236 folios

Simple Extended Advanced Compare

Search

also as initial part final part and case-insensitive

hits per page: 25 sort within corpora: not sorted Statistics: compile based on: word Show word picture Show map

Results: 2,440

GP 1994

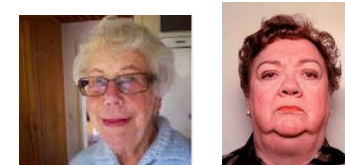
Man hade fått tjänsterevision och nu sparade hon till en ny... men fortfarande hade hon en månads hyra att betala för lägenheten...
Nu åkte en med ett par miljoner, en annan med en ny bil, en tredje med en resa till Kuba...
Ett flertal av de som hade varit i Sverige kom tillbaka till Sverige...
Jag hade svårt att ta mig ut och in i en bil...
Ett år senare hade jag varit i Frankrike i 120 år, äger en bil som inte alltid startar men har en desto mer fristående flyggt som h...
Till tonåringar i en stulen bil präjades en polistid och kraschade in i en annan vid en dramatisk bi...
Det är svårt att ta sig fram med bil...
En man står och tittar på en bil, en annan klipper gräs, men i borgen håller sig de flesta inomhus...
Familjen åkte till Italien - en resa som Olo och hans familj kan drömma om så här i e...
Och ute i landet dimper på till en bil med i soffbänket eller i mat för 30000 kronor...
Man väljer inte längre bara en bil vid endoknigt och viktiga ärenden, till a sjukhusvård...
En bil backade eller körde rakt in i butik och förstörde både skyltarna



Differences

LUs can have different extensions:

mormor: (maternal) grandmother } *Kinship*
 farmor: (paternal) grandmother }



Culture specific LUs... e.g.

surströmming: fermented fish } *Food*



Cultural differences: no juries in Sweden

more general definition fits better than a specific one

Jury_deliberation

Definition:

The Jury discuss the Case and its Charges in order to evaluate the Possible_sentence of the Accused.
 The jury DELIBERATED his guilt.

Deliberation	
domän	Gen
kärnelement	Case Charges Deliberating_group Possible_sentence
periferielement	Duration Place Time
exempel	<ul style="list-style-type: none"> [I morgon]Time väntas [domstolen]Deliberating_group [överlägga]LU [vilken dom som skall fällas över den andre huvudanklagade, den 33-åriga Imam Samudra]Possible_sentence . [Domstolen]Deliberating_group [höll]SUPP en [kort]Duration [överläggning]LU .
lus	vb överlägga ¹ nn överläggande ¹ överläggning ¹
kommentar	New frame that aims at broadening the FN frame Jury_deliberation for cultural reasons. Many countries, including Sweden, do not utilize juries. The deliberation is carried out by a group of persons selected by specific rules, different in different countries.





Differences

Inflection vs. Lexicon:

- English expresses activity in progress with the progressive form (-ing), not via frame.
 - FN is populated with LUs, not morphemes.
- Swedish expresses activity in progress with lexical expressions. SweFN has a frame for these expressions

Activity_in_progress

domän	Gen
kärnelement	Activity Entity Event
periferielement	Circumstances Depictive Duration Event_description Explanation Manner Means Place Purpose Time
exempel	<ul style="list-style-type: none"> ▪ [Gångbron till en färja kollapsade plötsligt]Event när [passagerarna]Entity [höll på]LU [att gå ombord.]Activity ▪ - Så [jag]Entity [håller på]LU [och lär mig]Activity , säger han ödmjukt. ▪ [Det]Event gör att [jag]Entity [i lugn och ro]Manner kan [hålla på]LU [och måla]Activity . ▪ [Vi]Agent [håller på]LU [och bygger om och bygger till]Activity och har verkligen att göra. ▪ [Hedeskolan]Entity [håller på]LU [och växer ut till en F-9-skola]Activity . ▪ [Man]Entity [är]LU [nu]Time [i färd]LU [med att finna lämpliga rutiner]Activity [för dess användning]Purpose . ▪ [Många skolor]Entity [är i färd]LU [med att utarbeta handlingsplaner mot mobbning]Activity .
lu-förslag	hålla på⁴ vara i färd¹
kommentar	Ny ram. These LUs do not establish any distinct semantic relations with a frame element in the subject position, called FE Entity, except for the relation of being involved in activities specified by an infinitive phrase that follows.





Differences

Compounds: Swedish writes compounds without a space between the parts; treating the compound's constituents separately is not appropriate. When SALDO lists the compound, SweFN treats it as a LU. The exact treatment of a compound depends on its transparency, or the level of its compositionality:

transparent: hästhov (horse+hoof) – 'horse hoof'

not transparent: hästhov – 'coltsfoot' (the flower)

Observable body parts	
domän	Med
semantisk typ	Body_part
kärnelement	Possessor
periferielement	Attachment Descriptor Orientational_location Subregion
exempel	<ul style="list-style-type: none"> [Handen]LU är bruten. [Armen]LU måste opereras. [Tusses]Possessor [brutna]Descriptor [arm]LU måste opereras. I synnerhet [tupparnas]Possessor [stora]Descriptor [kammars]LU var utsatta. Någonstans längre bort bakom husknutarna ekade klappret från [snabba]Descriptor [häst]Possessor [hovar]LU]LU ...
sms	Descriptor+LU Subregion+LU Possessor+LU
sms-exempel	Descriptor+LU_EX_ring finger, lill finger, pek finger, lång finger Subregion+LU_EX_under arm, över arm Possessor+LU_EX_häst hov

Plants	
domän	Gen
semantisk typ	Vegetal_entity
kärnelement	Plant
periferielement	Age Descriptor Origin Persistent_characteristics Systematics
exempel	<ul style="list-style-type: none"> [Luktärter]LU och [krasse]LU slingrar ut över grusgångarna. [Agave]LU (Agave americana) är en [suckulent]Persistent_characteristics växtart inom [agavesläktet]Systematics och familjen [agaveväxter]Systematics . [Agave americana]LU växer naturligt i [Mexiko]Origin och [södra USA]Origin . [Havre]LU (Avena sativa) är ett [kraftigt]Descriptor och oftast [högväxt]Descriptor [gräs]LU , inom släktet [havren]Systematics och familjen [gräs]Systematics , med kala strån som kan bli över en meter höga. Då kommer de första värtecknen i dikesrenarna och [hästhovarna]LU lyser som solar .



Some Applications

- Semantic Role Labeling
- Natural Language Generation
 - domain of art
- Search Applications
- Education/Teaching
- Information Extraction
 - medical domain: for medical language research, SweFN created several new frames





More Information <<http://spraakbanken.gu.se/eng/swefn/>>





LISTEN | PÅ SVENSKA | A-Ö



ABOUT US
RESOURCES
RESEARCH
PUBLICATIONS
PHD PROGRAM
STAFF

Home / Research / SweFN++

SweFN++

Search SweFN++

Development version

Documentation

Statistics

Error report

FrameNet Workshop 2013 >

NoDaLiDa 2013 workshop

Publications

Swedish FrameNet++ workshop 2014

SweFN++



[Malin Ahlberg](#), [Lars Borin](#), [Dana Dannélls](#), [Markus Forsberg](#), [Maria Toporowska Gronostaj](#), [Karin Friberg Heppin](#),
[Richard Johansson](#), [Dimitrios Kokkinakis](#), [Leif-Jöran Olsson](#), [Jonatan Uppström](#).

Follow the development via  [RSS](#).

Swedish FrameNet++ (SweFN++)

This project is funded by the Swedish Research Council for the years 2011-2014 (nr 2010-6013) and with a strategic research grant from the University of Gothenburg for the focus research area language technology (2009-2015).

The goal of the SweFN++ project is to build an open-content -- i.e., freely available and modifiable -- integrated lexical resource for Swedish -- so far lacking -- to be used as a basic infrastructural component in Swedish language technology (LT) research and in the development of LT applications for Swedish.

The resource -- Swedish FrameNet++ -- will consist of two main components:

1. a Swedish framenet covering at least 50,000 lexical units built on the same principles as the English Berkeley FrameNet (BFN) and to be developed in collaboration with the BFN team at ICSI Berkeley;
2. an integration of a number of existing free lexical resources, constructed by harmonizing, standardizing and merging these resources, and thereby reusing the valuable grammatical and semantic information painstakingly collected in these resources

Additionally, we aim to develop a methodology and workflow which makes maximal use of LT and other tools in order to minimize the human effort needed to build SweFN++.

SweFN++ will be a versatile basic lexical building block in Swedish LT research and LT applications, where two important areas in the near future will be LT-based eScience (particularly in the framework of CLARIN) and the processing of language data in connection with the Semantic Web. In both cases the semantic information in SweFN++ will be crucial to the realization of the full potential of those areas.

