Lexical Resources

Language Technology Resources (LT2304)

Markus Forsberg
Språkbanken
University of Gothenburg

2012-09-13/14
Overview of today’s and tomorrow’s lecture

- Korp demo – a corpus search interface
- Lexical resources for LT
- LMF (Lexical Markup Framework)
- Lexical semantics
  *Slides by Jurafsky*
- Princeton Wordnet
  *Slides by Jurafsky*
- Berkeley FrameNet
  *Slides by Fillmore*
- the SweFN++ project
- **Note: tomorrow’s lecture in L307**
Korp demo

▶ http://spraakbanken.gu.se/korp/#lang=en
Lexical resources refer to resources containing information about lexical units (e.g., formal descriptions, relations, combinations).

Lexical resources for LT differ from paper dictionaries in that they are:

- created for machines, not humans
- formal
- practically oriented (a lexical resource for LT that cannot be used in automatic text analysis is kinda worthless from a LT perspective).
LMF: Lexical Markup Framework

- LMF is the ISO standard (ISO-24613:2008) for representing LT lexical resources.

Example:

```
<LexicalResource dtdVersion="15">
  <GlobalInformation>
    <feat att="languageCoding" val="ISO 639-3"/>
  </GlobalInformation>
  <Lexicon>
    <feat att="language" val="eng"/>
    <LexicalEntry>
      <feat att="partOfSpeech" val="commonNoun"/>
      <Lemma>
        <feat att="writtenForm" val="clergyman"/>
      </Lemma>
      <WordForm>
        <feat att="writtenForm" val="clergyman"/>
        <feat att="grammaticalNumber" val="singular"/>
      </WordForm>
      <WordForm>
        <feat att="writtenForm" val="clergymen"/>
        <feat att="grammaticalNumber" val="plural"/>
      </WordForm>
    </LexicalEntry>
  </Lexicon>
</LexicalResource>
```

- Using a standard is never painless, but it beats having to create a special-purpose solution for every single lexical resource.
Three Perspectives on Meaning

1. **Lexical Semantics**
   - The meanings of *individual words*

2. **Formal Semantics** (or Compositional Semantics or Sentential Semantics)
   - How those meanings combine to make meanings for *individual sentences or utterances*

3. **Discourse or Pragmatics**
   - How those meanings combine with each other and with other facts about various kinds of context to make meanings for a *text or discourse* 

(+ Dialog or Conversational Semantics)
Relationships between word meanings

- Homonymy
- Polysemy
- Synonymy
- Antonymy
- Hypernomy
- Hyponomy
- Meronomy
First idea: The unit of meaning is called a Sense or wordsense

- One word “bank” can have multiple different meanings:
  - “Instead, a bank can hold the investments in a custodial account in the client’s name”
  - “But as agriculture burgeons on the east bank, the river will shrink even more”

- We say that a sense is a representation of one aspect of the meaning of a word.
  - Thus bank here has two senses
    - Bank1:
    - Bank2:
Some more terminology

- Lemmas and wordforms
  - A **lexeme** is an abstract pairing of meaning and form
  - A **lemma** or **citation form** is the grammatical form that is used to represent a **lexeme**.
    - *Carpet* is the lemma for *carpets*
    - *Dormir* is the lemma for *duermes*.
  - Specific surface forms *carpets, sung, duermes* are called **wordforms**

- The lemma **bank** has two **senses**:
  - Instead, a **bank** can hold the investments in a custodial account in the client’s name
  - But as agriculture burgeons on the east **bank**, the river will shrink even more.

- A **sense** is a discrete representation of one aspect of the meaning of a word
Homonymy

• **Homonymy:**
  - Lexemes that share a form
    - Phonological, orthographic or both
  - But have unrelated, distinct meanings
  - Clear example:
    - Bat (wooden stick-like thing) vs Bat (flying scary mammal thing)
    - Or bank (financial institution) versus bank (riverside)
• Can be homophones, homographs, or both:
  - Homophones:
    - Write and right
    - Piece and peace
Homonymy causes problems for NLP applications

- Text-to-Speech
  - Same orthographic form but different phonological form
    - bass vs bass

- Information retrieval
  - Different meanings same orthographic form
    - QUERY: bat care

- Machine Translation

- Speech recognition
  - Why?
Polysemy

1. The **bank** was constructed in 1875 out of local red brick.
2. I withdrew the money from the **bank**

Are those the same sense?

- We might call sense 2:
  - “A financial institution”

And sense 1

- “The building belonging to a financial institution”

Or consider the following example

- While some banks furnish sperm only to married women, others are less restrictive

Which sense of bank is this?
Polysemy

- We call **polysemy** the situation when a single word has multiple related meanings (bank the building, bank the financial institution, bank the biological repository)
- Most non-rare words have multiple meanings
Polysemy: A systematic relationship between senses

- Lots of types of polysemy are systematic
  - School, university, hospital
  - Can all be used to mean the institution or the building.
- We might say there is a relationship:
  - Building <-> Organization
- Other such kinds of systematic polysemy:

  Author (Jane Austen wrote Emma) <-> Works of Author (I really love Jane Austen)
  Animal (The chicken was domesticated in Asia) <-> Meat (The chicken was overcooked)
  Tree (Plums have beautiful blossoms) <-> Fruit (I ate a preserved plum yesterday)
How do we know when a word has more than one sense?

- Consider examples of the word “serve”:
  - Which flights serve breakfast?
  - Does America West serve Philadelphia?
- The “zeugma” test:
  - Does United serve breakfast and San Jose?

Since this sounds weird, we say that these are two different senses of “serve”
Synonyms

- Word that have the same meaning in some or all contexts.
  - filbert / hazelnut
  - couch / sofa
  - big / large
  - automobile / car
  - vomit / throw up
  - Water / H₂O
- Two lexemes are synonyms if they can be successfully substituted for each other in all situations
  - If so they have the same propositional meaning
Synonyms

• But there are few (or no) examples of perfect synonymy.
  • Why should that be?
  • Even if many aspects of meaning are identical
  • Still may not preserve the acceptability based on notions of politeness, slang, register, genre, etc.

• Example:
  • Water and H$_2$O
  • Big/large
  • Brave/courageous
Synonymy is a relation between senses rather than words

- Consider the words *big* and *large*
- Are they synonyms?
  - How *big* is that plane?
  - Would I be flying on a *large* or small plane?

- How about here:
  - Miss Nelson, for instance, became a kind of *big* sister to Benjamin.
  - Miss Nelson, for instance, became a kind of *large* sister to Benjamin.

- Why?
  - *big* has a sense that means being older, or grown up
  - *large* lacks this sense
Antonyms

- Senses that are opposites with respect to one feature of their meaning
- Otherwise, they are very similar!
  - dark / light
  - short / long
  - hot / cold
  - up / down
  - in / out
- More formally: antonyms can define a binary opposition or at opposite ends of a scale (long/short, fast/slow)
- Be **reversives**: rise/fall, up/down
Hyponymy

- One sense is a **hyponym** of another if the first sense is more specific, denoting a subclass of the other
  - *car* is a hyponym of *vehicle*
  - *dog* is a hyponym of *animal*
  - *mango* is a hyponym of *fruit*

- Conversely
  - *vehicle* is a hypernym/superordinate of *car*
  - *animal* is a hypernym of *dog*
  - *fruit* is a hypernym of *mango*
Hypernymy more formally

- Extensional:
  - The class denoted by the superordinate
  - Extensionally includes the class denoted by the hyponym

- Entailment:
  - A sense A is a hyponym of sense B if being an A entails being a B

- Hyponymy is usually transitive
  - (A hypo B and B hypo C entails A hypo C)
II. WordNet

- A hierarchically organized lexical database
- On-line thesaurus + aspects of a dictionary
  - Versions for other languages are under development

<table>
<thead>
<tr>
<th>Category</th>
<th>Unique Forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun</td>
<td>117,097</td>
</tr>
<tr>
<td>Verb</td>
<td>11,488</td>
</tr>
<tr>
<td>Adjective</td>
<td>22,141</td>
</tr>
<tr>
<td>Adverb</td>
<td>4,601</td>
</tr>
</tbody>
</table>
WordNet

• Where it is:
  • http://www.cogsci.princeton.edu/cgi-bin/webwn
Format of Wordnet Entries

The noun “bass” has 8 senses in WordNet.
1. bass\textsuperscript{1} - (the lowest part of the musical range)
2. bass\textsuperscript{2}, bass part\textsuperscript{1} - (the lowest part in polyphonic music)
3. bass\textsuperscript{3}, basso\textsuperscript{1} - (an adult male singer with the lowest voice)
4. sea bass\textsuperscript{1}, bass\textsuperscript{4} - (the lean flesh of a saltwater fish of the family Serranidae)
5. freshwater bass\textsuperscript{1}, bass\textsuperscript{5} - (any of various North American freshwater fish with lean flesh (especially of the genus Micropterus))
6. bass\textsuperscript{6}, bass voice\textsuperscript{1}, basso\textsuperscript{2} - (the lowest adult male singing voice)
7. bass\textsuperscript{7} - (the member with the lowest range of a family of musical instruments)
8. bass\textsuperscript{8} - (nontechnical name for any of numerous edible marine and freshwater spiny-finned fishes)

The adjective “bass” has 1 sense in WordNet.
1. bass\textsuperscript{1}, deep\textsuperscript{6} - (having or denoting a low vocal or instrumental range)
   "a deep voice"; "a bass voice is lower than a baritone voice";
   "a bass clarinet"
<table>
<thead>
<tr>
<th>Relation</th>
<th>Also called</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyponym</td>
<td>Superordinate</td>
<td>From concepts to superordinates</td>
<td>breakfast¹ → meal¹</td>
</tr>
<tr>
<td></td>
<td>Subordinate</td>
<td>From concepts to subtypes</td>
<td>meal¹ → lunch¹</td>
</tr>
<tr>
<td></td>
<td>Has-Member</td>
<td>From groups to their members</td>
<td>faculty² → professor¹</td>
</tr>
<tr>
<td></td>
<td></td>
<td>From concepts to instances of the concept</td>
<td>composer¹ → Bach¹</td>
</tr>
<tr>
<td>Has-Instance</td>
<td>Member-Of</td>
<td>From instances to their concepts</td>
<td>Austen¹ → author¹</td>
</tr>
<tr>
<td>Instance</td>
<td>Has-Part</td>
<td>From members to their groups</td>
<td>copilot¹ → crew¹</td>
</tr>
<tr>
<td></td>
<td>Part-Of</td>
<td>From wholes to parts</td>
<td>table² → leg³</td>
</tr>
<tr>
<td>Member Holonym</td>
<td></td>
<td>From parts to wholes</td>
<td>course⁷ → meal¹</td>
</tr>
<tr>
<td>Part Meronym</td>
<td></td>
<td>Opposites</td>
<td>leader¹ → follower¹</td>
</tr>
<tr>
<td>Part Holonym</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antonym</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relation</td>
<td>Definition</td>
<td>Example</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>-------------------------------------------------</td>
<td>--------------------------</td>
<td></td>
</tr>
<tr>
<td>Hypernym</td>
<td>From events to superordinate events</td>
<td>fly^9 \rightarrow travel^5</td>
<td></td>
</tr>
<tr>
<td>Troponym</td>
<td>From a verb (event) to a specific manner elaboration of that verb</td>
<td>walk^1 \rightarrow stroll^1</td>
<td></td>
</tr>
<tr>
<td>Entails</td>
<td>From verbs (events) to the verbs (events) they entail</td>
<td>snore^1 \rightarrow sleep^1</td>
<td></td>
</tr>
<tr>
<td>Antonym</td>
<td>Opposites</td>
<td>increase^1 \leftrightarrow decrease^1</td>
<td></td>
</tr>
</tbody>
</table>
WordNet Hierarchies

Sense 3
bass, basso --
(an adult male singer with the lowest voice)
=> singer, vocalist, vocalizer, vocaliser
    => musician, instrumentalist, player
    => performer, performing artist
    => entertainer
    => person, individual, someone...
    => organism, being
        => living thing, animate thing,
            => whole, unit
                => object, physical object
                    => physical entity
                        => entity
    => causal agent, cause, causal agency
        => physical entity
            => entity

Sense 7
bass --
(the member with the lowest range of a family of musical instruments)
=> musical instrument, instrument
   => device
       => instrumentality, instrumentation
           => artifact, artefact
               => whole, unit
                   => object, physical object
                       => physical entity
                           => entity
How is “sense” defined in WordNet?

- The set of near-synonyms for a WordNet sense is called a **synset** (synonym set); it’s their version of a sense or a concept
- Example: *chump* as a noun to mean
  - ‘a person who is gullible and easy to take advantage of’

\{chump\textsuperscript{1}, fool\textsuperscript{2}, gull\textsuperscript{1}, mark\textsuperscript{9}, patsy\textsuperscript{1}, fall guy\textsuperscript{1}, sucker\textsuperscript{1}, soft touch\textsuperscript{1}, mug\textsuperscript{2}\}

- Each of these senses share this same gloss
- Thus for WordNet, the meaning of this sense of *chump* is this list.
Princeton Wordnet is (probably) the most used LT resource. This because:

- it describes the English language;
- it is large-scale;
- it is freely available.

However, it has been criticized for having a too fine-grained sense distinction (actually, it more or less follows distinctions made in paper dictionaries).

If we cannot distinguish senses (the task called WSD: word sense disambiguation), then it is difficult/impossible to use the information in Wordnet.
Hanks (1992) divides lexicographers into “lumpers” or “splitters”.

As discussed in Ide and Wilks paper, LT needs more more lumping and less splitting.

Sense distinction is difficult for humans also, many report on around 80% inter-annotator agreement (e.g., Edmonds and Kilgarriff, 2002)
Lumping senses

1. paper – (a material made of cellulose pulp derived mainly from wood or rags or certain grasses)
2. composition, paper, report, theme – (an essay (especially one written as an assignment); “he got an A on his composition”)
3. newspaper, paper – (a daily or weekly publication on folded sheets; contains news and articles and advertisements; "he read his newspaper at breakfast")
4. paper – (a scholarly article describing the results of observations or stating hypotheses; "he has written many scientific papers")
5. paper – (medium for written communication; "the notion of an office running without paper is absurd")
6. newspaper, paper, newspaper publisher – (a business firm that publishes newspapers; "Murdoch owns many newspapers")
7. newspaper, paper – (a newspaper as a physical object; "when it began to rain he covered his head with a newspaper"

⇒

1. paper – material (1 and 5)
2. paper – composition, article (2 and 4)
3. paper – newspaper, publication, publisher (3,6,7)
Website Back Up :-)  
The scheduled maintenance was completed on Saturday, May 17th, and this website should now be up and fully functional. We have noticed that the startup of the top page is sometimes very slow, although the rest of the site seems to function normally after that. We are looking into the cause of this problem. If you notice any other problems with the website, or evidence of misuse, please contact us. Thank you for your patience.

Last Updated (May 20, 2008 at 11:30 AM)

Welcome to FrameNet

The Berkeley FrameNet project is creating an on-line lexical resource for English, based on frame semantics and supported by corpus evidence. The aim is to document the range of semantic and syntactic combinatory possibilities (valences) of each word in each of its senses, through computer-assisted annotation of example sentences and automatic tabulation and display of the annotation results. The major product of this work, the FrameNet lexical database, currently contains more than 10,000 lexical units (defined below), more than 6,100 of which are supported by example sentences.

The "Book" v1.3 Available for Download

Our main project document, "FrameNet II: Extended Theory and Practice" gives a basic introduction to frame semantics and offers guidelines for frame semantic annotation, discussing in detail the inventories of grammatical functions and phrase types that we use. The document includes a discussion of the frame development process and a consolidated list of frame-general, extra-thematic frame elements. Our latest version contains some general updates plus a new section devoted to the relationship of FrameNet to other lexical resources in the framework of the Annotated Corpus Lexicon framework.
FrameNet is a lexicography project. That means we’re making a dictionary.

But hasn’t that already been done? What’s wrong with the dictionaries we’ve got?
What’s wrong with dictionaries?

• Dictionaries only indirectly give access to the conceptual structures underlying word meanings. Human intelligence is required to find connections hidden in defining phrases.
Are online thesauri helpful?

- Thesauri show us that certain groups of words are semantically related, but only indirectly how they are related; and they show nothing of a word’s combinatorial behavior.
What about WordNet?

- WordNet is a vast online lexical resource that combines the work of dictionaries and thesauri, with the flaws of each, but provides essentially no information about the combinatory possibilities of the words.
The FN web page welcomes you with a description of the goals and products of this decade-old project and encourages you to follow links that will lead to more information.
In particular

That opening paragraph explains that we are building a lexicon
(a) based on corpus evidence,
(b) with words defined in terms of the frames they evoke,
(c) documenting the combinatorial properties of each word through frame-relevant annotations, and
(d) automatically producing summaries as lexical entries.
Back to the question: What’s wrong with the dictionaries we’ve got?
Let’s look at what dictionaries do with the nouns
pedestrian
customer
land
chicken
pedestrian

some definitions:

• ‘a person walking along a road or a developed area’ (Mac OSX)
• ‘someone who travels by foot’ (WN)
• ‘one who is travelling on foot, especially in an area also used by cars’ (MS)
• ‘WALKER: person on foot; Opposite: driver’ (OALD)
WordNet Hierarchy

- traveller
  (one who moves from one place to another)

- pedestrian, walker, footer
  (one who travels on foot)

- straggler, tramp, rambler, jaywalker, ...
  (one who walks in a particular way)
What does it mean to “hit a pedestrian”?

- Suppose you overheard me say: 
  
  \textit{I hit a pedestrian on my way to work.}

- Which of these is the more likely interpretation?
  - I’m a belligerent guy, and punched a man who was walking next to me.
  - I was driving a car or bicycle, and my vehicle collided with someone on the road.
How is the word used?

• Example sentences (from the web)
  • Mr. Peguero was not aware that *his vehicle had struck a pedestrian*.
  • Specifically, if *you hit a pedestrian* while driving at 20 mph, the pedestrian has a 95% chance of survival.
  • If *you hit a pedestrian*, it is your fault even if you have the right of way.
  • *A sports car hit a pedestrian* trying to cross the road.
How to describe the frame?

A frame-informed definition (tentative):

There are spaces shared competitively by moving vehicles - buses, cars, bicycles, etc. - and people moving on foot. *Within that context*, the word designating the people moving about on foot is **pedestrian**.

Use of the word always implies the actual or potential co-presence of people and vehicles. If you die of a heart attack while walking in the park, your name will **not** get listed in the city’s statistics on **pedestrian casualties**.
Common compounds

pedestrian casualties
pedestrian fatalities
pedestrian accidents
pedestrian zones
pedestrian safety
pedestrian-friendly roads
pedestrian crossing
There were hints of the frame in the definitions.

• ‘a person walking along a road or a developed area’ (Mac OSX)
• ‘someone who travels by foot’ (WN)
• ‘one who is travelling on foot, especially in an area also used by cars’ (MS)
• ‘WALKER: person on foot; Opposite: driver’ (OALD)
Are there other words like that?

Thousands. Here are a few, with typical definitions

customer
‘someone who buys something at a shop or business’

land
‘the solid part of the earth’s surface’

chicken
‘(mass noun): the meat of a chicken or chickens’
The noun customer is typically defined as ‘someone who buys something in a shop or business.’ That includes everyone I know over the age of 5.

Suppose you overhear somebody say

Sue tends to be rude to customers.

What situation do you imagine?
The noun **land** tends to be defined as something like ‘the solid part of the earth’s surface’. WordNet adds **earth** and **ground** as **synonyms**.

Suppose you find in an article about sea birds ...

**Auks build their nests on land.**

(Is that informative? Where else can a bird build a nest?)

The word **land** expresses the contrast with **sea**. Auks spend most of their time at sea. Nobody would say - though it is true - **Sparrows build their nests on land.**
The noun chicken, as a *count noun*, is the name of a well-known domestic bird. As a mass noun it is defined as ‘the meat of a chicken or chickens’.

What’s wrong with the following sentence?

*The fox that lives near our farm likes chicken.*  *(compare: likes chickens)*

The image you might get is of a fox eating fried chicken, holding a knife and a fork, and a napkin, in its paws.
• The products of the lexical construction that yields mass noun uses of chicken, lamb, duck, turkey, etc., refer to meats prepared as part of human cuisine.

*The wolf that lives near our ranch prefers lamb.
We make sense of the things we see around us by fitting them into the ready-made frames we have in our heads. We invoke those frames, whether or not any words associated with the events have been used.

When we encounter a word that “belongs to” a particular frame, we can say that the word evokes the frame in our minds.

As a lexicographer, I ask of the frame-evoking words, what kinds of information about them should be recorded in an adequate dictionary.
Invoking and Evoking Frames

- **People invoke** (summon up from their memory) frames, to make sense of their **experience**, linguistic or otherwise.
- a cognitive act

- **Words evoke** categories and knowledge structures that shape interpreters’ understanding of a text.
- a cognitive experience

*Warning: this is not a standard use of these words.*
So,

We need to describe words in terms of the “framal” background.
If we don’t understand the frame, we don’t understand 
the word, 
or why the language needs this word, 
or why the speaker chose to use it.
• This point is easiest to see in the case of specialist vocabulary:
  • Freud, psychoanalytic theory
    • *id*, *ego*, *superego*, *transference*
  • Tectonic plate theory
    • *subduction zone*, *oceanic trench*, *transform boundary*
  • Baseball - game, conditions, strategies
    • *inning*, *short porch*, *gap hitter*
Frame semantics assumes that all content words in a language -- not just the special-domain vocabulary, but also the words everybody knows -- are best explained by appealing to the conceptual backgrounds (the frames) that underlie their meanings and motivate their use.
Double-decker dictionary

• Ideally, then, a dictionary entry for content words should have two parts:
  • One to explain the “framal” background of the meaning of a word.
  • One to explain the role of that word within or against that background.
The ideal dictionary should let you

1. Look up a word
2. Get a link to a description of the relevant frame, for each of its meanings, and see the names of the frame’s components
3. See a display of its combinatory affordances, its **valence** possibilities, both semantic and syntactic
4. Find a collection of example sentences illustrating all of its main combinatory patterns
5. Find a list of other words that evoke the same frame
6. Link to other semantically related frames
What kind of description do we end up with?

• We don’t have to describe one word at a time if we recognize that many words evoke the same frame, and so we begin by collecting words into frames.

• Accompanying each frame description, we provide a list of frame elements -- something like the “cases” of old -- as naming the kinds of things worth talking about within the given frame.
Background: FrameNet

The FrameNet project itself began in the late 90’s with ambitions I shared with lexicographer Sue Atkins on

(1) using corpus data for discovering the combinatory possibilities of “frame-bearing” English words and
(2) describing their “valences” in terms of

(a) semantic roles
(b) syntactic properties and functions,
using “frame semantic” notions for (a) and more-or-less theory-neutral notions for (b).
Frame examples: Risk

Taking_a_risk:

- Protagonist, Action, Harm, Asset
  - I’m going to risk a swim in the sea.
  - You’ll risk bankruptcy if you make that investment.
  - You’re risking your reputation by doing that.
  - You’re taking a big risk.

Being_at_risk:

- Protagonist, Harm, Asset
  2. Newborns in this hospital run the risk of hypothermia.
  3. We risk our lives every day.
  4. I am at risk of stroke.
Frame examples: Explanation

Communication.explanation:
- Speaker, Addressee, Mystery, Account
  - The coach explained the problem to the team.
  - The coach explained that they hadn’t learned the maneuvers.
  - What’s your explanation of these facts?
  - The defense lawyer gave an inadequate explanation.

Cognition.explanation:
- Mystery, Account
  - What can explain these facts?
  - A history of unrestricted logging explains the erosion pattern we see here.
Next let’s look at a small family of visualizable frames, with a temporal structure, involving a Role (a place, a job, a title), an Old Occupant, a New Occupant, and, for some cases, an Agent who brings about a change in role occupancy.
Reading the Pictures

- The boxes refer to five-part scenarios consisting of an initial state, a transition, an intermediary state, another transition, and a final state.
- The writing under the pictures abbreviates particular role names and gives verbs that evoke instances of the scenario.
- The bold borders indicate a profiling of some portion of the event.

| state | transition | state | transition | state |
The first examples have only one “thing”, and it just goes away and comes back.
He returned to Hong Kong.
He returned Tuesday evening after a week’s trip to Australia.
He returned to his home for a few days.

The verb RETURN profiles the time of arrival, but it evokes the entire frame; other information in the sentences can fill in some of the details of the larger scenario.
I returned your books this morning.

I returned to your desk the books that I had borrowed last week.

After the earthquake we replaced all the books on the shelf.
Now examples with two objects, changing places.
J'ai remplacé la N, O; replace, substitute (for)

Jim has replaced me.

Jim has replaced me on your examination committee.

Nobody will ever replace you in my heart.

You are irreplaceable.
Introducing an outside Agent, and alternative profiles.
A, N, O; replace (O with N); substitute (N for O)

You’d better work harder; you can be replaced.
We’d better replace these weak batteries.
After the scandal, we immediately replaced Ed with Morgan.

*Here a two-step act is profiled: removal and placement.*
Three years after Smith retired, the department finally replaced him. Somebody stole my bicycle; it was a year before I could afford to replace it.
The last two show a difference in profiling, recognizing the possibility in one case that the Agent who put the New thing in place is not responsible for removing the Old thing.
A, N, O; replace (O with N); substitute (N for O)
Some ambiguities
John replaced me.
John replaced the telephone.
Replacing

Agentive_replacing:
Agent, New, Old [Place]
1. The chairman replaced me with my worst enemy.
2. I think we can replace sugar with honey in this recipe.

Non-agentive_replacing:
New, Old [Place]
• Smith has replaced me on the committee.
• Honey can replace sugar in this recipe.
REVENGE

One FN frame that is simple enough to describe completely, and just complex enough to be interesting, is the so-called Revenge frame, the nature of which requires understanding a kind of history. In that history, one person (we call him the Offender) did something to harm another person (what he did we call the Offense and his victim we call the Injured_party); reacting to that act, someone (the Avenger, possibly the same individual as the Injured_party) acts so as to do harm to the Offender, and what he does we call the Punishment.
frame elements

- **Participants and Sub-events**
  - **Avenger**: the one who enacts revenge
  - **Offender**: the original offender
  - **Injured_party**: the offender’s victim
  - **Injury**: the offender’s act
  - **Punishment**: the avenger’s act
grammar

• Components of linguistic form for expressing the FEs (defining valence).
  • Subject
  • Direct Object
  • Prepositional marking (by, for, with, on, at, against)
  • Subordinate clause marking (for DOING, by DOING)
From start to finish: Revenge

The **Revenge** frame involves a situation in which

a) A has done something to harm B and

b) C - who cares about B - takes action to harm A in turn

c) C’s action is carried out independently of any legal or other institutional setting.

d) B and C may or may not be the same individual.
We avenged the insult by setting fire to his village.
We avenged the insult by setting fire to his village.
We avenged the insult by setting fire to his village.
We avenged the insult by setting fire to his village.
Find other words in the same frame. And then do the same with each of them.

REVENGE FRAME

wordlist

avenge, revenge, retaliate, get back at, pay back, get even, ...

revenge, vengeance, retaliation, retribution, reprisal, ...

vengeful, retaliatory, retributive; in revenge, in retaliation, ...

take revenge, wreak vengeance, exact retribution, ...
Annotation Reports > Valence descriptions

Annotation is in layers (“standoff”)
1. target (frame-bearing) word (here, a verb)
2. semantic role labels
3. grammatical function labels
4. syntactic form (phrase type) labels

Example of one valence for **avenge**:
{Avenger:subject:NP,
 Injured:object:NP,
 Punishment:Oblique:by+VPing}

Another possibility is Injured_party as object: **He avenged his brother…**
Swedish lexical LT resources

created, but not shared  created, but “forgotten”

- Språkbanken, with a long history as a resource dumping site, is the keeper of many “forgotten” (bit-rotten) resources.
The SweFN++ project

- SweFN++ is a project conducted at Språkbanken with two main objectives:
  - to create of a new lexical resource: a Swedish framenet covering at least **50,000 lexical units** built on the same principles as the English Berkeley FrameNet;
  - to **curate** and **integrate** existing free Swedish lexical resources into one “super lexical resource”.

- the SweFN++ plans involve:
  - building a diachronic lexical resource
  - exploring language technology-based methods for automatizing the acquisition of lexical information from corpora and other linguistic resources
  - making the resulting resource open-content
The SweFN++ project

- in particular, we would like to explore
  - automatic selection of ‘good’ examples from corpora
  - addition of senses and discovery of related frames via lexical-semantic links
  - how to link general semantic roles (from GLDB/SDB) to frame elements
  - (principled) inclusion of multi-word units
  - inclusion of constructions
  - (ongoing) longer-term plan: the creation of an infrastructure for lexical resources and corpora material (an under-researched area).
SALDO is the pivot

- In SweFN++ we are using the SALDO resource as the pivot to which everything else is linked.
- SALDO is a lexical resource with associative semantic relations.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>wordforms</td>
<td>1,825,261</td>
</tr>
<tr>
<td>senses</td>
<td>118,402</td>
</tr>
<tr>
<td>lemgrams</td>
<td>115,490</td>
</tr>
<tr>
<td>paradigms</td>
<td>1,323</td>
</tr>
<tr>
<td>PoS</td>
<td>37</td>
</tr>
</tbody>
</table>
SALDO “PIDs”

- SALDO has id’s for:
  - senses (grad..1)
  - lemgrams (grad..nn.1)
  - parts of speech (nn)
  - paradigms/inflection tables (nn_3u_film)

- the id’s are designed to be
  - unique (no other id’s should be necessary, e.g., database keys)
  - atomic (no built-in assumptions about sense–subsense relationships, etc.)
  - **usable in Semantic Web formalisms** (RDF, OWL): id’s are well-formed XML names
  - human-readable (makes resources easier to work with)
<table>
<thead>
<tr>
<th>lex:</th>
<th>fisk</th>
</tr>
</thead>
<tbody>
<tr>
<td>fm:</td>
<td>djur</td>
</tr>
<tr>
<td>fp:</td>
<td>vatten</td>
</tr>
<tr>
<td>PRIM:</td>
<td>abborre benfisk berggylta braxen bröding firre glasbult gylta&lt;sup&gt;2&lt;/sup&gt; gädda gös haj havskatt havsmus hälleflundra id iktyologisk kantnål klumpfisk knorrhane koffertfisk kolja kummel kvastfening kvidd lake lansettfisk lax lumpfisk lungfisk långa makrill marulk mört nejonöga nors nototenioïd nätting rom&lt;sup&gt;2&lt;/sup&gt; sarv sill skolast&lt;sup&gt;2&lt;/sup&gt; smörbult stör&lt;sup&gt;2&lt;/sup&gt; tobakspipa&lt;sup&gt;2&lt;/sup&gt; tonfisk torsk ål</td>
</tr>
<tr>
<td>afrikansk:</td>
<td>ciklidfisk</td>
</tr>
<tr>
<td>andas:</td>
<td>gäl</td>
</tr>
<tr>
<td>brosk:</td>
<td>broskfiskar</td>
</tr>
<tr>
<td>damm:</td>
<td>karp</td>
</tr>
<tr>
<td>djuphav:</td>
<td>djuphavsfisk</td>
</tr>
<tr>
<td>flyga:</td>
<td>flygfisk</td>
</tr>
<tr>
<td>giftig:</td>
<td>drakhuvudsfisk</td>
</tr>
<tr>
<td>grå:</td>
<td>gräsej</td>
</tr>
<tr>
<td>hav:</td>
<td>saltvattensfisk</td>
</tr>
<tr>
<td>hästliknande:</td>
<td>sjöhäst</td>
</tr>
<tr>
<td>insjö:</td>
<td>insjöfisk</td>
</tr>
</tbody>
</table>
### SALDO

<table>
<thead>
<tr>
<th>lex:</th>
<th>PRIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>fm:</td>
<td>*</td>
</tr>
<tr>
<td>fp:</td>
<td>*</td>
</tr>
<tr>
<td>mf(51):</td>
<td>PRIM: all annan använda att bara bra den en för framme färg för förbi före genom göra ha hur hända i ja just kunna ljud ljus med men mycken måste namn natur när och om om på rak röra säga tal till tänka vad var vara varm vem veta vid vilja öppen</td>
</tr>
</tbody>
</table>

| pf(42480): | * |

---

SALDO, 2
<table>
<thead>
<tr>
<th>lex:</th>
<th>spelautomat</th>
</tr>
</thead>
<tbody>
<tr>
<td>fm:</td>
<td>automat</td>
</tr>
<tr>
<td>fp:</td>
<td>hasardspel</td>
</tr>
<tr>
<td>mf(2):</td>
<td>PRIM: arkadspel enarmad_bandit</td>
</tr>
<tr>
<td>pf(1):</td>
<td>vinst: jackpot</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>word class</th>
<th>inherently feature(s)</th>
<th>inflection table</th>
</tr>
</thead>
<tbody>
<tr>
<td>sg indef nom</td>
<td>enarmad bandit</td>
<td></td>
</tr>
<tr>
<td>sg indef gen</td>
<td>enarmad bandits</td>
<td></td>
</tr>
<tr>
<td>sg def nom</td>
<td>enarmade banditen</td>
<td></td>
</tr>
<tr>
<td>sg def gen</td>
<td>enarmade banditens</td>
<td></td>
</tr>
<tr>
<td>pl indef nom</td>
<td>enarmade banditer</td>
<td></td>
</tr>
<tr>
<td>pl indef gen</td>
<td>enarmade banditers</td>
<td></td>
</tr>
<tr>
<td>pl def nom</td>
<td>enarmade banditerna</td>
<td></td>
</tr>
<tr>
<td>pl def gen</td>
<td>enarmade banditernas</td>
<td></td>
</tr>
</tbody>
</table>
merging existing resources

- existing resources in many formats and with heterogeneous content categories
- minimally we need two kinds of common categories:
  1. (lexical) senses
  2. lemgrams (with accompanying inflectional paradigms)
- our goal is to link all resources using the senses or the lemgrams or both (and senses are linked to lemgrams in SALDO)
- all information must be explicit and unambiguous
- we will use SALDO identifiers
merging lexical resources

- format harmonization can by and large be automatized
- how much of content merging can be automatized and how much manual work is needed?
- enter Zipf
George Kingsley Zipf (1902–1950)
(from Wikipedia)
senses/lemma in SALDO
lemmas $\Rightarrow$ senses in SALDO
with Zipf towards the future

- our hypothesis: since the majority of lemmas generally correspond to only one sense each in our lexical resources, it will be possible to merge the resources largely automatically with an acceptably small manual post-processing effort,
- and, further, the merged resource will be of acceptable quality for practical applications
- in fact, preliminary experiments largely bear this out
Berkeley FrameNet: Operate vehicle

Operate_vehicle

Definition:
The words in this frame describe motion involving a Vehicle and someone who controls it, the Driver. Some words normally allow the Vehicle to be expressed as a separate constituent.

Tom DROVE his car all the way across North America.

Other words in this domain are based on the names of vehicles, and do not normally allow the Vehicle to be expressed as a separate constituent.

The group BIKEd all the way across the country.

However, a separate Vehicle constituent can occur if it adds information not included in the verb.

Kim BIKEd across the country on an old 10-speed.

FES:

Core:

Area [Area]

Driver [Driver]
Semantic Type: Sentient
Excludes: Area

Goal [Goal]
Semantic Type: Goal

This frame element is used for expressions which describe a general area in which motion takes place when the motion is understood to be irregular and not to consist of a single linear path. Locative setting adjuncts of motion expressions may also be assigned this frame element.

The bikers RODE all over the place.

This is the being, typically human, that controls the Vehicle as it moves.

Kim DROVE my old car cross-country.

Any expression which tells where the moving object(s) ends up as a result of the motion expresses the frame element Goal. Some particles imply the existence of a Goal which is understood in the context of utterance.

Kim DROVE into the parking lot.

Kim DROVE in.
**SweFN: Operate vehicle frame**

<table>
<thead>
<tr>
<th>ram</th>
<th>Operate_vehicle</th>
</tr>
</thead>
<tbody>
<tr>
<td>domän</td>
<td>Gen</td>
</tr>
<tr>
<td>semantisk typ</td>
<td>Move</td>
</tr>
<tr>
<td>känelement</td>
<td>Area, Driver, Goal, Path, Source, Vehicle</td>
</tr>
<tr>
<td>periferielement</td>
<td>Circumstances, Cotheine, Degree, Depictive, Distance, Duration, Event, Explanation, Frequency, Manner, Means, Particular_iteration, Place, Purpose, Result, Route, Speed, Time</td>
</tr>
<tr>
<td>sms</td>
<td>Distance+LU, Manner+LU, Means+LU, Path+LU, Purpose+LU, Time+LU, Vehicle+LU,</td>
</tr>
<tr>
<td>sms-exempel</td>
<td>Distance+LU_EX_distansflykning, Manner+LU_EX_blaoköra, lagflyga, promonadköra, premårrköra, varmköra, Means+LU_EX_blandflykning, instrumentflykning, Path+LU_EX_vårdsjöomseglning, Purpose+LU_EX_provkör, spion]flykning, tävlingsssegla, övningsköra, Time+LU_EX_fyllsynsinsflykning, Vehicle+LU_EX_blandsegla</td>
</tr>
<tr>
<td>salto</td>
<td>vb: avsegla, 1 bila, 3 blåkör, 1 brådesegla, 1 busa, 1 cykla, 1 flyga, 1 flyga_upp, 1 insega, 1 kanola, 1 kappsegla, 1 krypkör, 1 kör, 1 kör, 2 köra_om, 1 länsa, 2 lagflyga, 1 paddla, 1 premårrköra, 1 promonadköra, 1 provköra, 1 rata, 1 råda, 1 råda_barhacka, 1 ro_2 Samantha, 1 segflyga, 1 segla, 1 skjuta, 1 slöra, 1 styr, 2 laxi, 2 tolk, 3 lärningssegla, 1 varmkör, 1 övningsköra, 1</td>
</tr>
<tr>
<td>salto (nya)</td>
<td>nn: antflygning, 1 avsegling, 1 bilde, 2 blandflygning, 1 blandsegla, 1 brådesegla, 1 busande, 1 cyklande, 1 distansflykning, 1 ensamflygning, 1 flygande_1 flygning, 1 fundflygning, 1 flyttsegling, 1 havskapssegling, 1 inflygning, 1 inseglade, 1 instrumentflykning, 1 kanotand, 1 kappsegling, 1 konflygning, 1 kyrkoflyga, 1 körande, 1 länsegla, 2 omseglare, 1 omseglning, 1 paddande, 1 premårrköra, 1 promonadköra, 1 provköra, 1 ratande, 1 risande, 1 rundflygning, 1 segflygning, 1 segflygning, 1 segande_1 segling, 1 skjutsande, 1 sörande, 1 spolflygning, 1 storskapssegling, 1 styrande_2 taxande, 1 tolkande, 3 tävlingssegla, 1 tävlingssegling, 1 varmkörande, 1 vårdsoseglning, 1 överflygning, 1</td>
</tr>
<tr>
<td>kommentar</td>
<td>MTG</td>
</tr>
<tr>
<td>skapad</td>
<td>2010-10-01</td>
</tr>
<tr>
<td>modifierad</td>
<td>2011-04-29</td>
</tr>
</tbody>
</table>
The linking

- We typically link to SALDO’s sense identifiers.
- However, most resources associate lexical information to PoS-tagged baseforms, where the information is not always valid for all senses of current baseform.
- In other words: we are creating new resources where we are able to be more precise.
The diversity of the lexical information may be used to improve the quality of its parts. E.g., syntactic valency information can be mirrored against the semantic valency information. We are currently working on a unified test bench for expressing these kinds of dependencies.
SweFN++ has a diachronic dimension (1)

- The starting point is four digitized paper dictionaries: Dalin (19th century Swedish), and three Old Swedish dictionaries.
- Linking SALDO and Dalin is relatively straightforward.
- The vocabulary differences are mainly in the compounds, e.g.:
  - bäfverhund ‘dog used for beaver hunt’
  - bäfverhund → bäverhund → bäver..1+hund..1
SweFN++ has a diachronic dimension (II)

- Linking Old Swedish to SALDO is a much more challenging task. An illustrative example:
  - *bakvapi* ‘fatal accident resulting from a sword being struck backwards without the striker looking in that direction beforehand’
- Naturally, there is no modern variant of this word, so it is an open, empirical question where it is most beneficial to link.
An important theme of the project is openness.

The theme is a philosophical stance — we believe that research should be carried out in the open to enable scrutinization and increased collaboration.

It is also our attempt of not being part of the problem we are trying to solve...
1. Openness from day one

- To make resources and related information accessible as soon as possible, preferably at day one.
- A project such as this has its main activity during its project time.
- This rather obvious observation implies that to enable someone to influence and contribute to the project, they need access as soon as possible.
2. Daily updates

- To deliver development versions of the resources, tools and related information regularly.
- This goal is related to the first one, since the input of others is only relevant if they have access to up-to-date information.
- Instantaneous updates would be preferred, but for technical reasons we have settled for daily updates.
3. Open content and open standards

- To deliver resources with an open content license and to use open standards for them
  - Licenses: CC-BY-SA 3.0 and LGPL 3.0 (use the resources however you like, but share your modifications)
  - Important standards: LMF, ISOCat, OLAC/CMDI
  - Necessary requirements to enable someone to make good use of the resources, or to continue the work that the SweFN++ project now started.
4. Web services

- To make the resources and tools available through web service APIs
- Web services are convenient ways of making resources and tools available.
- Enable instantaneous updates and a clean and platform-independent interface.
- Web services still suffer from network latency; batch processing using web services is only feasible for small materials.
- However, the network speed has increased drastically the last few years, so this will probably not be an issue in a not-so-distant future.
Openness in practise

- We have started the work on a lexical infrastructure to reach the aforementioned goals.
- SBLEX that became Karp.
Versioning

- The versioning system with anonymous access is our delivery channel for the lexical resources.
- The use of a versioning system has the advantage that not only the latest version of a resource is available but all of its history.
- Not to mention the added value of using a versioning system in a collaborative environment such as a research project.
SBLEX, our lexical infrastructure, is generic.

- Adding a new resource to the infrastructure only requires write access to the repository.
- The resource must be in a compatible format (+ a few pieces of additional information such as localization).
- The new (or updated) resources are imported every night into the XML database.
Searching for vinka 'wave v.' in SBLEX
### Lexicon resources

<table>
<thead>
<tr>
<th>SALDO</th>
<th>License</th>
<th>CC-BY-SA 3.0, LGPL 3.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version</td>
<td></td>
<td>development</td>
</tr>
<tr>
<td>Homepage (swe)</td>
<td><a href="http://sprakbanken.gu.se/swe/saldo">Link</a></td>
<td></td>
</tr>
<tr>
<td>Homepage (eng)</td>
<td><a href="http://sprakbanken.gu.se/eng/saldo/">Link</a></td>
<td></td>
</tr>
<tr>
<td>Downloads</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XML</td>
<td><a href="https://svn.spraakdata.gu.se/sb-arkiv/pub/lexikon/saldo/saldo.xml">Link</a></td>
<td></td>
</tr>
<tr>
<td>LMF</td>
<td><a href="https://svn.spraakdata.gu.se/sb-arkiv/pub/lfm/saldo/saldo.xml">Link</a></td>
<td></td>
</tr>
<tr>
<td>CMDI</td>
<td><a href="https://svn.spraakdata.gu.se/sb-arkiv/pub/lexikon/saldo/saldo-cmdi.xml">Link</a></td>
<td></td>
</tr>
</tbody>
</table>

### Swedish FrameNet

<table>
<thead>
<tr>
<th>License</th>
<th>CC-BY-SA 3.0, LGPL 3.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version</td>
<td>development</td>
</tr>
<tr>
<td>Homepage (swe)</td>
<td><a href="http://sprakbanken.gu.se/swe/swnet">Link</a></td>
</tr>
<tr>
<td>Homepage (eng)</td>
<td><a href="http://sprakbanken.gu.se/eng/swnet">Link</a></td>
</tr>
<tr>
<td>Downloads</td>
<td></td>
</tr>
<tr>
<td>XML</td>
<td><a href="https://svn.spraakdata.gu.se/sb-arkiv/pub/lexikon/swnet/swnet.xml">Link</a></td>
</tr>
<tr>
<td>CMDI</td>
<td><a href="https://svn.spraakdata.gu.se/sb-arkiv/pub/lexikon/swnet/swnet-cmdi.xml">Link</a></td>
</tr>
</tbody>
</table>

### Swesaurus

<table>
<thead>
<tr>
<th>License</th>
<th>CC-BY-SA 3.0, LGPL 3.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version</td>
<td>development</td>
</tr>
<tr>
<td>Homepage (swe)</td>
<td><a href="http://sprakbanken.gu.se/swe/forskning/swesaurus">Link</a></td>
</tr>
</tbody>
</table>
About openness

- Openness implies that all members of the SweFN++ project work in plain sight.
- This can be quite disconcerting at first, but we have experienced nothing but positive effects: we feel that the work has improved in terms of quality and relevance, and that the general interest of the project has increased.
Constructions
Constructions

- General rules
- Grammars
- Lexical resources
- Particular words
Constructions

Grammars

Lexical resources

general rules

particular words
Construction

Grammars

Lexical resources

general rules

constructions

particular words
Constructions

“The number of idioms and constructions that speakers know is of a comparable order of magnitude to the number of words, and the frequency of such constructions in text and conversation is very high” (Jackendoff 2007:57)
Constructicon

- SweCxn – a Swedish Constructicon
- a large-scale (eventually) collection of Swedish constructions
- a freely available resource for linguistic, lexicographic, and educational purposes, as well as for language technology applications
- an addition to the Swedish FrameNet, integrated with the linguistic resource network of Språkbanken
Why?

- fill an empirical gap
- address "a pain in the neck for NLP" (Sag et al. 2002)
- improved base for L2 education (in particular)
- enhance (e-) lexicography
- cross-linguistic application (constructicons under way for English, Japanese, and Brazilian Portuguese, hopefully with German and Spanish soon to come)
Constructions – conventionalized pairings of form and meaning/function

Grammars

Lexical resources

general rules

constructions

particular words
Constructions – conventionalized pairings of form and meaning/function

Grammars

Lexical resources

(constructions)

(particular words)

(general rules)
# Constructions – conventionalized pairings of form and meaning/function

<table>
<thead>
<tr>
<th>Degrees of schematicity</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>fully schematic</td>
<td>[V NP]$<em>{VP}$, [NP VP]$</em>{S}$, stemV-PAST (e.g. <em>walk-ed</em>, <em>smell-ed</em>)</td>
</tr>
<tr>
<td>partially schematic</td>
<td><em>the</em> [AdjP] (<em>the rich/hungry/young</em>), [time expression] <em>ago</em> (<em>six days/beers ago</em>)</td>
</tr>
<tr>
<td>fully filled and partially flexible</td>
<td><em>go</em>[tense] <em>postal</em>, <em>hit</em>[tense] <em>the road</em></td>
</tr>
<tr>
<td>fully filled and fixed</td>
<td><em>blue moon</em>, <em>by and large</em>, <em>children</em>, <em>ink</em>, <em>blue</em></td>
</tr>
</tbody>
</table>
SweCxn – A Swedish Constructicon

- similar to Berkeley cxn and FrameNet
- as simple cx descriptions as possible, to facilitate large-scale coverage
- formalized to enable LT application
- free text definition of dictionary type + simple structure sketch + annotated examples
- cx elements explicitly included in the cx definition and tagged in both definition and examples
### Example adapted from the Berkeley FrameNet Constructicon (core)

<table>
<thead>
<tr>
<th>id</th>
<th>adj_as_nom.people</th>
</tr>
</thead>
<tbody>
<tr>
<td>definition</td>
<td>A noun phrase denoting the generic ([\text{set of people}]<em>{\text{Entity}}) with a particular ([\text{property}]</em>{\text{Property}}) is formed from the ([\text{Definite_determiner the}]<em>{\text{Det}}) and an ([\text{Adjective_phrase}]</em>{\text{Property}}) that identifies that property. The noun phrase so created is morphologically plural.</td>
</tr>
<tr>
<td>structure</td>
<td>the^{1} AP</td>
</tr>
</tbody>
</table>
| construction_elements | * [Cat=Det msd=def, lu=the^{1}]  
* [Role=Property, Cat=AP]  
* [Role=Entity, Cat=GNI] |
| examples | The greatest danger for \([\text{the}]_{\text{Det}} \[\text{very young}]_{\text{Property}} \]adj_as_nom.people is to be separated from the nest because they are helpless and can quickly die of cold. In their tomb portrayals \([\text{the}]_{\text{Det}} \[\text{deceased}]_{\text{Property}} \]adj_as_nom.people are always seen as young and fit, the image of eternal youth. |
SweCxn – A Swedish Constructicon

- cx elements as feature sets
- features (roles etc.) globally defined – as separate cxn entries
- collostructional elements (cf. Stefanowitsch and Gries)
- cross-linguistic representation – preliminary approximation by frames
Constructicon entries

<table>
<thead>
<tr>
<th>id</th>
<th>name of the construction (unique ID)</th>
</tr>
</thead>
<tbody>
<tr>
<td>type</td>
<td>Cx (cf. Role, Cat)</td>
</tr>
<tr>
<td>cat</td>
<td>grammatical category of cx (N, VP etc.)</td>
</tr>
<tr>
<td>inheritance</td>
<td>more general cx from which this cx inherits</td>
</tr>
<tr>
<td>evokes</td>
<td>closest corresponding Frame (if applicable)</td>
</tr>
<tr>
<td>definition</td>
<td>free text definition, cx elements explicit and tagged</td>
</tr>
<tr>
<td>structure</td>
<td>simple schematic constituent structure</td>
</tr>
<tr>
<td>cee</td>
<td>cx evoking element</td>
</tr>
<tr>
<td>coll</td>
<td>collostructional elements (cf. Stefanowitsch &amp; Gries)</td>
</tr>
<tr>
<td>construction_elements</td>
<td>cx elements: role, cat, LU, msd (internal / external)</td>
</tr>
<tr>
<td>examples</td>
<td>annotated examples (cx, ce)</td>
</tr>
<tr>
<td>comment</td>
<td>any kind of relevant information</td>
</tr>
<tr>
<td>reference</td>
<td>source of analysis (if applicable)</td>
</tr>
<tr>
<td>id</td>
<td>adj_as_nom.people</td>
</tr>
<tr>
<td>--------</td>
<td>-------------------</td>
</tr>
<tr>
<td>type</td>
<td>Cx</td>
</tr>
<tr>
<td>cat</td>
<td>NP:plur</td>
</tr>
<tr>
<td>inheritance</td>
<td>(adjective_as_nominal)</td>
</tr>
<tr>
<td>evokes</td>
<td>n/a</td>
</tr>
<tr>
<td>definition</td>
<td>A noun phrase denoting the generic [set of people]Entity with a particular [property]Property is formed from the [Definite_determiner the]Det and an [Adjective_phrase]Property that identifies that property. The noun phrase so created is morphologically plural.</td>
</tr>
<tr>
<td>structure</td>
<td>the\textsuperscript{1} AP</td>
</tr>
<tr>
<td>cee</td>
<td>the\textsuperscript{1}</td>
</tr>
<tr>
<td>coll</td>
<td>rich\textsuperscript{1}, poor\textsuperscript{1}, elderly\textsuperscript{1}, etc.</td>
</tr>
</tbody>
</table>
| construction_elements, internal | * [Cat=Det msd=def, lu=the\textsuperscript{1}]  
|                             | * [Role=Property, Cat=AP]  
| construction_elements, external | * [Role=Entity, Cat=GNI]  
| examples | The greatest danger for [ [the]\text{Det} [very young]\text{Property} ]\text{adj\_as\_nom,people} is to be separated from the nest because they are helpless and can quickly die of cold.  
|             | In their tomb portrayals [ [the]\text{Det} [deceased]\text{Property} ]\text{adj\_as\_nom,people} are always seen as young and fit, the image of eternal youth.  
| comment | –  
| reference | Berkeley FrameNet Constructicon |
reflexiv_resultativ

type | Cx  
category | vbm  
evokes | Causation_scenario  
definition | [Någon]_{\text{Actor}} \text{eller} [\text{något}]_{\text{Theme}} \text{utför eller undergår [en aktion]}_{\text{Activity}} \text{som leder (eller antas leda) till att [aktören]}_{\text{Actor}} / [\text{temat}]_{\text{Theme}} \text{, uttryckt med reflexiv, uppnår [ett tillstånd]}_{\text{Result}}.  
structure | vb refl AP  
cee | refl  
coll | \{äta}^{1} : mätt^{1} \{supa}^{1} : full^{2} \{skrika}^{1} : hes^{1} \{springa}^{1}  

| internal | construction | elements |  
| role: name=Activity cat=vb  
| role: cx=refl name=Actor  
| role: cx=refl name=Theme  
| role: name=Result cat=AP  

| external | construction | elements |  
| role: name=Actor cat=NP  
| role: name=Theme cat=NP  

| examples |  
| [Vi åskådare]_{\text{Actor}} [springer]_{\text{Activity}} [oss]_{\text{Actor}} inte [varma]_{\text{Result}}_{\text{reflexiv}} direkt.  
| [Kornet och havren]_{\text{Theme}} får [frysa]_{\text{Activity}} [sig]_{\text{Theme}} [mogen]_{\text{Result}}_{\text{reflexiv}}  
| [Drick]_{\text{Activity}} [dig]_{\text{Actor}} [smal]_{\text{Result}}_{\text{reflexiv}} i vår.  


| http://hdl.handle.net/2077/19000  
SweFN++ demo

http://spraakbanken.gu.se