Lexical Semantics

Dana Dannélls
dana.dannells@svenska.gu.se

Språkbanken, Department of Swedish Language
University of Gothenburg

12 February 2014

Today's lecture

Lexical Semantics

- Word senses
- Word sense relations
- Thematic roles
- Lexical semantics resources

Computational lexical semantics

- Word Sense Disambiguation
- WSD evaluation
- WSD methods

Lexical Semantics

What is a word

Lemma

The basic lexical unit

- grammatical form for representing a *lexeme*
  - noun lemmas are often singular indefinite (e.g *dog*)
  - verb lemmas are often in infinitive form (e.g *sing*)
- possible variants of a lemma are called *word forms* (e.g *dogs*, *sang*, *sung*)

Lemmatization

The process of mapping a wordform to a lemma

- often ambiguous and depends on the context (e.g *found* can be mapped to *find* 'to locate' or to *found* 'to create an institution')
- similar to Part Of Speech (POS) tagging

Word Senses

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

The senses of the word might not have any particular semantic connection – *Homonymy*

bank^1: "financial institution"
bank^2: "sloping mound" (unrelated with bank^1)

The senses of the word have a semantic connection – *Polysemy*

bank^3: "biological repository" (related with bank^1)

There is a systematic connection between the senses of the word – *Metonymy*

bank^4: "the building belonging to a bank" (bank^1)

Major relations between senses

Synonymy

Antonymy

Hyponymy

Meronymy

Synonymy

The relation of semantic identity between words. Two senses of two different words are synonymous if they can be substituted in a given context.

couch – sofa
vomit – throw up
big – large
car – automobile
water – H2O

But not all synonymous words are substituted in every contexts.

big sister – large sister
a glass of water – a glass of H2O

Antonymy

The relation of oppositeness between words. Two senses can be antonyms if they define a binary opposition or are at opposite ends of some scale.

in – out
fast – slow
big – small
hot – cold
good – bad
dead – alive

Antonymy is not negation!

big $\neq$ not small
hot $\neq$ not cold

Hyponymy (is-a)

The relation that holds between a more general and a more specific word. The more specific word is called *hyponym* and the more general word is called *hypernym* (sometimes also 'hyperonym').

hyponym		**hypernym**
car	is-a	vehicle
mango	is-a	fruit
cat	is-a	mammal
mammal	is-a	animal

Other terms for hyponym and hypernym are superordinate and subordinate.

Hyponymy (is-a)

Hyponymy is usually a transitive relation:

> if A is hyponym of B and B is hyponym of C then
> A is hyponym of C

A hierarchical structure of hyponyms and hyperonyms is a *taxonomy*

Meronymy (part-of)

The relation that holds between two words that are part of each other. The word that is the part of the other word is called *meronym*, the word that contains the other word is called *holonym*.

meronym		**holonym**
leg	part-of	chair
wheel	part-of	car
elbow	part-of	arm

Meronymy is not a unitary relation, it comprises a number of subtypes.

pilot	member-of	crew
gin	substance-of	dry martini
keyboard	component-of	computer
wood	material-of	door

WordNet

A large-scale electronic lexicon developed in 1985 at Princeton University

- organized according to word sense relations
  - on average 2.95 senses for polysemous words
  - between 2.50 for adverbs, and 3.57 for verbs
- nouns, verbs, adjectives and adverbs are grouped into synonym sets (synsets)
- Current version is 3.1 (December 2013)

WordNet synsets

Basic semantic unit is the *synset*:

- the set of near-synonyms for a word sense:
  {day^2, daytime^1, daylight^1}
- each synset has a definition:
  "the time after sunrise and before sunset while it is light outside"
- many synsets also have examples:
  "the dawn turned night into day", "it is easier to make the repairs in the daytime", "there are 30,000 passengers per day"

WordNet synsets

Synsets can be very large.

synset: {batch^{2}, deal^{3}, flock^{3}, good deal^{1}, great deal^{1}, hatful^{1}, heap^{2}, lot^{1}, mass^{2}, mess^{6}, mickle^{1}, mint^{1}, mountain^{2}, muckle^{1}, passel^{1}, peck^{1}, pile^{2}, plenty^{2}, pot^{5}, quite a little^{1}, raft^{2}, sight^{7}, slew^{1}, spate^{1}, stack^{2}, tidy sum^{1}, wad^{2}}

definition: “(often followed by ’of’) a large number or amount or extent”

examples: “a batch of letters”, “a deal of trouble”, “a lot of money”, “he made a mint on the stock market”, “see the rest of the winners in our huge passel of photos”, “it must have cost plenty”, ...

WordNet synsets

Synsets can be very similar.

coach[3]: “a railcar where passengers ride”

coach[4]: “a carriage pulled by four horses with one driver”

coach[5]: “a vehicle carrying many passengers; used for public transport”

WordNet synsets

`http://wordnetweb.princeton.edu/perl/webwn`

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"

WN 3.0 entry for the noun *bass*.

WordNet relations (nouns)

Relation	Also Called	Definition	Example
Hypernym	Superordinate	From concepts to superordinates	$breakfast^1 \rightarrow meal^1$
Hyponym	Subordinate	From concepts to subtypes	$meal^1 \rightarrow lunch^1$
Instance Hypernym	Instance	From instances to their concepts	$Austen^1 \rightarrow author^1$
Instance Hyponym	Has-Instance	From concepts to concept instances	$composer^1 \rightarrow Bach^1$
Member Meronym	Has-Member	From groups to their members	$faculty^2 \rightarrow professor^1$
Member Holonym	Member-Of	From members to their groups	$copilot^1 \rightarrow crew^1$
Part Meronym	Has-Part	From wholes to parts	$table^2 \rightarrow leg^3$
Part Holonym	Part-Of	From parts to wholes	$course^7 \rightarrow meal^1$
Substance Meronym		From substances to their subparts	$water^1 \rightarrow oxygen^1$
Substance Holonym		From parts of substances to wholes	$gin^1 \rightarrow martini^1$
Antonym		Semantic opposition between lemmas	$leader^1 \Longleftrightarrow follower^1$
Derivationally Related Form		Lemmas w/same morphological root	$destruction^1 \Longleftrightarrow destroy^1$

WordNet relations (verbs)

Relation	Definition	Example
Hypernym	From events to superordinate events	$fly^9 \rightarrow travel^5$
Troponym	From events to subordinate event (often via specific manner)	$walk^1 \rightarrow stroll^1$
Entails	From verbs (events) to the verbs (events) they entail	$snore^1 \rightarrow sleep^1$
Antonym	Semantic opposition between lemmas	$increase^1 \Longleftrightarrow decrease^1$
Derivationally Related Form	Lemmas with same morphological root	$destroy^1 \Longleftrightarrow destruction^1$

WordNet relations hierarchy

```
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
```

Hyponymy chains for two separate senses of the lemma *bass*.

SALDO

An electronic semantic lexicon for Swedish developed in 2008 at Språkbanken

- organized according to an associative network with a hierarchical structure
- each word is associated with:
  - a primary descriptor (obligatory)
  - a secondary descriptor (optional)
- the morphological description of a word contains:
  - dictionary information (lemma, word class, paradigm form, lemma id)
  - inflection information (paradigm table)

SALDO word senses

http://spraakbanken.gu.se/saldo

stol

sammansättningsanalys

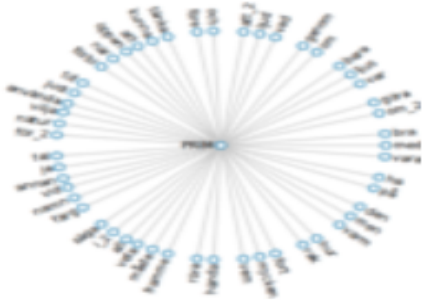
sticka_under_stol	dölja^2	sticka_under_stol (vbm)	korpus
stol	möbel + sitta	stol (nn)	korpus
stol^2	ämbete	stol (nn)	korpus

Semantic relations in SALDO

stol

⇧[8] | möbel | ↑ | sitta

stol (nn)



[korpus]

⇩ 32			↓ 4	
	PRIM	armstöd pall stolsits		dyna stolsdyna
	armstöd	armstol		massera massagestol
	bar[2]	barstol		rad stolsrad
	bekväm	länstol		rygg stolsrygg
	besökare	besöksstol		
	bäbis	babysitter		
	bära	bärstol		
	flygplan	flygstol		
	fläta	korgstol		
	frisörsalong	frisörstol		
	gunga	gungstol		
	hjul	rullstol		

Thematic roles

Tend to capture the meaning of the syntactic arguments for a sentence.

"John broke the window with a rock"

AGENT – the subject
John

THEME – direct object
window

INSTRUMENT – indirect object
rock

But despite this potential benefit, it has proved quite difficult to come up with a standard set of roles.

Proposition Bank (PropBank)

A resource of sentences annotated with semantic roles.

Core roles:

Arg0: Prototypical Agent
Arg1: Prototypical Patient
Arg2: indirect object/instrument/attribute/end state
Arg3: start point/instrument/attribute
Arg4: end point

$[_{Arg0}$ Big Fruit Co.] increased $[_{Arg1}$ the price of bananas]

The PropBank semantic roles can be useful in convening shallow semantic information about verbal arguments.

FrameNet (FN)

FrameNet is a lexical resource that is based on the theory of frame semantics.

Describes frames by analysing Lexical Units (a pairing of a word with a meaning).

A frame in FrameNet consists of

- A definition
- Frame Elements (FEs)
- Lexical Units (LUs)
- Sentences that are extracted from a corpus

The Berkeley FrameNet (BFN) project

An electronic lexical resource for English.

Current Version 1.5 contains over 1,000 frames.

`https://framenet.icsi.berkeley.edu/fndrupal/`

A semantic frame from the BFN database

Frame: **Cause_to_be_dry**

Definition: An Agent causes a Dryee (either a surface or an entire entity, inside and out) to become dry. This should not include examples like “drying tears” or “drying spills” as these are in the Removing frame.

Core FEs: Agent, Cause, Dryee

Non-Core FEs: Degree, Duration, Instrument, Manner, Means, Place, Purpose, Subregion, Temperature, Time

LUs: anhydrate.v, dehumidification.n, dehumidify.v, dehydrate.v, desiccate.v, desiccation.n, dry_off.v, dry_out.v, dry_up.v, dry.v

The Swedish FrameNet (SweFN)

SweFN is part of the Swedish FrameNet++ project.

An electronic lexical resource for Swedish.

Follows the same principles as in Berkeley FrameNet.

The Swedish FrameNet database has more than 1000 frames and over 26,000 LUs.

`http://spraakbanken.gu.se/eng/swefn`

It is integrated with other lexical resources both modern and historical.

`http://spraakbanken.gu.se/karp/`

A semantic frame from the SweFN database

Frame: **Cause_to_be_dry**

Domain: Gen

Core FEs: Agent, Cause, Dryee

Non-Core FEs: Degree, Duration, Instrument, Manner, Means, Place, Purpose, Subregion, Temperature, Time

LUs: vb: avfukta, föna, hässja, torka[3], torka bort[2], torktumla, torrlägga, ugnstorka; nn: hässjning, torktumlande, torrläggning; vbm: torka ut[2]

sms: Cause+LU luft|torka, sol|torka, Instrument+LU ugns|torka

Computational lexical semantics

Word Sense Disambiguation (WSD)

WSD is the process of determining the right sense of a word in its context.

Who need WSD?

- Machine Translation
- Question Answering
- Information Retrieval
- Text classification
- Textual entailment
- Parsing
- Summarization
- etc.

Two variants of WSD

Lexical sample

- limited nr. of words to disambiguate
- slightly artificial problem

All-words task

- decide the word sense of all words in the input (like POS tagging with larger set of tags)
- a much harder problem than the lexical sample
- but also much more useful

WSD approaches

Supervised machine leaning

Dictionary and thesaurus methods

Unsupervised machine leaning: Bootstrapping techniques

Feature vectors for supervised leaning

Feature vector

A set of extracted numeric or nominal values representing different linguistic context features.

Context: "An electric guitar and **bass** player stand off to one side, not really part of the scene, just as a sort of nod .."

Collocational feature vector

keeps sentence word order

$[w_{i-2}, POS_{i-2}, w_{i-1}, POS_{i-1}, w_{i+1}, POS_{i+1}, w_{i+2}, POS_{i+2}]$
$\Rightarrow$ [guitar, NN, and, CC, player, NN, stand, VB]

Bag-of-words feature vector

uses context words without any particular word order

[fishing, big, sound, player, fly, rod, pound, double, runs, playing, guitar, band]
$\Rightarrow$ [0,0,0,1,0,0,0,0,0,0,1,0]

Naive Bayes classifier

Choose the best sense s' out of a set of possible senses s for a feature vector f_n.

$$\begin{aligned} \mathrm{s}' &= \mathrm{argmax}_s\ \mathrm{P}(\mathrm{s} \mid \mathrm{f}_1 \ldots \mathrm{f}_n) \\ &= \mathrm{argmax}_s\ \mathrm{P}(\mathrm{s})\ \mathrm{P}(\mathrm{f}_1 \ldots \mathrm{f}_n \mid \mathrm{s}) \\ &\approx \mathrm{argmax}_s\ \mathrm{P}(\mathrm{s}) \prod_j \mathrm{P}(\mathrm{f}_j \mid \mathrm{s}) \end{aligned}$$

Maximum Likelihood Estimation (MLE)

Count each of the individual feature probabilities for a given sense.

$$\mathrm{P}(\mathrm{f}_j \mid \mathrm{s}) = \frac{Count(f_j, s)}{Count(s)}$$

$$\mathrm{P}(\mathrm{guitar} \mid \mathrm{bass}^1) = \frac{3}{60} = 0.05$$

Naive Bayes classifier and HMM

Naive Bayes classifier

$$\begin{aligned} \mathrm{s}' &= \mathrm{argmax}_s\ \mathrm{P}(\mathrm{s} \mid \mathrm{f}_1 \ldots \mathrm{f}_n) \\ &= \mathrm{argmax}_s\ \mathrm{P}(\mathrm{s})\ \mathrm{P}(\mathrm{f}_1 \ldots \mathrm{f}_n \mid \mathrm{s}) \\ &\approx \mathrm{argmax}_s\ \mathrm{P}(\mathrm{s}) \prod_j \mathrm{P}(\mathrm{f}_j \mid \mathrm{s}) \end{aligned}$$

HMM tagging

$$\begin{aligned} \mathsf{T}' &= \mathrm{argmax}_{\mathsf{T}}\ \mathrm{P}(\mathsf{T} \mid \mathrm{w}_1 \ldots \mathrm{w}_n) \\ &= \mathrm{argmax}_{\mathsf{T}}\ \mathrm{P}(\mathsf{T})\ \mathrm{P}(\mathrm{w}_1 \ldots \mathrm{w}_n \mid \mathsf{T}) \\ &\approx \mathrm{argmax}_{\mathsf{T}}\ \mathrm{P}(\mathsf{T}) \prod_j \mathrm{P}(\mathrm{w}_j \mid \mathsf{T}_j) \end{aligned}$$

Decision list classifier

A sequence of tests on the feature vector

can be learned automatically

NLTK uses decision trees instead

Rule		Sense
fish within window	⇒	**bass**^{1}
striped bass	⇒	**bass**^{1}
guitar within window	⇒	**bass**^{2}
bass player	⇒	**bass**^{2}
piano within window	⇒	**bass**^{2}
tenor within window	⇒	**bass**^{2}
sea bass	⇒	**bass**^{1}
play/V *bass*	⇒	**bass**^{2}
river within window	⇒	**bass**^{1}
violin within window	⇒	**bass**^{2}
salmon within window	⇒	**bass**^{1}
on bass	⇒	**bass**^{2}
bass are	⇒	**bass**^{1}

An abbreviated decision list for disambiguating the fish sense of *bass* from the museum sense.

WSD evaluation

Baseline

- most frequent sense
- the Lesk algorithm

Ceiling

- human inter-annotator agreement

  for WordNet-style senses: 75–80%
  for more coarse-grained, binary senses: 90%

Dictionary methods: Simplified Lesk

Disambiguate “bank” in this sentence:
“The bank can guarantee deposits will eventually cover future tuition costs because it invests in adjustable-rate mortgage securities.”

Given these two definitions:

bank^1 = a financial institution that accepts deposits and channels the money into lending activities.
Examples: “he cashed a check at the bank”, “that bank holds the mortgage on my home”

bank^2 = sloping land (especially the slope beside a body of water).
Examples: “they pulled the canoe up on the bank”, “he sat on the bank of the river and watched the currents”

$\Rightarrow$ bank^1 has 2 overlapping non-stopwords; bank^2 has 0

The Simplified Lesk algorithm

function SIMPLIFIED LESK(*word*, *sentence*) **returns** best sense of *word*

best-sense ← most frequent sense for *word*
max-overlap ← 0
context ← set of words in *sentence*
for each *sense* **in** senses of *word* **do**
  signature ← set of words in the gloss and examples of *sense*
  overlap ← COMPUTEOVERLAP(*signature*, *context*)
  if *overlap* > *max-overlap* **then**
    max-overlap ← *overlap*
    best-sense ← *sense*
end
return(*best-sense*)

Selectional restrictions

One of the earliest knowledge source for WSD.

Selectional preference strength because of semantic restrictions on the verb arguments:

$$\text{THEME} = [+\text{FOOD}] \text{ for the verb } \textit{eat}$$

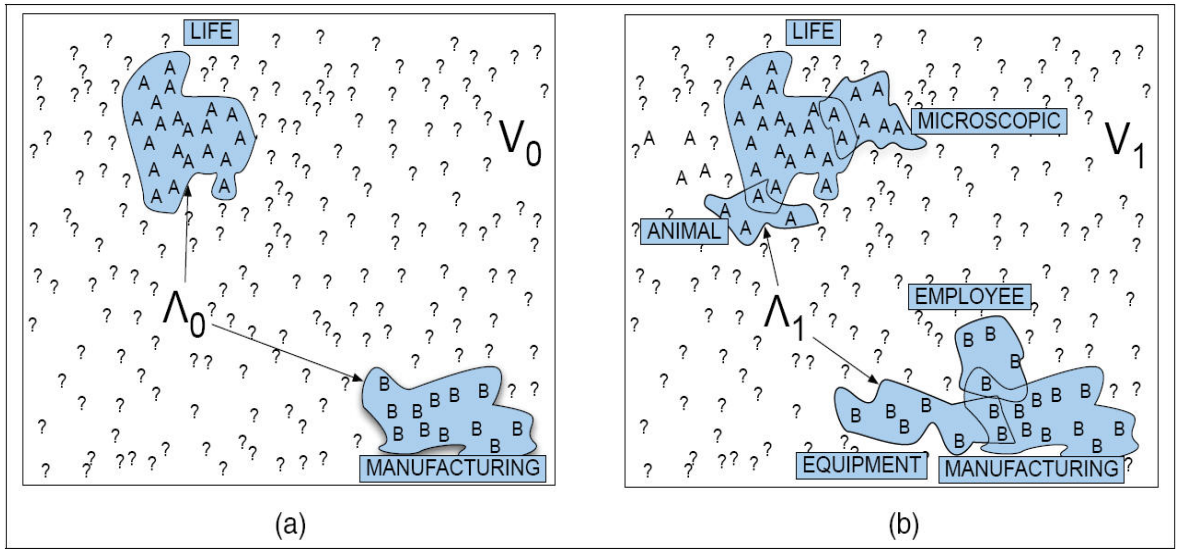
Selection associations for verbs and some WN semantic classes of their direct objects:

	Direct Object		Direct Object	
Verb	**Semantic Class**	**Assoc**	**Semantic Class**	**Assoc**
read	WRITING	6.80	ACTIVITY	-.20
write	WRITING	7.26	COMMERCE	0
see	ENTITY	5.79	METHOD	-0.01

Bootstrapping for WSD

Word sense disambiguation in a large completely unannotated data.

Only a small hand-labeled annotated data (a small seed set) is needed

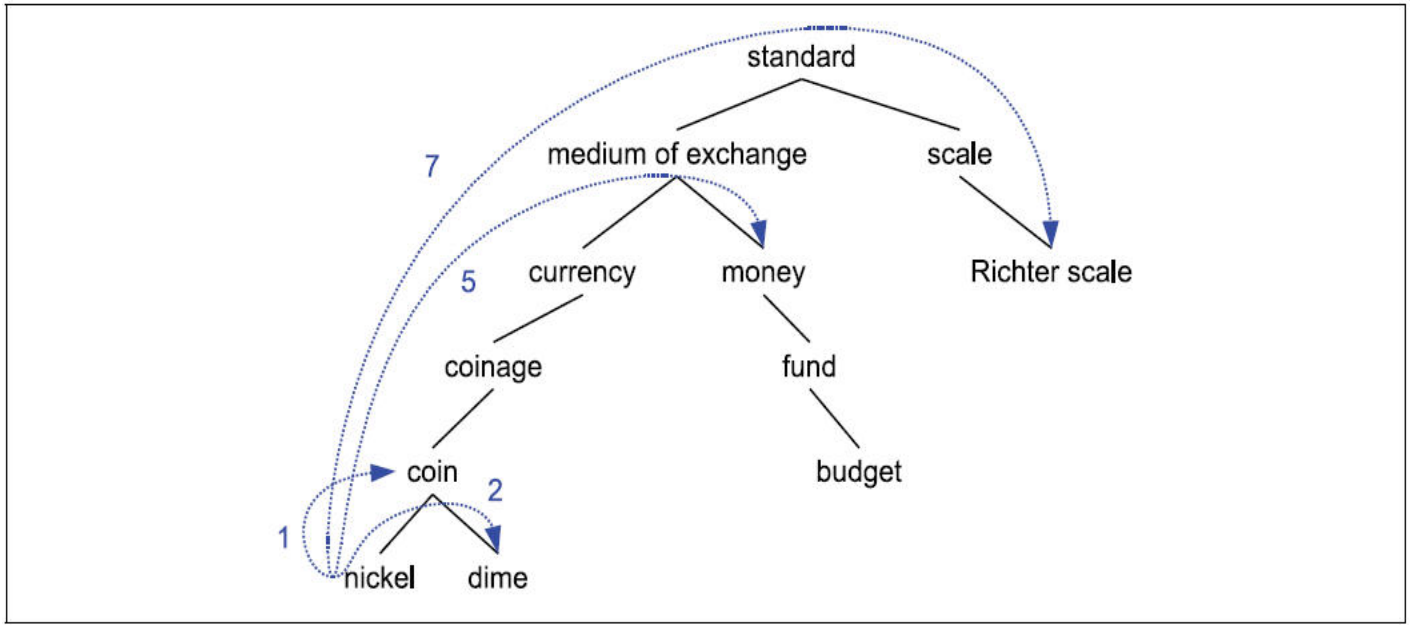


The Yarowsky algorithm disambiguating *plant* at two stages.

Word similarity

Two words are similar if they share more features of meaning or are near synonyms.

- Thesaurus methods
- Distributional methods



A fragment of the WN hpernym hierarchy.

Semantic Role Labeling (SRL)

The process of finding the semantic roles for each predicate in a sentence automatically.

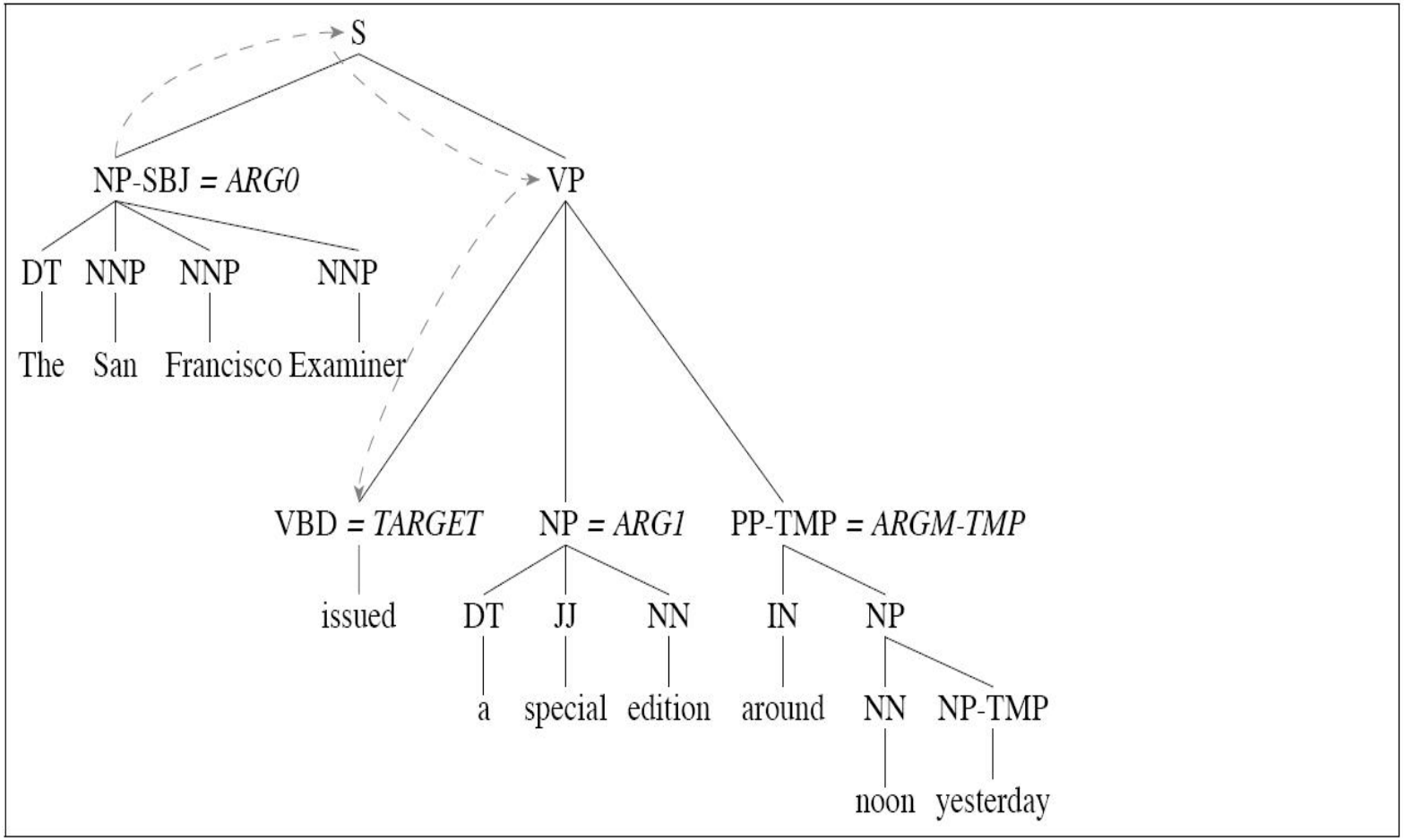
Useful resources are FrameNet and PropBank

```
function SEMANTICROLELABEL(words) returns labeled tree

    parse ← PARSE(words)
    for each predicate in parse do
        for each node in parse do
            featurevector ← EXTRACTFEATURES(node, predicate, parse)
            CLASSIFYNODE(node, featurevector, parse)
```

An example of an algorithm that can be trained on semantic labeled data.

SRL example



Parse tree for a PropBank sentence with the PropBank argument labels.