GÖTEBORGS UNIVERSITET
Språk-BANKEN
CLT

Lexical Resources

Language Technology Resources (LT2304)

Markus Forsberg
Språkbanken
University of Gothenburg

2011-09-21

GÖTEBORGS UNIVERSITET
Språk-BANKEN
CLT

Overview of today's and tomorrow's lecture

- Lexical resources for LT
- Lexical Markup Framework
- Lexical semantics
- Princeton Wordnet
- Berkeley FrameNet
- SweFN++
- (Slides borrowed from Jurafsky & Fillmore)

GÖTEBORGS UNIVERSITET
Språk-BANKEN
CLT

Lexical resources for LT

- Lexical resources refers to resources containing information about lexical units (e.g., formal description, 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).

GÖTEBORGS UNIVERSITET
Språk-BANKEN
CLT

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 / H_20
- 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_20
  - 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*

superordinate	vehicle	fruit	furniture	mammal
hyponym	car	mango	chair	dog

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

Category	Unique Forms
Noun	**117,097**
Verb	**11,488**
Adjective	**22,141**
Adverb	**4,601**

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

The adjective "bass" has 1 sense in WordNet.
1. bass[1], deep[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"

WordNet Noun Relations

Relation	Also called	Definition	Example
Hypernym	Superordinate	From concepts to superordinates	*breakfast*[1] → *meal*[1]
Hyponym	Subordinate	From concepts to subtypes	*meal*[1] → *lunch*[1]
Member Meronym	Has-Member	From groups to their members	*faculty*[2] → *professor*[1]
Has-Instance		From concepts to instances of the concept	*composer*[1] → *Bach*[1]
Instance		From instances to their concepts	*Austen*[1] → *author*[1]
Member Holonym	Member-Of	From members to their groups	*copilot*[1] → *crew*[1]
Part Meronym	Has-Part	From wholes to parts	*table*[2] → *leg*[3]
Part Holonym	Part-Of	From parts to wholes	*course*[7] → *meal*[1]
Antonym		Opposites	*leader*[1] → *follower*[1]

WordNet Verb Relations

Relation	Definition	Example
Hypernym	From events to superordinate events	$fly^9 \rightarrow travel^5$
Troponym	From a verb (event) to a specific manner elaboration of that verb	$walk^1 \rightarrow stroll^1$
Entails	From verbs (events) to the verbs (events) they entail	$snore^1 \rightarrow sleep^1$
Antonym	Opposites	$increase^1 \Longleftrightarrow decrease^1$

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^1, fool^2, gull^1, mark^9, patsy^1, fall guy^1, sucker^1, soft touch^1, mug^2}

- Each of these senses share this same gloss
- Thus for WordNet, the meaning of this sense of chump *is* this list.

GÖTEBORGS UNIVERSITET

Språkbanken

CLT

Princeton Wordnet: final comments

- 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.

GÖTEBORGS UNIVERSITET

Språk-BANKEN

CLT

Sense inventory

- 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)

GÖTEBORGS UNIVERSITET

Språk-BANKEN

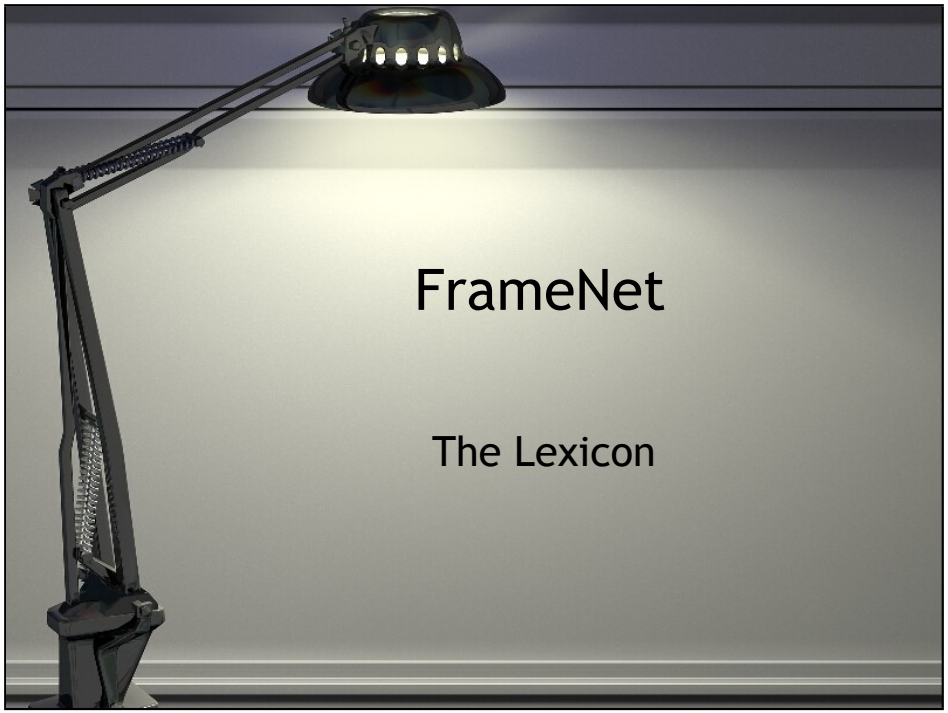
CLT

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)

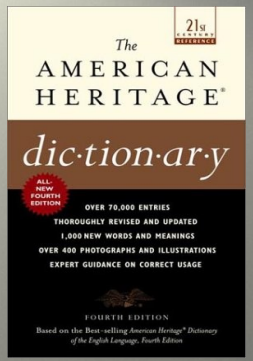


FrameNet is a lexicography project. That means we're making a dictionary.

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

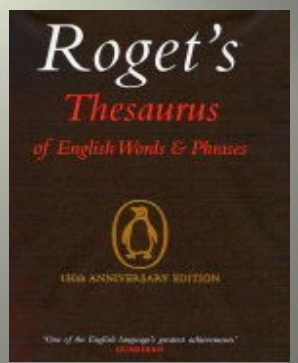
What's wrong with dictionaries?

- Dictionaries only indirectly give access to the conceptual structures underlying word meanings. Human intelligence is required to find connections hidden in defining phrases.



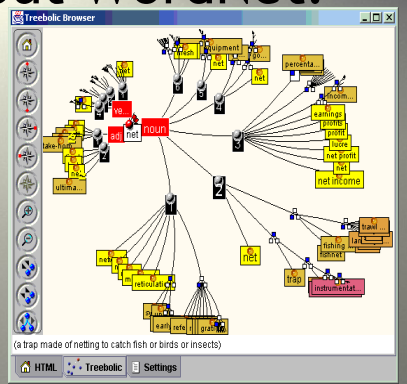
Are online thesauri helpful?

- Thesauri shows us <u>that</u> certain groups of words are semantically related, but only indirectly <u>how</u> they are related; and they show nothing of a word's combinatorial behavior.



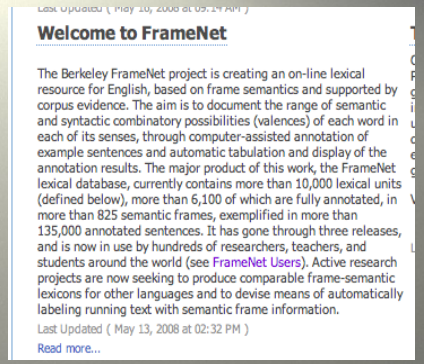
What about WordNet?

- 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 rhe words.



http://famenet.icsi.berkeley.edu

- 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.



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 fully annotated, in more than 825 semantic frames, exemplified in more than 135,000 annotated sentences. It has gone through three releases, and is now in use by hundreds of researchers, teachers, and students around the world (see FrameNet Users). Active research projects are now seeking to produce comparable frame-semantic lexicons for other languages and to devise means of automatically labeling running text with semantic frame information.

Last Updated (May 13, 2008 at 02:32 PM)

Read more...

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 **combinatory 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:

  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'

customer

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?

land

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.

chicken *(mass noun)*

- 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.**

Back to Frames

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 <u>in</u>voke** (summon up from their memory) frames, to make sense of their **experience,** linguistic or otherwise.
- a cognitive act

- **Words <u>e</u>voke** categories and knowledge structures that shape interpreters' understanding of a text.
- a cognitive experience

Warning: this 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

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

1. Buildings in California risk destruction by earthquakes.
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:

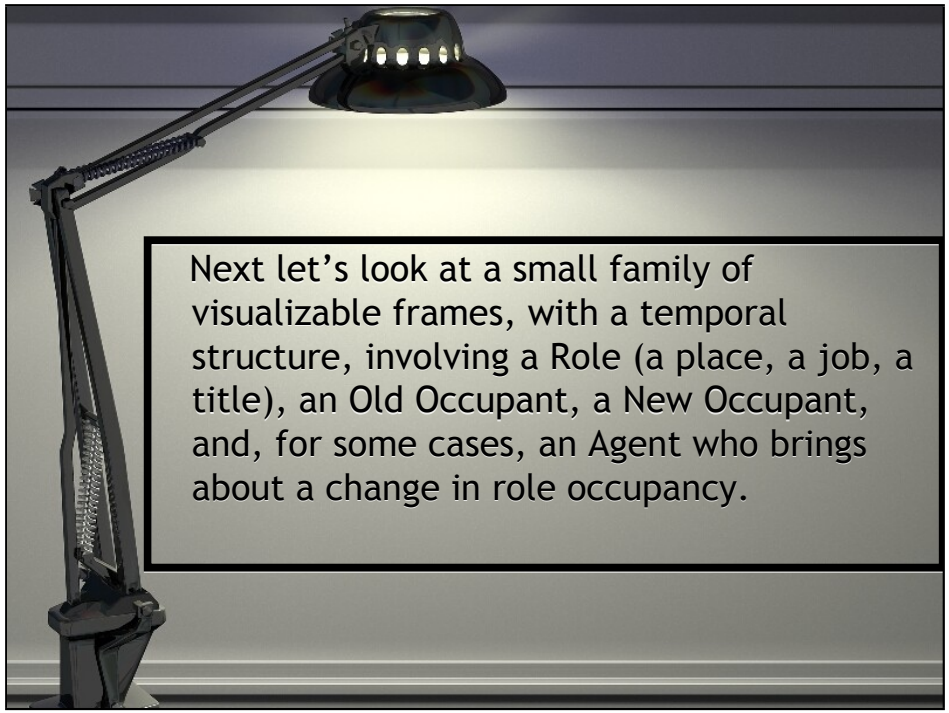
Speaker, Addressee, Mystery, Account

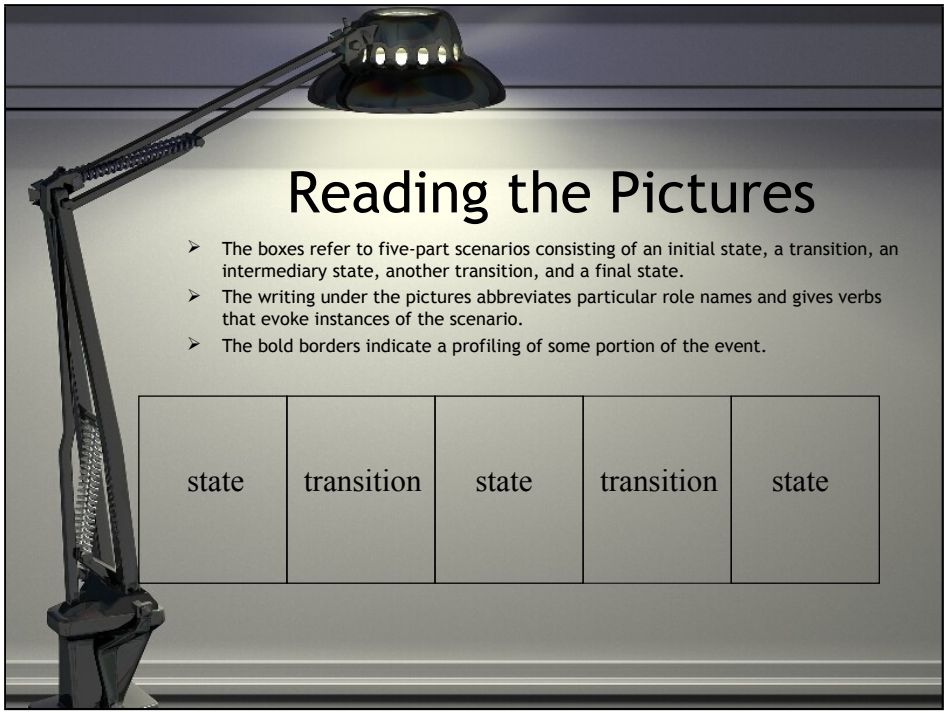
- The coach explained the problem to the team.
- The coach explained that they hadn't learned the maneuvers.
- What's your explanation of these facts?
- The defense lawyer *gave* an inadequate explanation.

Cognition.explanation:

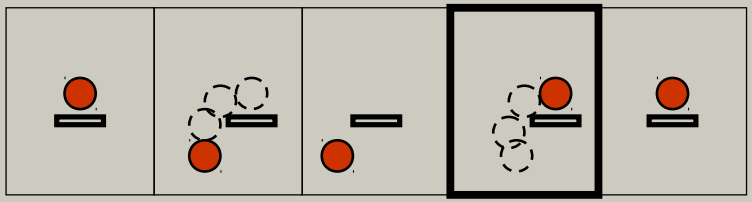
Mystery, Account

- What can explain these facts?
- A history of unrestricted logging explains the erosion pattern we see here.





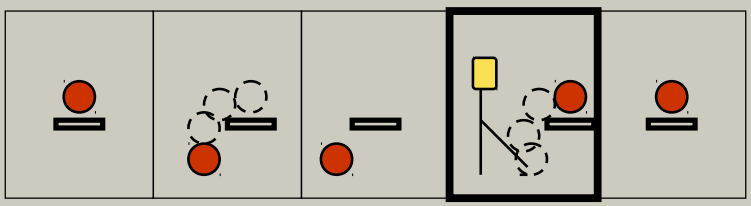
The first examples have only one “thing”, and it just goes away and comes back.



X; return, go back, come 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.*



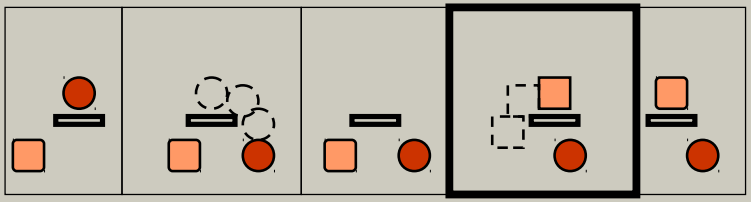
A, X; return, replace, put back

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.



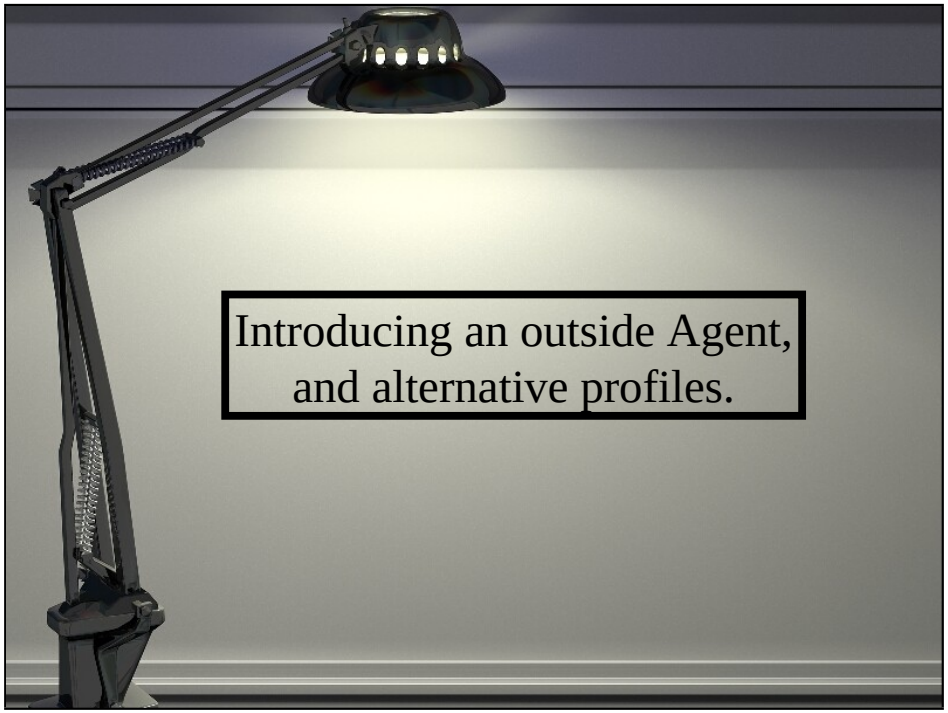
N, O; replace, substitute (for)

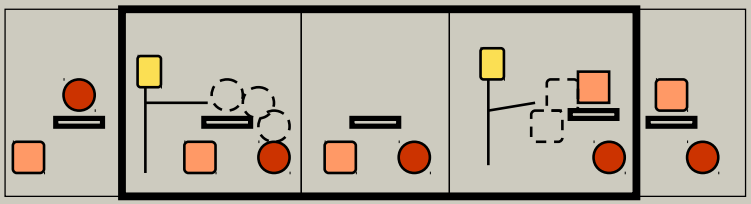
Jim has replaced me.

Jim has replaced me on your examination committee.

Nobody will ever replace you in my heart.

You are irreplaceable.

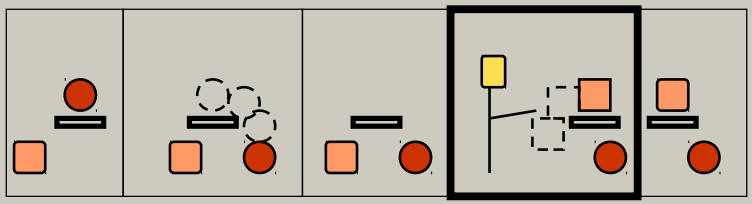




A, N, O; replace (O with N); substitute (N for O)

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.

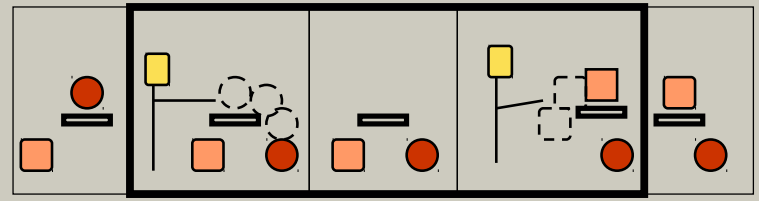
Here a two-step act is profiled: removal and placement.



A, N, O; replace (O with N); substitute (N for O)

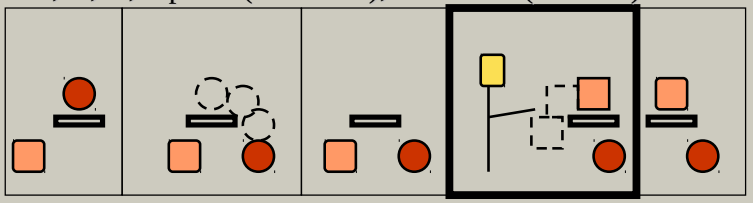
Three years after Smith retired, the department finally replaced him.
Somebody stole my bicycle; it was a year before I could afford to replace it.

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.

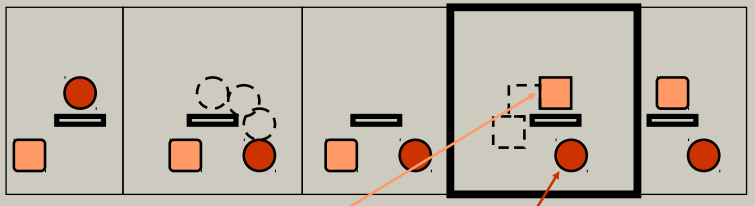


A, N, O; replace (O with N); substitute (N for O)

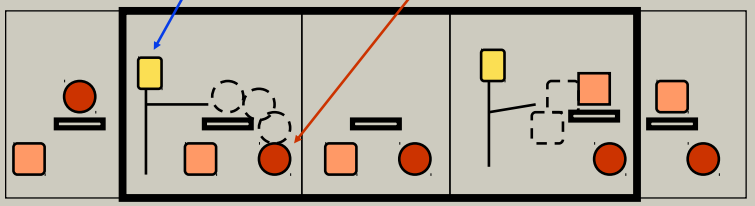
A, N, O; replace (O with N); substitute (N for O)

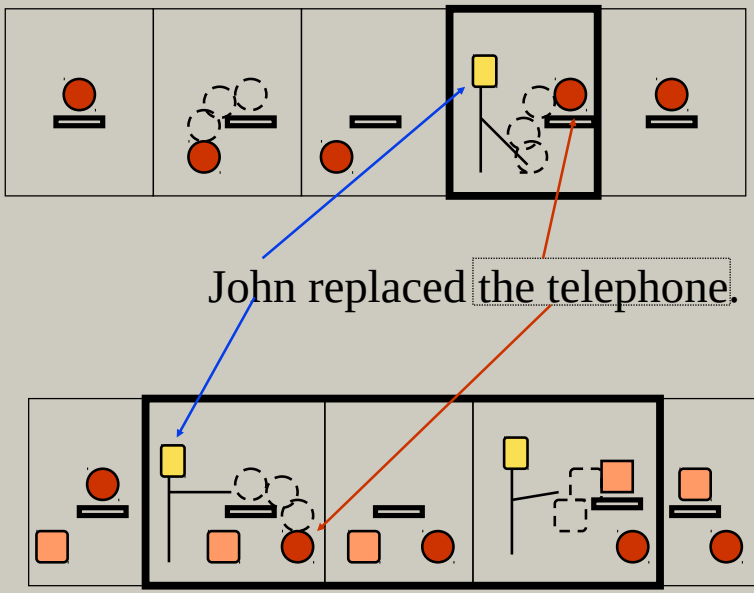


Some ambiguities



John replaced me.





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

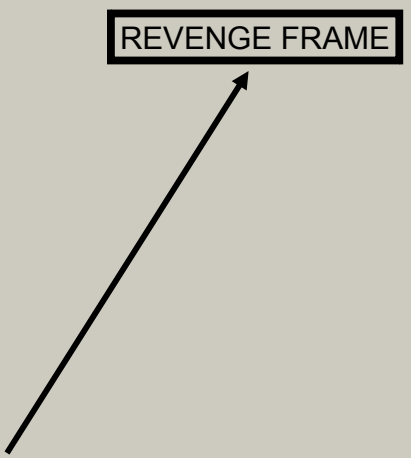
c) C's action is carried out independently of any legal or other institutional setting.

(?) B and C may or may not be the same individual.

Suppose we're working on *avenge*.
We find sentences that show it in use.

We avenged the insult by setting fire to his village.

Identify its frame.



We avenged the insult by setting fire to his village.

Characterize the elements of the frame.

REVENGE FRAME
frame elements

Avenger
Offender
Injury
Injured Party
Punishment

We avenged the insult by setting fire to his village.

Tag phrases that express frame elements.
Do this for many examples of each target.

REVENGE FRAME
annotation

Avenger
Offender - unexpressed
Injury
Injured Party - unexpressed
Punishment

Notice that some FEs do not get expressed.

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.

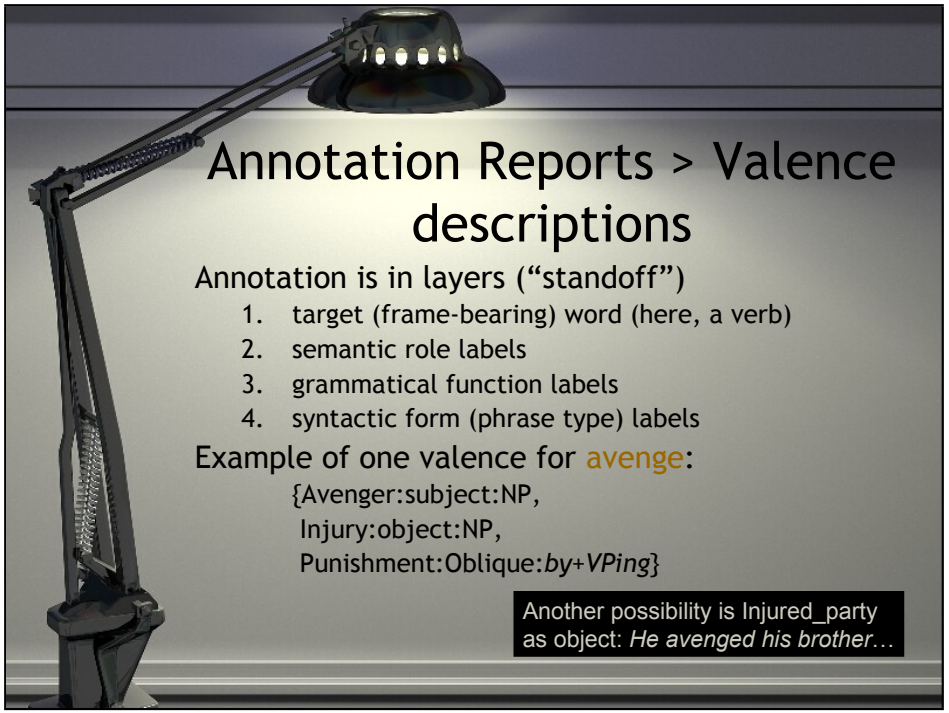
REVENGE FRAME
wordlist

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,
Injury:object:NP,
Punishment:Oblique:*by+VPing*}

Another possibility is Injured_party as object: *He avenged his brother…*

GÖTEBORGS UNIVERSITET
Språk-BANKEN
CLT

SweFN++

Language Technology Resources (LT2304)

Markus Forsberg
Språkbanken
University of Gothenburg

2011-09-22

SweFN++ in sum

- the SweFN++ plans involve:
  - reuse of existing free lexical resources
  - building a Swedish framenet
  - 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

methodology, 2

- 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 resources with associative semantic relations.

wordforms	1,779,314
senses	116,279
lemgrams (lexemes)	113,217
paradigms	1,290
PoS	37

GÖTEBORGS UNIVERSITET
Språk-BANKEN
CLT

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)

GÖTEBORGS UNIVERSITET
Språk-BANKEN
CLT

SALDO

lex:	**fisk**
fm:	djur
fp:	vatten
PRIM:	abborre benfisk berggylta braxen bröding firre glasbult gylta[2] 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 nototenioid nätting rom[2] sarv sill skoläst[2] smörbult stör[2] tobakspipa[2] tonfisk torsk ål
afrikansk:	ciklidfisk
andas:	gäl
brosk:	broskfiskar
damm:	karp
djuphav:	djuphavsfisk
flyga:	flygfisk
giftig:	drakhuvudfisk
grå:	gråsej
hav:	saltvattensfisk
hästliknande:	sjöhäst
insjö:	insjöfisk

GÖTEBORGS UNIVERSITET
Språk-BANKEN
CLT

SALDO, 2

SALDO

skicka

lex:	**PRIM**
fm:	*
fp:	*
mf(51):	**PRIM:** all annan använda att bara bra den en[2] fort framme färg för[2] förbi före genom göra ha hur hända i[2] ja just kunna ljud ljus med men mycken måste namn natur när och om om[2] på rak röra säga tal till tänka vad var vara varm vem veta vid vilja öppen
pf(42480):	*

GÖTEBORGS UNIVERSITET
Språk-BANKEN
CLT

SALDO, 3

SALDO

skicka

lex:	**spelautomat**
fm:	automat
fp:	hasardspel
mf(2):	**PRIM:** arkadspel enarmad_bandit
pf(1):	**vinst:** jackpot

SALDO

skicka

word class	nnm
inherent feature(s)	u
inflection table	
sg indef nom	enarmad bandit
sg indef gen	enarmad bandits
sg def nom	enarmade banditen
sg def gen	enarmade banditens
pl indef nom	enarmade banditer
pl indef gen	enarmade banditers
pl def nom	enarmade banditerna
pl def gen	enarmade banditernas

GÖTEBORGS UNIVERSITET
Språk-BANKEN
CLT

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

GÖTEBORGS UNIVERSITET
Språk-BANKEN
CLT

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**

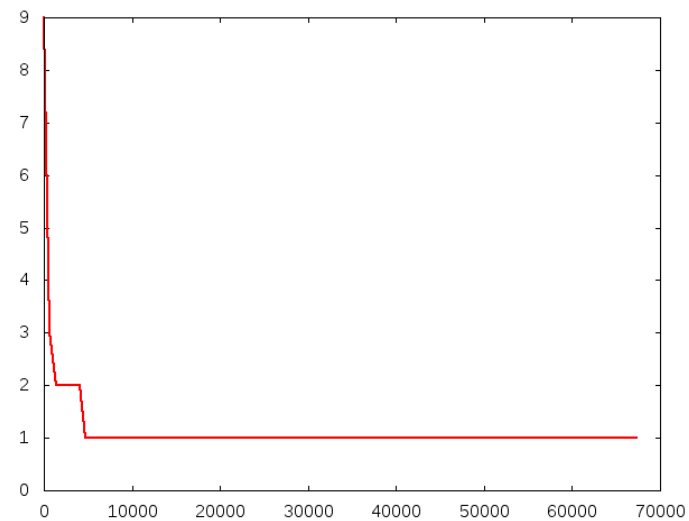
GÖTEBORGS UNIVERSITET
Språk-BANKEN
CLT

George Kingsley Zipf (1902–1950)

(from Wikipedia)

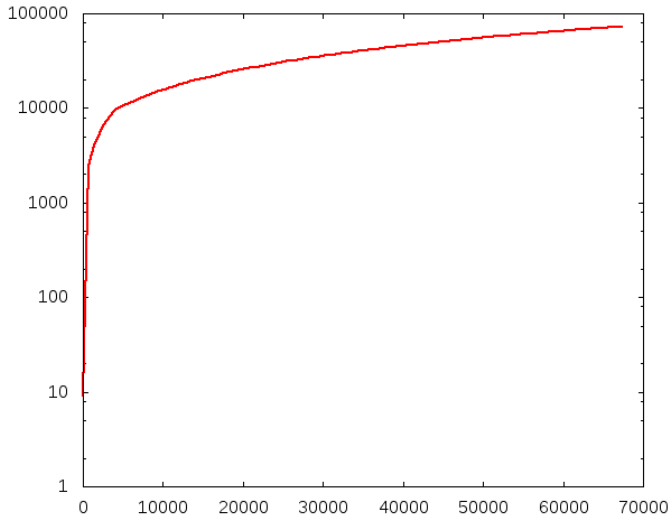
GÖTEBORGS UNIVERSITET
Språk-BANKEN
CLT

senses/lemma in SALDO



GÖTEBORGS UNIVERSITET
Språk-BANKEN
CLT

lemmas ⇒ senses in SALDO



GÖTEBORGS UNIVERSITET
Språk-BANKEN
CLT

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**

GÖTEBORGS UNIVERSITET
Språk-BANKEN
CLT

SweFN++ demo

- http://spraakbanken.gu.se