

Introduction to lexical resources

Bolette Sandford Pedersen
Centre for Language Technology
University of Copenhagen
bspedersen@hum.ku.dk



Outline 1

- Why are lexical semantic resources relevant for semantic annotation?
- What is the relation between traditional lexicography and lexical resources for annotation purposes?



Outline 2

- WordNet, primary focus on vertical, taxonomical relations
- VerbNet, primary focus on horizontal, syntagmatic relations, thematic roles
- PropBank, primary focus on horizontal, syntagmatic relations, argument structure
- FrameNet, primary focus on horizontal, syntagmatic relations, thematic roles (= frame elements)
- PAROLE/SIMPLE - Semantic Information for Multifunctional, PluriLingual Lexicons - based on the Generative Lexicon (Pustejovsky 1995)



Outline 2

- WordNet, primary focus on vertical, taxonomical relations
- VerbNet, primary focus on horizontal, syntagmatic relations, thematic roles
- PropBank, primary focus on horizontal, syntagmatic relations, argument structure
- FrameNet, primary focus on horizontal, syntagmatic relations, thematic roles
- PAROLE/SIMPLE - Semantic Information for Multifunctional, PluriLingual Lexicons - based on the Generative Lexicon (Pustejovsky 1995)



Outline 2

- WordNet, primary focus on vertical, taxonomical relations
- VerbNet, primary focus on horizontal, syntagmatic relations, thematic roles
- PropBank, primary focus on horizontal, syntagmatic relations, argument structure
- FrameNet, primary focus on horizontal, syntagmatic relations, thematic roles
- PAROLE/SIMPLE - Semantic Information for Multifunctional, PluriLingual Lexicons - based on the Generative Lexicon (Pustejovsky 1995)



Outline 2

- WordNet, primary focus on vertical, taxonomical relations
- VerbNet, primary focus on horizontal, syntagmatic relations, thematic roles
- PropBank, primary focus on horizontal, syntagmatic relations, argument structure
- FrameNet, primary focus on horizontal, syntagmatic relations, thematic roles
- PAROLE/SIMPLE - Semantic Information for Multifunctional, PluriLingual Lexicons - based on the Generative Lexicon (Pustejovsky 1995)



Outline 4

- Ambiguity and granularity – eventual merging of senses



Semantic annotation of proper names

In some contexts, semantic annotation refers to annotation of proper names such as places, persons, organisations and events

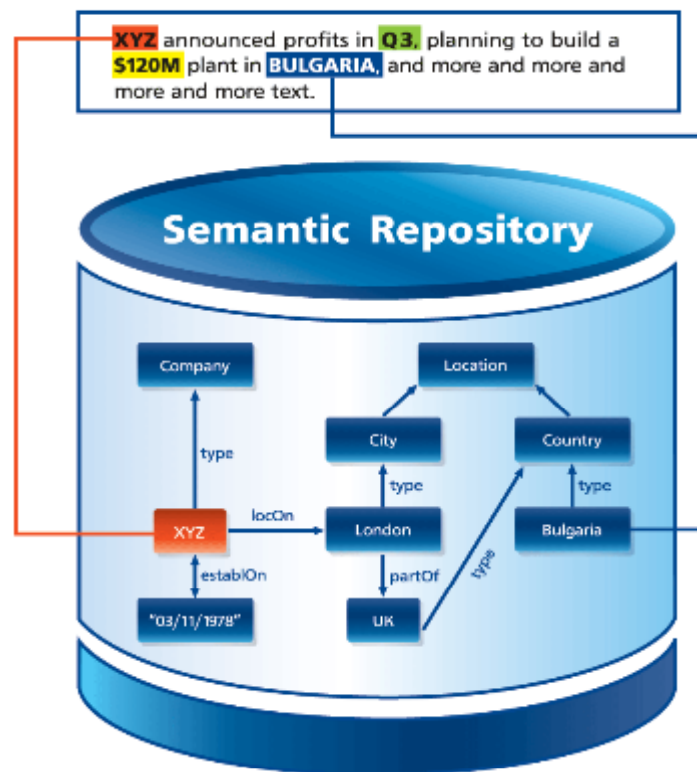
Many media demand these kinds of annotations to facilitate their users

Gazateers are available for several interests

Many media collect their own repositories by collecting lists of proper names and by dynamically extending them

Semantic annotation of proper names

Example from OntoText:





Why are lexical semantic resources relevant for semantic annotation?

- If you want to do supervised learning in lexical semantics, the lexical resource is your guiding principle
- Your annotations should ideally refer to a consistent and well motivated source
- Annotations/groupings of similar examples on the fly are possible, but it is very hard to be consistent (lexicographers' work)
- The lexical source should ideally be corpus-based



The link between the corpus and the lexicon

at dramaets forløb med at sige sig for den høje **pande**. "I Tøge Gatt-artalen", fortsætter John Formand uanfægtet, ved i en kaotisk situation foran Frem-målet at sætte **panden** til bolden langt ude fra og sende den forbi den periode. Et skud på hjelmen gav ham en flænge i **panden**. Efter et par minutters pause var han tilbage på isen der nu står frem. Han har gang på gang løbet **panden** mod en mur i arbejdet for at forbedre og koordinere at det vil give de lokale politikere tænkefuger i **panden**, men stadsingeniøren og hans tekniske medarbejdere på råk form. Den lille grå mand med "djævehornene" i **panden** var et typisk eksempel på den vittige side af Olaf fint. Varm olie og smør godt op på en stor **pande** og steg grønsagerne let. De må ikke blive bløde. Rør ca. 10 minutter. Varm fedtstoffet godt op på en stor **pande** og brun fileterne på begge sider over ret stærk varme, og læg, og skal derefter stå og småsimre på **panden** en 20 minutters tid. Kog til sidst **panden** af med småsimre på panden en 20 minutters tid. Kog til sidst **panden** af med lidt piskefløde, det giver en dejlig sauce. Kalvemellem: skål. Varm olie og smør godt op på en stor **pande**. Hæld urterne i, rør rundt og steg blandingen 3-4 minutter, få mere glæde af at uddanne sig end at løbe **panden** mod uddannelses-systemets betonkolde mur. Slaget om de brutal. Han er solbrændt med markante rynker i den høje **pande** over de stærke brune øjne, og hans hår og overskæg sider. Olien hældes bort og kødet lægges tilbage på **panden**. Der tilsættes hvidvin og når vinen er kogt ind, lidt og fra det der sker bag statsledernes bekvemrede **pander**. Overalt i verden er der f.eks. organisationer som søger at

Pande_kitchen aid

Pande_idiom

Pande_bodypart

The link between the corpus and the lexicon

at dramaets forløb med at sige sig for den nøje **pande**. "I Tøge Gatt-artalen", fortsætter John Formand uanfægtet, ved i en kaotisk situation foran Frem-målet at sætte **panden** til bolden langt ude fra og sende den forbi den periode. Et skud på hjelmen gav ham en flænge i **panden**. Efter et par minutters pause var han tilbage på isen der nu står frem. Han har gang på gang løbet **panden** mod en mur i arbejdet for at forbedre og koordinere at det vil give de lokale politikere tænkefuger i **panden**, men stadsingeniøren og hans tekniske medarbejdere på råk form. Den lille grå mand med "djævehornene" i **panden** var et typisk eksempel på den vittige side af Olaf fint. Varm olie og smør godt op på en stor **pande** og steg grønsagerne let. De må ikke blive bløde. Rør ca. 10 minutter. Varm fedtstoffet godt op på en stor **pande** og brun fileterne på begge sider over ret stærk varme. og løg, og skal derefter stå og småsimre på **panden** en 20 minutters tid. Kog til sidst **panden** af med småsimre på **panden** en 20 minutters tid. Kog til sidst **panden** af med lidt piskefløde, det giver en dejlig sauce. Kalvemellem: skål. Varm olie og smør godt op på en stor **pande**. Hæld urterne i, rør rundt og steg blandingen 3-4 minutter, få mere glæde af at uddanne sig end at løbe **panden** mod uddannelses-systemets betonkolde mur. Slaget om de brutal. Han er solbrændt med markante rynker i den høje **pande** over de stærke brune øjne, og hans hår og overskæg sider. Olien hældes bort og kødet lægges tilbage på **panden**. Der tilsættes hvidvin og når vinen er kogt ind, lidt og fra det der sker bag statsledernes bekvemrede **pander**. Overalt i verden er der f.eks. organisationer som søger at

Pande_kitchen aid

Pande_idiom

Pande_bodypart



Challenge: To contradict "old truths" regarding lexicography and Human Language Technology (HLT):

It's too difficult to reuse lexicographical data for HLT:

- Information is too inconsistent *Ide & Véronis (1995)*
- Unclear criteria for sense distinction *Killgarraf (1997)*
- Wordnet community: More economic to generate wordnets from Princeton WordNet (via translations) *Rigau & Agirre (2002)*



HLT resources vs traditional lexicography

Formal linguistics – traditionally views language as a production system:

- Chomsky's generative grammar
- Related to the development of logical systems of mathematics and logics
- Lexicon not primary interest of formal linguistics:
"the lexicon is a wastebin of irregularities"



HLT resources vs traditional lexicography

Lexical objects of interest in **formal semantics**: the elements in language that have counterparts in logic (operators, quantifiers); function words

- *Determiners*
- *Conjunctions*
- *Adverbs*
- *Numerals*



HLT resources vs traditional lexicography

In classical approaches to formal semantics, content words have typically been underspecified

One exception is argument structure

Function words are not vocabulary of particular interest in lexicography



Formal lexical semantics

Formal approaches to **content words**: nouns, verbs, adjectives in terms of theories (such as i.e. The Generative lexicon, and development projects: AQUILEX, SIMPLE etc.)

Needs for a formal apparatus that enables to express:

- Different meaning components of content words
- Word ambiguity
- Relations between concepts
- Semantic similarity

These aspects of meaning (of content words) are also dealt with in lexicography



Some HLT tools that are changing lexicography

Early advances:

- Lexicographical development tools which enabled more systematic approach to word description
- Large electronic corpora and concordance corpus tools which enable lexicographers to scan corpus material to ensure corpus-based establishment of senses



Some HLT tools that are changing lexicography

Later advances:

- Advances corpus tools like **Sketch Engine** (Kilgarrif et al.) and **DeepDict** (Bick et al.) which provide so-called "word sketches" built on parsed corpora

Sketch Engine is used in the development of:

- [Macmillan English Dictionary for Advanced Learners](#)
- [FrameNet](#)
- However, some backlog must be foreseen; huge lexicon projects are not launched every day.

eat (verb) British National Corpus freq = 13823

<u>and/or</u>	<u>674</u>	<u>0.5</u>
drink	171	9.61
sleep	53	8.61
cook	22	7.4
eat	37	6.54
smoke	7	6.4
kill	25	5.28
rest	7	5.15
dress	5	4.51
sit	16	4.36
talk	12	4.29
live	10	3.51
catch	7	3.47
wear	5	3.18
enjoy	5	2.95
try	12	2.7
work	10	2.54
come	22	2.43
go	28	2.31
play	5	1.7
find	6	0.38

<u>unary rels</u>		
np_sfin	75	9.6
np_np	290	6.1

<u>np_adj_comp</u>	<u>79</u>	<u>5.5</u>
alive	16	7.09
raw	2	6.74
rich	5	4.58
cold	5	4.12

<u>object</u>	<u>5083</u>	<u>5.4</u>
meal	220	8.85
meat	143	8.75
sandwich	93	8.6
breakfast	114	8.46
food	354	8.31
lunch	93	8.02
bread	79	7.87
dinner	92	7.81
supper	48	7.81
disorder	56	7.52
fish	116	7.51
pie	42	7.47
biscuit	39	7.46
chocolate	43	7.43
cake	56	7.39
diet	63	7.3
fruit	62	7.29
apple	39	7.05
cheese	36	6.9
sweet	22	6.84
egg	50	6.82
steak	21	6.81
toast	22	6.8
vegetable	31	6.79
grass	37	6.71

<u>part_trans</u>	<u>200</u>	<u>3.8</u>
up	135	4.75
out	47	3.44
off	8	2.56

<u>pp_at-p</u>	<u>166</u>	<u>2.8</u>
restaurant	11	5.34
table	18	4.41
home	7	2.01
time	19	1.51

<u>subject</u>	<u>1373</u>	<u>2.7</u>
swallow	12	7.8
shark	8	6.89
caterpillar	6	6.69
snail	5	6.24
predator	7	6.11
snake	6	6.01
owl	8	5.92
fish	28	5.74
dog	31	5.72
goat	5	5.7
human	8	5.62
pop	6	5.47
rat	6	5.37
Ruth	5	5.13
cattle	6	5.11
bird	16	5.09
Adam	5	5.08
cat	7	4.68
fruit	6	4.41
animal	15	4.38
passenger	5	4.32
kid	5	4.23
Thomas	6	4.08
child	43	3.87
sister	7	3.86



Corpus: **British National Corpus**

Hits: 5

H8R

that's simply Corruption of the facts. Adam **ate** the apple. Eve ate Adam. The serpent ate

CB8

Ja-ack? If I do the braising steak, will Adam **eat** a baked potato or should I do roast? </p>

CDB

. He couldn't eat another mouthful. Adam **ate** . He ate his way through bread and butter

CDB

while you were in Greece? </p><p> Adam could **eat** nothing. The other time, he remembered,

F99

soft sand above the tide line while Adam **ate** ; there had always been fewer stones here

Lexical Computi

spise (verb)

total of 71129 relations

Hide Frequencies

Subjects:

PERS: den, vi, man, jeg, de, du, han, hun, der, I

7.77:8 <H> ·

7.19:8 **PROP-hum** ·

6.22:7 <HH> ·

6.09:7 <Hprof> ·

6.07:7 <Hnat> ·

3.97:6 <Hfam> ·

5.08:3 **PROP** ·

0.96:7 **barn** ·

1.79:6 **dansker** ·

0.44:6 **folk** ·

1.34:5 **hund** ·

1.86:4 <Hideo> ·

1.39:4 **PROP-tit** ·

1.29:4 **PROP-org** ·

1.25:2 **nordbo** ·

0.65:2 **vegetar** ·

0.4:1 **vandnymfe** ·

0.4:1 **eremitskreb** ·

0.11:1 **knopsvane**

Accusative objects:

PERS: den

10.82:9 <food-h> · 9.26:9 <food-m> ·

9.15:9 <food> · 8.9:9 <food-m-h> ·

8.4:9 <food-c-h> · 7.48:8 <cc-h> ·

7.46:8 <Aich> · 7.16:8 <occ> ·

7.1:8 <fruit-c> · 6.69:8 <H> ·

6.67:8 **frokost** · 6.52:7 <amount> ·

6.46:7 <temp> · 6.38:7 <Azo> ·

6.33:7 **morgenmad** · 6.28:7 <food-c> ·

5.03:8 **mad** · 5.91:7 ·

5.53:7 <Adom> · 5.26:7 <cm> ·

5.15:7 **middag** · 5.23:6 **aftensmad** ·

5.08:6 **slik-2** · 5.08:6 <ac> · 3.75:7 **fisk**

· 4.61:6 <Aent> · 4.61:6 <cm-liq-h> ·

4.46:6 <cc-org> · 4.32:6 <Aorn> ·

4.24:6 <HH> · 4.19:6 <cm-h> ·

4.17:6 <A> · 6.1:4 **PROP** ·

3.94:6 **PROP-hum** · 3.66:5 <drink-h> ·

3.52:5 <unit> · 3.52:5 <cc> ·

3.45:5 <drink-c-h> · 3.41:5 <Hprof> ·

3.29:5 <anorg> · 3.29:5 <cm-liq> ·

3.25:5 <fruit> · 2.98:5 <sem-r> ·

2.93:5 <am> · 2.88:5 <anmov> ·

2.5:5 <tool> · 2.44:5 <ac-cat> ·

2.44:5 <sem-w> · 2.37:4 <Lh> ·

2.1:4 <act-c> · 1.69:4 <an> ·

1.59:4 <meta> · 1.59:4 <domain> ·

1.5:4 <Ltop> · 1.5:4 <sick> ·

1.29:4 <con> · 1.29:4 <Hfam> ·

1.29:4 <ling> · 0.11:3 **PROP-org**

Forms	Abs Freq	Rel Freq
Total	461	100.00%
børn -> spiser	35	7.59%
børn -> spise	19	4.12%
mennesker -> spiser	19	4.12%
børnene -> spiser	18	3.90%
kvinder -> spiser	13	2.82%
børnene -> spise	13	2.82%

[Show all forms...](#)

Concordances for: <H>_N -> spise_V

ID	Text
inf140-53377	Den nye arbejder spiser Grandiosa , rejser på campingferie , køber billig sprut ved Svinesund og stemmer Frp.
c2000-paynjnqw111	Kostundersøgelsen fra 1995 viser , at børn og voksne spiser næsten ens .
c2000-paynjnqw111	Den nye arbejder spiser Grandiosa , rejser på campingferie , køber billig sprut ved Svinesund og stemmer Frp.

mor -> spiste	5	7.25%
mor -> spiser	4	5.80%
far -> spiser	4	5.80%
forfædre -> spiste	3	4.35%
mor -> spise	3	4.35%
mor -> spist	2	2.90%

[Show all forms...](#)

Concordances for: <Hfam>_N -> spise_V

ID	Text
c90-PEGPRIMG	Det bord er reserveret til os selv , og der stilles rent service på det , når alle gæsterne er nået til desserten eller kaffen , og så spiser vi , far og mor og min søster Sanna og jeg , og somme tider Bettino .
inf130-64643	For så ville datteren slet ikke spise noget . Men den afmagrede teenager var gået med til at gå i terapi inde i København .
inf130-120191	Min mor spiste slet ikke noget i en uge . " I det nye brev stod , at :
inf00-161005	Når drengene har enist deres proteinbøffer dyrker de kondiløb med sikkerhedsseler nå



Some HLT tools that are changing lexicography

These tools will have great impact on future lexicography

- Lexicographers will be more faithful to corpus data, i.e. less idiosyncratic
- More systematic
- Better, prototypical examples
- Better and more current collocational information
- Clearer sense distinctions?
- Clearer definitions?

Will make traditional lexicons more suitable for HLT



Wordnets

- Wordnets originate from ideas about the structure of the mental lexicon
- Wordnets have become popular in HLT because they can work as word models for computers
- Wordnets are being developed for more than 40 languages (i.e. EuroWordNet)
- Easy access: Princeton and some others are open source



Hypotheses behind wordnet

Separability hypothesis:

The mental lexicon has an independent status compared to our other language competences

Meaningful to study vocabulary independently



Hypotheses behind wordnet

We wouldn't be capable of mastering so many words without use of patterns

Pattern hypothesis: Psycholinguists had a hypothesis that our knowledge of word meaning is stored in terms of a network with internal relations between concepts; some concepts are superconcepts to others



Hypotheses behind wordnet

Storing and inheritance hypothesis:

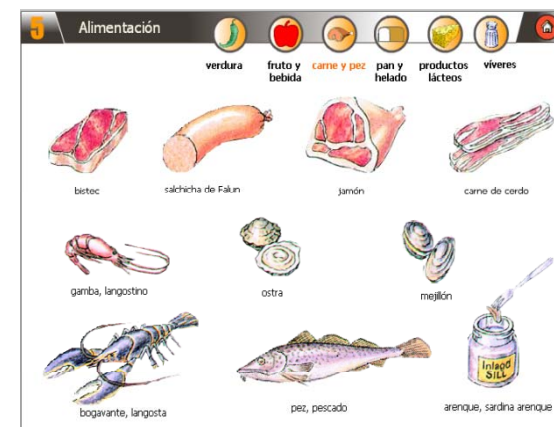
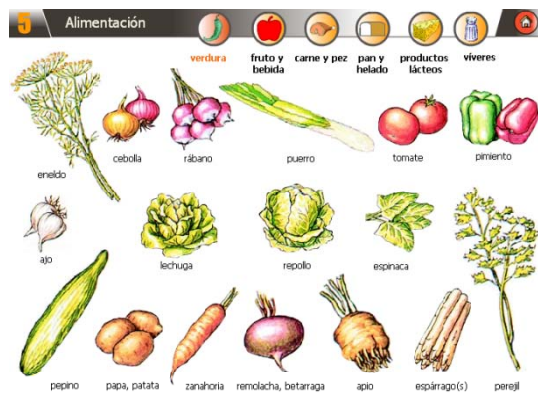
Knowledge is stored at the highest possible node and inherited by lower (more specific) concepts rather than being multiply stored

Can birds fly? Short reaction time

Do canaries fly? Longer reaction time

Do canaries have a heart? Long reaction time

Same hypotheses used in language learning (Lexin project)





The architecture of wordnets

A wordnet is a collection of word senses where senses denoting the same concepts are organised in so-called **synsets (= synonym sets)**

Example of synset:

{computer, datamat, datamaskine}



The architecture of wordnets

Synsets are connected to each other by means of the semantic relations that hold between them, such as

- Hyponymy and hypernymy (e.g. cup — tea cup) and
- Meronymy - parts and wholes (e.g. handle — tea cup)



Princeton WordNet: instruments

WordNet Search - 3.0 - [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for: Display Options:

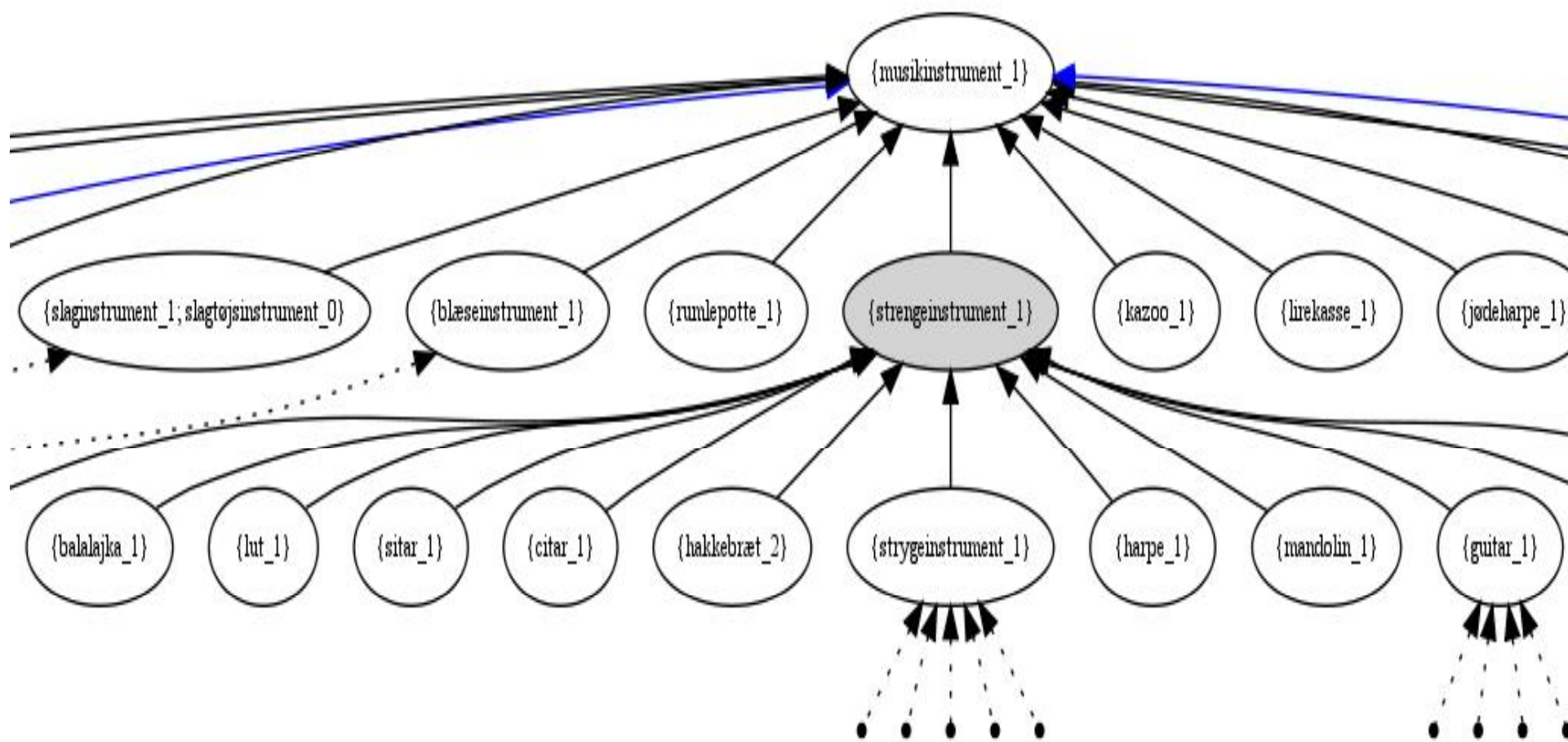
Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Noun

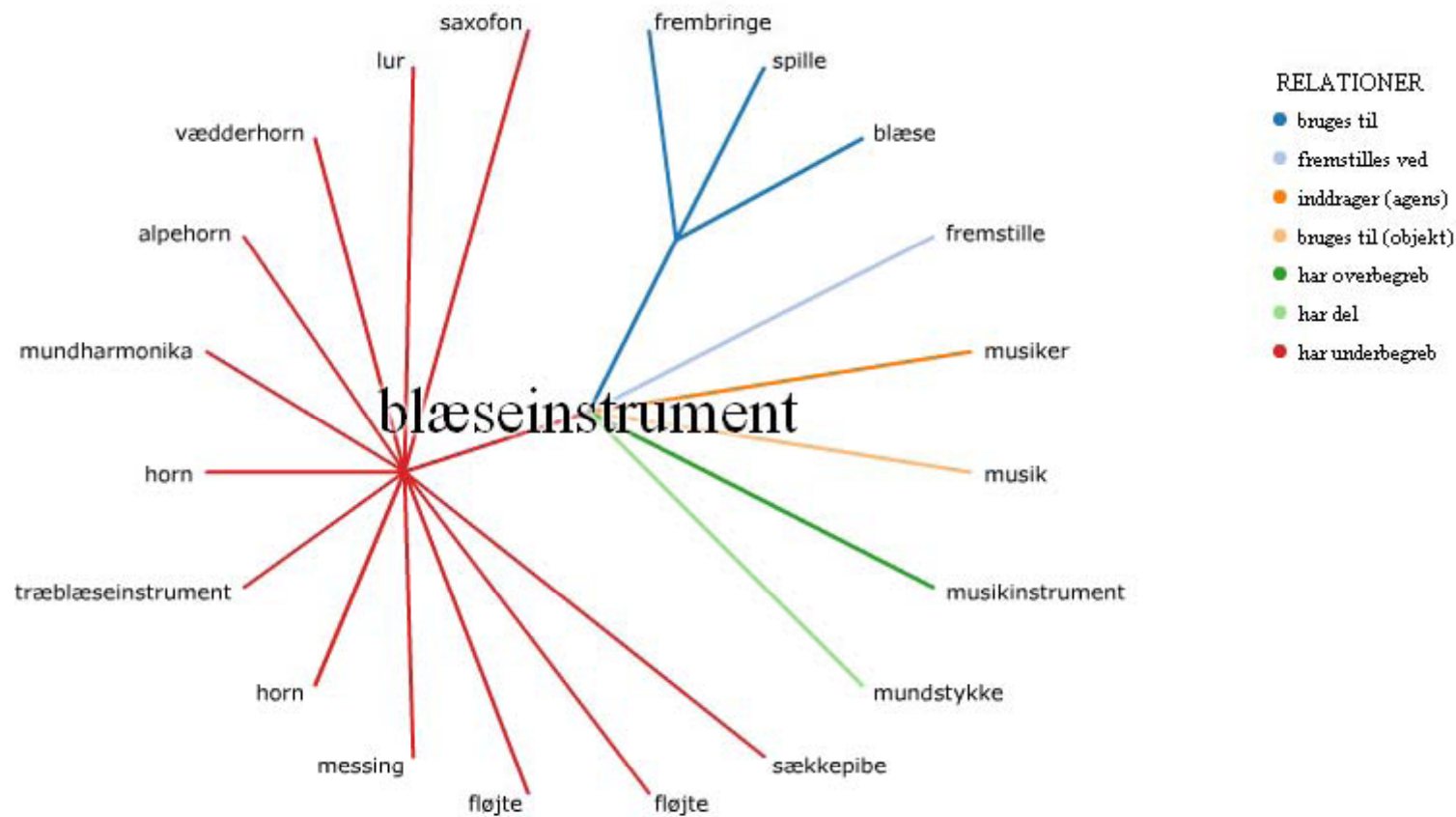
- **S:** (n) **wind instrument**, [wind](#) (a musical instrument in which the sound is produced by an enclosed column of air that is moved by the
 - **direct hyponym** / **full hyponym**
 - **S:** (n) [brass](#), [brass instrument](#) (a wind instrument that consists of a brass tube (usually of variable length) that is blown by
 - **S:** (n) [free-reed instrument](#) (a wind instrument with a free reed)
 - **S:** (n) [kazoo](#) (a toy wind instrument that has a membrane that makes a sound when you hum into the mouthpiece)
 - **S:** (n) [ocarina](#), [sweet potato](#) (egg-shaped terra cotta wind instrument with a mouthpiece and finger holes)
 - **S:** (n) [organ](#), [pipe organ](#) (wind instrument whose sound is produced by means of pipes arranged in sets supplied with air
 - **S:** (n) [organ pipe](#), [pipe](#), [pipework](#) (the flues and stops on a pipe organ)
 - **S:** (n) [pipe](#) (a tubular wind instrument)
 - **S:** (n) [post horn](#) (wind instrument used by postilions of the 18th and 19th centuries)
 - **S:** (n) [whistle](#) (a small wind instrument that produces a whistling sound by blowing into it)
 - **S:** (n) [woodwind](#), [woodwind instrument](#), [wood](#) (any wind instrument other than the brass instruments)
 - **part meronym**
 - **S:** (n) [bell](#) (the flared opening of a tubular device)
 - **S:** (n) [mouthpiece](#), [embouchure](#) (the aperture of a wind instrument into which the player blows directly)
 - **direct hypernym** / **inherited hypernym** / **sister term**
 - **S:** (n) [musical instrument](#), [instrument](#) (any of various devices or contrivances that can be used to produce musical tones)

[WordNet home page](#)

DanNet wordnet: instruments



Other relations: wind instruments





How do wordnets differ from ontologies?

- Wordnets differ from formal ontologies by being anchored in language expressions
- Wordnets are meant to express lexical-semantic knowledge in a systematic way
- Wordnets refer to concrete languages like *Danish* or *Spanish* and can therefore account for language specific structures

Ex:

Dedos – *dedos (de manos)* (fingers), *dedos de pies* (toes)
Lemmer (*limbs*) – *arme, fødder, tæer, fingre* (arms, feet, toes, fingers)

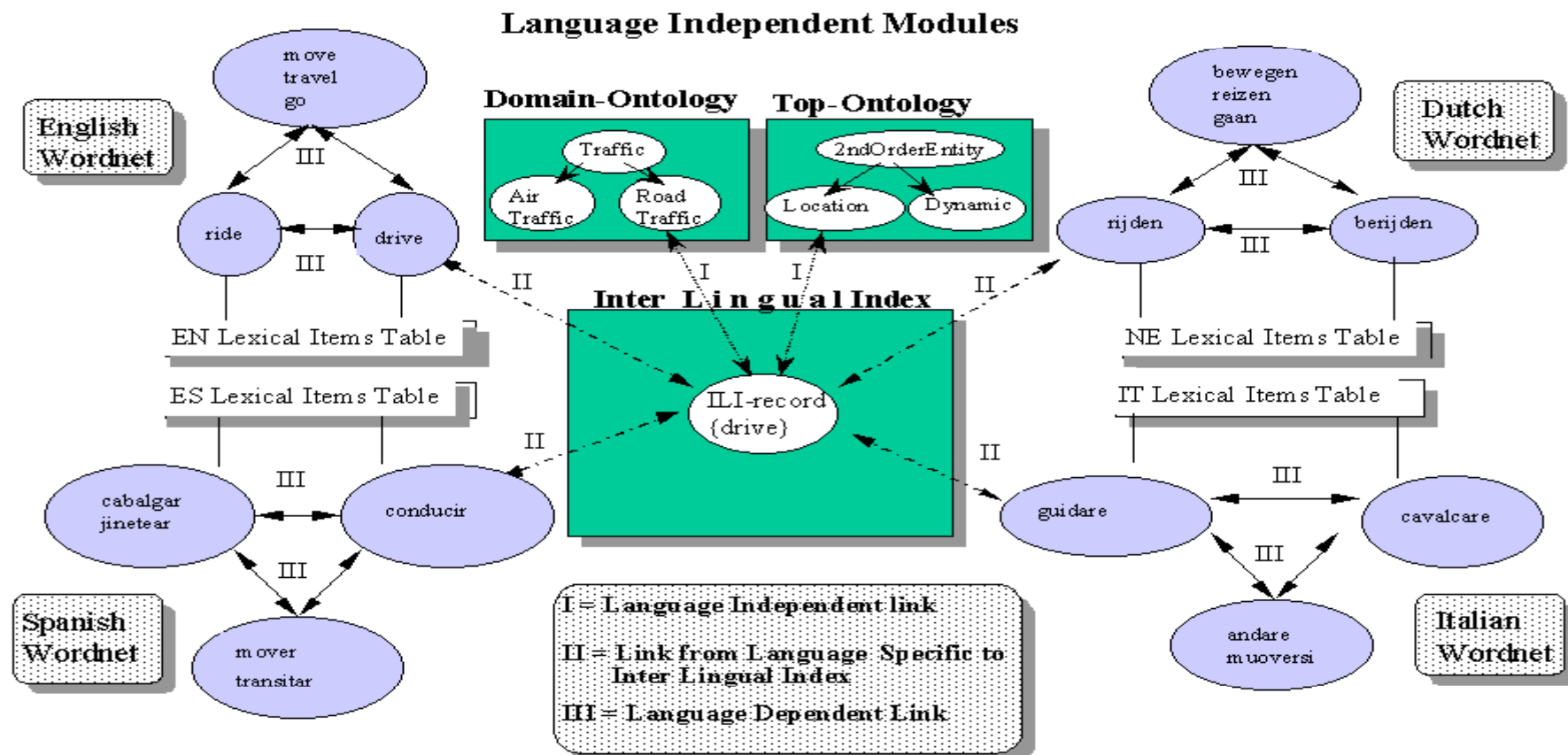


How do wordnets differ from ontologies?

- In wordnets we talk about senses and synonym sets, in ontology we talk about classes, i.e. an abstraction over word level
- Classical wordnets do not include **axiomatic expressions in terms of logical statements**:
 - *There exist x, y : $instance(x, Person)$ & $instance(y, Animal)$ & $pet(x, y)$*
- Several attempts to connect wordnets to formal ontologies (SUMO, DOLCE)

To which extent are wordnets multilingual?

Architecture of the EuroWordNet Data Structure





Limitations of wordnets (1)

Not (very) corpus-based; not necessarily built on monolingual grounds

Some wordnets are much too fine-grained, not always textually motivated

Taxonomy may be a central semantic dimension of nouns, but not for other word classes



Taxonomy and verbs

X is a way of Y

Events seem to be better organized along the dimensions of the manner relation, i.e. *troponymy* (Fellbaum 1998, 2002)

Examples:

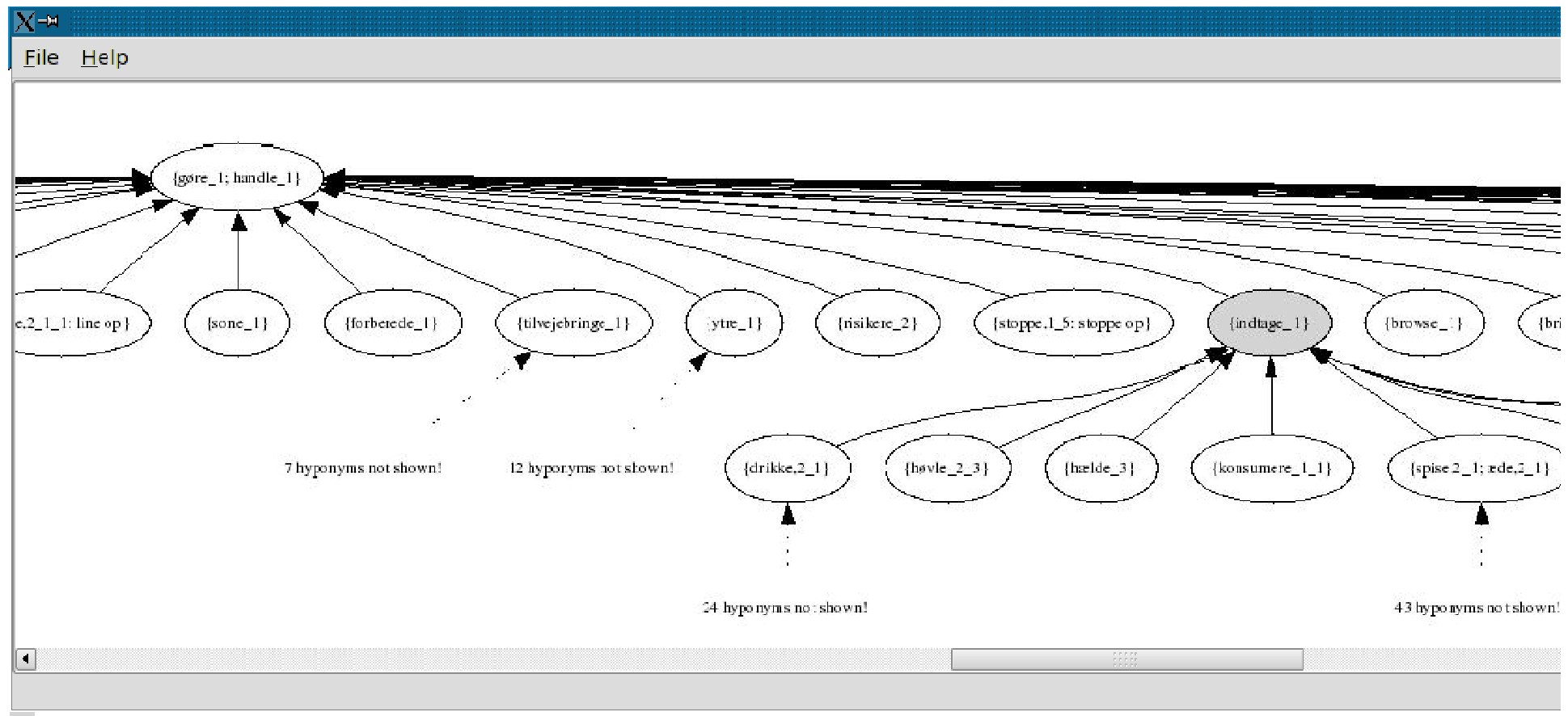
to scoff is a quick and rough way of eating

to swim is a way of moving in water

to reply is way of communicating

Taxonomy and verbs

(scoff is way of eating which is a way of consuming which is a way of acting)





The manner relation is highly polysemous

The differentiae distinguishing the superordinate from the more specific subordinate may be i.e.

direction *move* - *rise*

speed *walk* - *run*

volume *talk* - *scream*

intensity *persuade* - *brainwash*

..

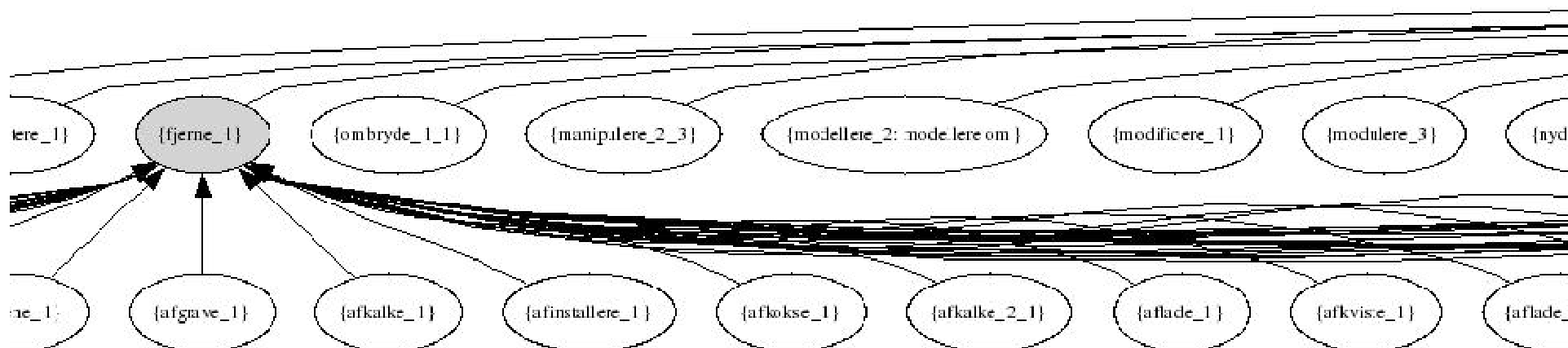
Fellbaum 2002:p. 27

Pedersen and Nimb 2008



Taxonomy and verbs

remove -> decalcify, descale, delouse, dehydrate, discolour)





Taxonomy and verbs

*remove -> decalcify, descale, delouse, dehydrate,
discolour*

Not so much ways of removing but rather different
material that is removed (calcium, lice, humidity,
colour..)



Limitations of wordnets (2)

- Wordnets lack syntagmatic relations, do not capture some of the basic semantic dimensions of verbs and deverbal nouns (argument structure, selectional restrictions)
- Wordnets generally lack compositional components (for example those included by telicity)



VerbNet

Inspired by Levin's semantic classes

Largest online verb lexicon for English

Approx. 3960 verbs distributed over 471 classes

- Semantic roles
- Syntactic frames
- Selectional restrictions
- Semantic predicates (Jackendoff inspired)

<http://verbs.colorado.edu/~mpalmer/projects/verbnet.html>

Palmer, Gildea & Xue "Semantic Role Labelling"

Kipper, Dang & Palmer "Building a class-based verb lexicon using TAGs"



Class Hit-18.1			
Roles and Restrictions: Agent[+int_control] Patient[+concrete] Instrument[+concrete]			
Members: bang, bash, hit, kick, ...			
Frames:			
Name	Example	Syntax	Semantics
Basic Transitive	Paula hit the ball	Agent V Patient	cause(Agent, E)manner(during(E), directedmotion, Agent) !contact(during(E), Agent, Patient) manner(end(E),forceful, Agent) contact(end(E), Agent, Patient)

pour-9.5

Members: 8, Frames: 5

CLASS HIERARCHY
POUR - 9.5
NO SUBCLASSES

MEMBERS

REF KEY

DRIBBLE (FN 1; WN 1, 2)	SPEW (FN 1; WN 1, 2, 3)
DRIP (FN 1; WN 1, 2)	SPILL (FN 1; WN 1, 2, 3)
POUR (FN 1; WN 1, 3, 4)	TRICKLE (WN 1)
SLOP (WN 1)	
SLOSH (WN 3)	

ROLES

REF

- AGENT [+ANIMATE]
- THEME [+SUBSTANCE] [+CONCRETE & +PURAL]
- LOCATION [+LOCATION & -REGION]
- SOURCE [+LOCATION & -REGION]

FRAMES

REF KEY

NP V NP PP.LOCATION

EXAMPLE "Tamara poured water into the bowl."

SYNTAX AGENT **V** THEME [{ +PATH & -DEST_DIR }] LOCATION

SEMANTICS MOTION (DURING(E), THEME) NOT (PREP(START(E), THEME, LOCATION)) PREP(E, THEME, LOCATION)
CAUSE(AGENT, E)

NP V NP ADVP

EXAMPLE "Tamara poured water here."

SYNTAX AGENT **V** THEME LOCATION <+ADV_LOC>

SEMANTICS MOTION (DURING(E), THEME) NOT (PREP(START(E), THEME, LOCATION)) PREP(E, THEME, LOCATION)

NP V PP.LOCATION

EXAMPLE "Water poured onto the plants."

SYNTAX THEME **V** [{ +PATH & -DEST_DIR }] LOCATION

SEMANTICS MOTION (DURING(E), THEME) NOT (PREP(START(E), THEME, LOCATION)) PREP(E, THEME, LOCATION)

NP V NP PP.SOURCE PP.LOCATION

EXAMPLE "Maria poured water from the bowl into the cup."

SYNTAX AGENT **V** THEME [{ +SRC }] SOURCE [{ +DEST_CONF }] LOCATION

SEMANTICS NOT (PREP(START(E), THEME, LOCATION)) PREP(E, THEME, SOURCE) PREP(E, THEME, LOCATION) CAUSE(AGENT, E)

NP V PP.SOURCE PP.LOCATION

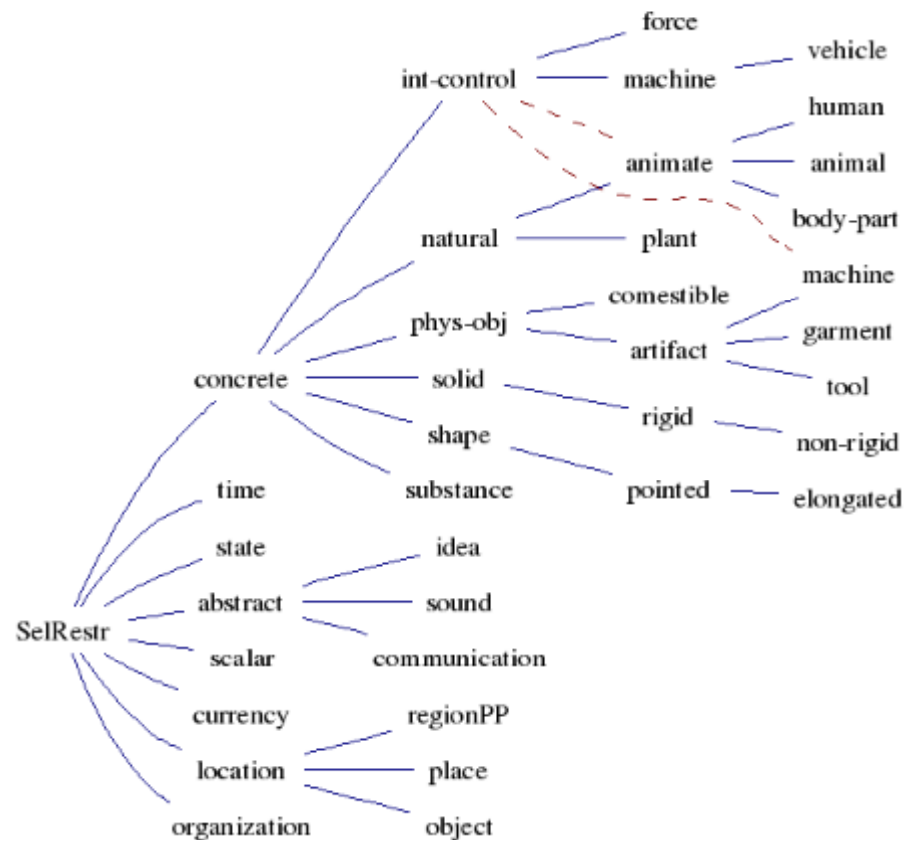
EXAMPLE "Water poured from the bowl into the cup."

SYNTAX THEME **V** [{ +SRC }] SOURCE [{ +DEST_CONF }] LOCATION

SEMANTICS NOT (PREP(START(E), THEME, LOCATION)) PREP(E, THEME, SOURCE) PREP(E, THEME, LOCATION)

Actor:	used for some communication classes (e.g., Chitchat-37.6, Marry-36.2, Meet-36.2) when both arguments can be considered symmetrical (pseudo-agents).
Agent:	generally a human or an animate subject. Used mostly as a volitional agent, but also used in VerbNet for internally controlled subjects such as forces and machines.
Asset:	used for the Sum of Money Alternation, present in classes such as Build-26.1, Get-13.5.1, and Obtain-13.5.2 with 'currency' as a selectional restriction.
Attribute:	attribute of Patient/Theme refers to a quality of something that is being changed, as in (The price)att of oil soared. At the moment, we have only one class using this role Calibratable cos-45.6 to capture the Possessor Subject Possessor-Attribute Factoring Alternation. The selectional restriction 'scalar' (defined as a quantity, such as mass, length, time, or temperature, which is completely specified by a number on an appropriate scale) ensures the nature of Attribute.
Beneficiary:	the entity that benefits from some action. Used by such classes as Build-26.1, Get-13.5.1, Performance-26.7, Preparing-26.3, and Steal-10.5. Generally introduced by the preposition 'for', or double object variant in the benefactive alternation.
Cause:	used mostly by classes involving Psychological Verbs and Verbs Involving the Body.
Location, Destination, Source:	used for spatial locations.
Destination:	end point of the motion, or direction towards which the motion is directed. Used with a 'to' prepositional phrase by classes of change of location, such as Banish-10.2, and Verbs of Sending and Carrying. Also used as location direct objects in classes where the concept of destination is implicit (and location could not be Source), such as Butter-9.9, or Image impression-25.1.
Source:	start point of the motion. Usually introduced by a source prepositional phrase (mostly headed by 'from' or 'out of'). It is also used as a direct object in such classes as Clear-10.3, Leave-51.2, and Wipe instr-10.4.2.
Location:	underspecified destination, source, or place, in general introduced by a locative or path prepositional phrase.
Experiencer:	used for a participant that is aware or experiencing something. In VerbNet it is used by classes involving Psychological Verbs, Verbs of Perception, Touch, and Verbs Involving the Body.
Extent:	used only in the Calibratable-45.6 class, to specify the range or degree of change, as in The price of oil soared (10%)ext. This role may be added to other classes.
Instrument:	used for objects (or forces) that come in contact with an object and cause some change in them. Generally introduced by a 'with' prepositional phrase. Also used as a subject in the Instrument Subject Alternation and as a direct object in the Poke-19 class for the Through/With Alternation and in the Hit-18.1 class for the With/Against Alternation.
Material and Product:	used in the Build and Grow classes to capture the key semantic components of the arguments. Used by classes from Verbs of Creation and Transformation that allow for the Material/Product Alternation.
Material:	start point of transformation.
Product:	end result of transformation.
Patient:	used for participants that are undergoing a process or that have been affected in some way. Verbs that explicitly (or implicitly) express changes of state have Patient as their usual direct object. We also use Patient1 and Patient2 for some classes of Verbs of Combining and Attaching and Verbs of Separating and Disassembling, where there are two roles that undergo some change with no clear distinction between them.
Predicate:	used for classes with a predicative complement.
Recipient:	target of the transfer. Used by some classes of Verbs of Change of Possession, Verbs of Communication, and Verbs Involving the Body. The selection restrictions on this role always allow for animate and sometimes for organization recipients.
Stimulus:	used by Verbs of Perception for events or objects that elicit some response from an xperiencer. This role usually imposes no restrictions.
Theme:	used for participants in a location or undergoing a change of location. Also, Theme1 and Theme2 are used for a few classes where there seems to be no distinction between the arguments, such as Differ-23.4 and Exchange-13.6 classes.
Time:	class-specific role, used in Begin-55.1 class to express time.
Topic:	topic of communication verbs to handle theme/topic of the conversation or transfer of message. In some cases, like the verbs in the Say-37.7 class, it would seem better to have 'Message' instead of 'Topic', but we decided not to proliferate the number of roles.

VerbNet – selectional restrictions



FrameNet



- Approx 950 semantic frames (fx sell and buy-frame, apply heat-frame etc.)
- Approx 2.500 frame elements (each frame has its own set of frame elements)
- 7000 lexical units
- 130.000 annotated sentences annotated Frame Element (FE), phrase type (PT) and grammatical function (GF).
- Core
- Peripheral
- Extra-thematic (but not free adjuncts)

Different differentiation than what is used in Levin's verb taxonomy, which is built on syntactic patterns.

<http://framenet.icsi.berkeley.edu/>

Fillmore et al. "FrameNet II"



FrameNet: A three-layer model

Motion Frame

Text	<i>The kitchen</i>	<i>swarmed</i>	<i>with ants</i>	<i>during summer</i>
FE	Location		Mover	Time
PT	NP		PP	PP
GF	Subject		Oblique	Adverbial

- Null instantiation: conceptually present in the phrase but not expressed
- Hierarchy of frames which share certain frame elements



FrameNet: Annotations for *swarm*

swarm.v

Frame Element	Core Type
Degree	Peripheral
Depictive	Extra-Thematic
Location	Core
Theme	Core
Time	Peripheral

[Turn Colors Off](#)

- 730-s20-ppwith
 1. The seas today SWARM with crustacean arthropods ranging in size from t
 2. The Great Barrier Reef SWARMS with life .
 3. The narrow streets SWARMED with young people .
 4. Kaffa and its ships SWARMED with rats ; plague was quickly disseminate
 5. Chapter 21 does not deal merely with a desert SWARMING with snakes ,
- 780-s20-ppother
 1. As they entered the boardroom , executives SWARMED around Damian .
 2. They SWARMED like lice over the camp , and there was a battle being w
 3. Raimundo 's shaggy lurchers SWARMED round Perdita as she staggered g
 4. A group of Dutch tourists SWARMED out of a restaurant behind Rain and
- 810-s20-pother
 1. Above them , Hunters SWARMED in increasing numbers .DNI
- 880-s20-intrans-simple
 1. Like ants over their nest they SWARMED , black and confused .INI
- 890-s20-intrans-adverb
 1. Men SWARM everywhere , the cafés are crowded with them , sitting at d
 2. As the boat docked , police SWARMED on board .
 3. Enemy aircraft SWARMED overhead , methodically searching each gully :
- 900-s20-other
 1. Frogs love it , and in springtime birds SWARM all over it , making off with
 2. As the little rodents SWARMED all over him , he swung his arm back , po
 3. Thousands of civilians SWARMED over several military garrisons , seizing

Comparing verb descriptions

Palmer et al. P. 27



Table 2.1: A comparison of current lexical resources: FrameNet, VerbNet and PropBank

Attribute	FrameNet	VerbNet	PropBank
Lexical Units (senses)	11,600	5,733	6,204
Lexemes (lemmas)	6,000	3,965 (verbs)	5,213 (verbs)
categories	960 Frames	471 Classes	
Semantic Roles	2,500+ Frame Elements	24 Thematic Role types	16 Args, 6000+ (verb specific roles)
Annotated Data	150,000 sentences covering 6,800 lexical units	1M word WSJ Treebank (90% token coverage)	1.75M word WSJ/BN/BC, etc. (all verbs)



Exercise to WordNet

- Sit in groups of two to three
- Check out in Princeton WordNet (wordnet.princeton.edu) the concepts
snack food, fast food, hot dog and *burger*. Draw on a piece of paper the partial taxonomy/taxonomies (hypernymy, hyponymy)
- Discuss whether you could suggest an improved taxonomy for these concepts?
- Do you consider any of the involved concepts as not fitting into the taxonomy?

(For comparison, consider *burger* in DanNet (andreord.dk))



Exercise to WordNet and VerbNet

- Sit in groups of two to three
- Check out in Princeton WordNet (<http://wordnet.princeton.edu>) the verb *bake*
- Unfold all relations (including sentence frame) and discuss
- Compare to *bake* in VerbNet (<http://verbs.colorado.edu/verb-index/index.php>)
- Discuss differences



Exercise to VerbNet and FrameNet

- Sit in groups of two to three
- Check out *cook* in VerbNet (<http://verbs.colorado.edu/verb-index/index.php>)
- Compare to *cook* in FrameNet (<http://framenet.icsi.berkeley.edu/>)
(Click View FrameNet data; click Lexical units)



PAROLE/SIMPLE

- Aim of EU project: provide harmonised semantic lexicons for 12 EU languages
- 10.000 semantic entries for each language
- A unified , ontology-based semantic model: the SIMPLE model (Lenci *et al.* 2000)
- Web page with 100 samples for each language:
<http://www.ub.es/gilcub/SIMPLE/simple.html>





Theoretical background of SIMPLE

- Argument structure
- Event structure
- Qualia structure
- Lexical inheritance structure (linguistic ontology)



Pustejovsky 1995



Qualia structure

- Formal (is_a - hyponymy)
- Constitutive (form – meronymy)
- Telic (purpose – used_for)
- Agentive (origin – made_by)



Qualia structure for *cake*

- Formal (is_a – *pastry* → *food* → *artifact* → *physobj*)
- Constitutive (form – made of *flour, sugar..*)
- Telic (purpose – used_for *eat*)
- Agentive (origin – made_by *bake*)

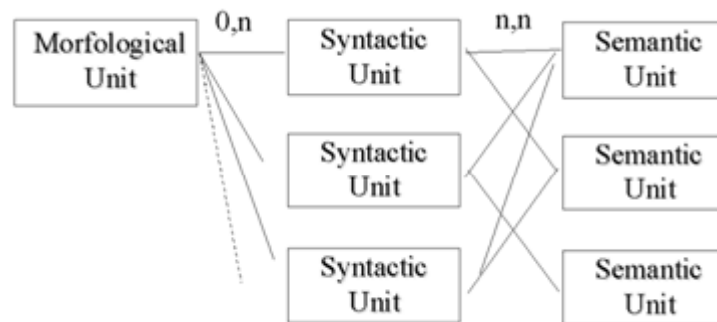


Integrating the levels

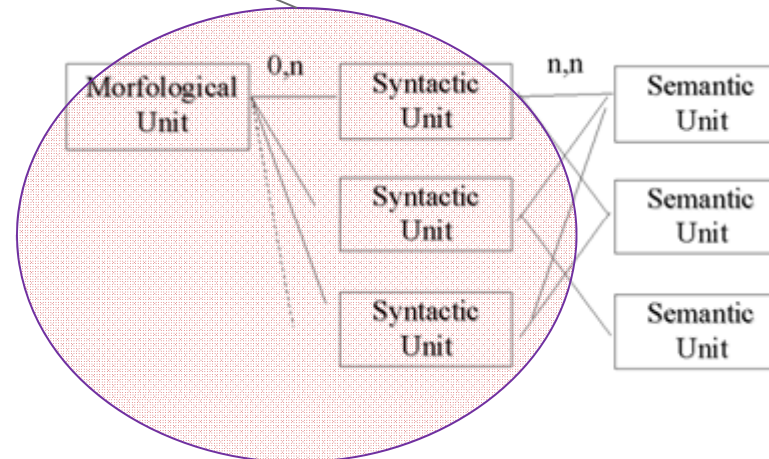
$$\left[\begin{array}{l} \text{bake a cake} \\ \text{EVENTSTR} = \left[\begin{array}{l} E_1 = e_1:\text{process} \\ E_2 = e_2:\text{state} \\ \text{RESTR} = <_{\infty} \\ \text{HEAD} = e_1 \end{array} \right] \\ \text{ARGSTR} = \left[\begin{array}{l} \text{ARG1} = \boxed{1} \left[\begin{array}{l} \text{animate_ind} \\ \text{FORMAL} = \text{physobj} \end{array} \right] \\ \text{ARG2} = \boxed{2} \left[\begin{array}{l} \text{artifact} \\ \text{CONST} = \boxed{3} \\ \text{FORMAL} = \text{physobj} \end{array} \right] \\ \text{D-ARG1} = \boxed{3} \left[\begin{array}{l} \text{material} \\ \text{FORMAL} = \text{mass} \end{array} \right] \end{array} \right] \\ \text{QUALIA} = \left[\begin{array}{l} \text{create_lcp} \\ \text{FORMAL} = \text{exist}(e_2, \boxed{2}) \\ \text{AGENTIVE} = \text{bake_act}(e_1, \boxed{1}, \boxed{3}) \end{array} \right] \end{array} \right]$$



PAROLE/SIMPLE

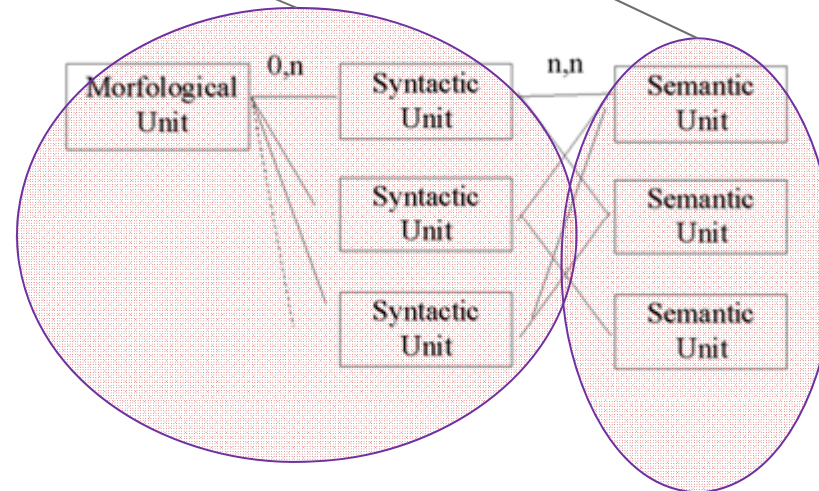


PAROLE/SIMPLE



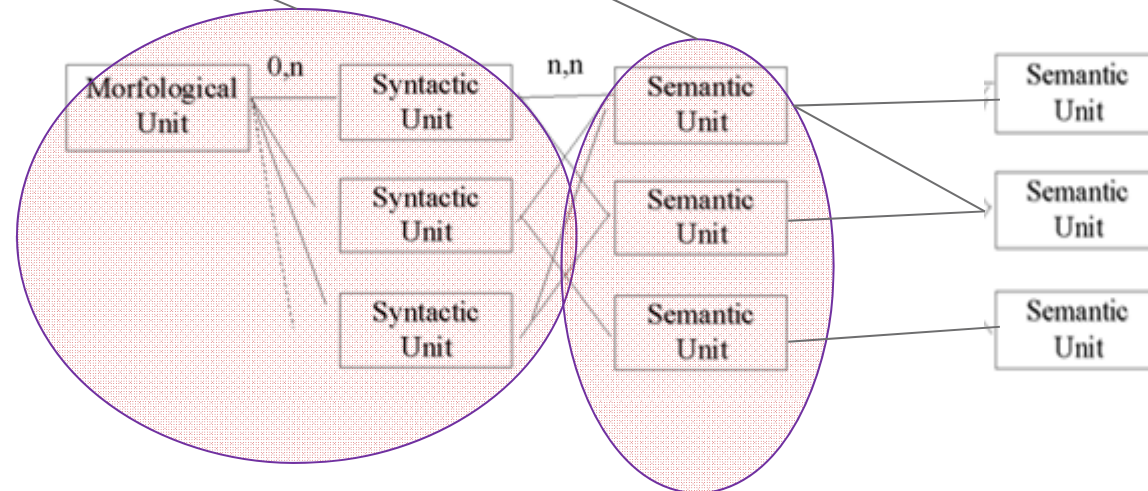


PAROLE/SIMPLE

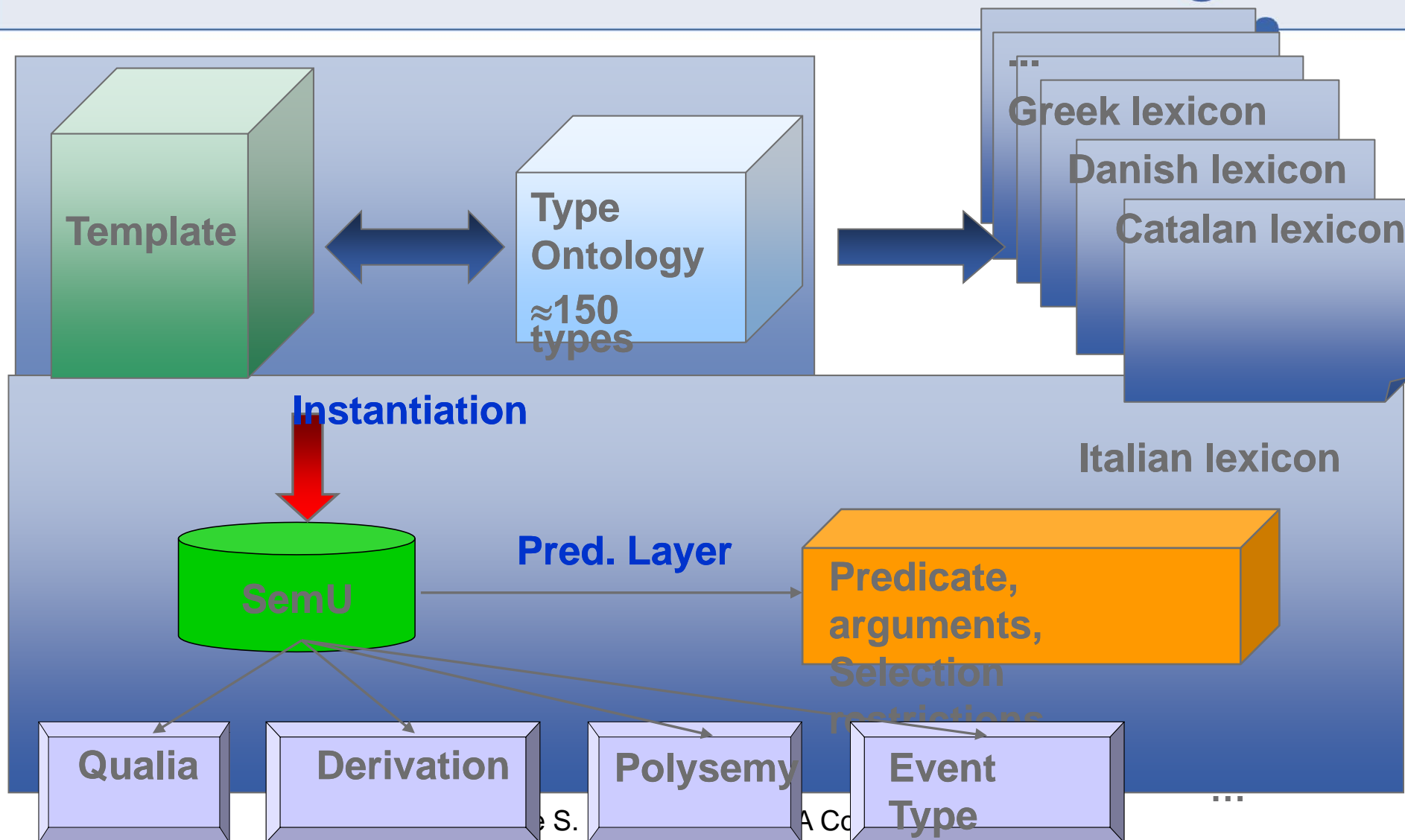




PAROLE/SIMPLE - MULTILINGUAL



Overall Organization (Calzolari)





PAROLE/SIMPLE Verbs

Semantic Unit	<i>krydse_CHL</i> (cross)
Definition:	<i>Bevæge sig tværs over et åbent område</i> (Nudansk Ordbog) (move across an open area)
Corpus example:	<i>Drengen krydsede sporet ved stationen, men så ikke toget</i> 'The boy crossed the rails at the station but he didn't see the train'
Semantic type:	Change of location
Unification Path:	Change/Agentive
Domain:	General
Argument Structure	ARG1 ARGDirection
Selectional Restrictions	ARG1= Human OR Animal OR Vehicle Direction = Concrete
EventType	Transition
Formal quale:	<i>Is_a = ændring</i> (change)
Agentive quale:	Agentive = <i>bevæge_sig</i> (move)
Telic quale:	Nil
Constitutive quale:	Resulting_State = <i>være</i> (be) Direction= forwards
Systematic Polysemy	Nil
Synonymy	Nil



PAROLE/SIMPLE Nouns

Semantic Unit	<i>puslespil</i> (jigsaw puzzle)
Definition:	<i>et spil med træ- el. papbrikker i forskellige faconer som skal lægges sammen så de danner et hele</i> (NDO)
Corpus example:	<i>nu var hun næsten ved at være færdig med det puslespil, hun var begyndt på lige efter påske</i>
Ontological type:	Artifact
Unification Path	Concrete_Entity Agentive Telic
Domain:	General
Formal quale:	<i>is_a = spil</i> (game)
Agentive quale:	<i>created_by = udskære</i> (cutting out)
Telic quale:	<i>used_for = samle til et hele</i> (collect to a whole)
Constitutive quale:	<i>has_as_parts = træbrikker OR papbrikker</i>



Excercises SIMPLE

Build the qualia structure for *orange* and *house*

Formal (is_a - hyponymy)

Constitutive (form – meronymy, has_as_parts, is_a_part_of)

Telic (purpose – used_for/purpose_of)

Agentive (origin – made_by)

Check the concepts in Princeton Wordnet
and see which relations correlate

<http://wordnet.princeton.edu/>

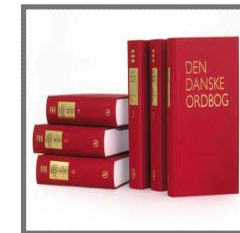


DanNet – a WordNet that includes qualia

**The Danish Society for Language and
Literature
University of Copenhagen**



Background resources:
The Danish Dictionary
SIMPLE-DK



International inspiration:
Princeton WordNet, EuroWordNet
SIMPLE, The Generative Lexicon



Supported by: The Danish Research Council



Driving reuse factor: Genus proximum

	Expanded le	Pos	GenProx	Definition
29	dispenser_1	sb.	beholder	beholder der afgiver sit ind
30	drikkehorn_1	sb.	beholder	beholder til at drikke af, la
31	dunk,1_1	sb.	beholder	beholder til at opbevare v
32	dybfryser_1	sb.	beholder	skabs- el. kummeformet b
33	dykkerklokke	sb.	beholder	stor, opr. klokkeformet be
34	dåse_1	sb.	beholder	hermetisk lukket, ofte cyli
35	dåse_2	sb.	beholder	cylinderformet beholder m
36	ekspansionsl	sb.	beholder	beholder i et centralvarme
37	fjerding_2_1	sb.	beholder	beholder (tønde el. anker)
38	flaske,1_1	sb.	beholder	aflang, rund beholder af gl
39	forlag,2_1	sb.	beholder	beholder hvori et destillat
40	form_5	sb.	beholder	beholder hvori der hældes
41	frostboks_2	sb.	beholder	beholder hvor genstande r
42	fryseboks_1	sb.	beholder	beholder hvor fødevarer n
43	fryser_1	sb.	beholder	beholder hvor fødevarer k
44	fuglebad_1	sb.	beholder	beholder el. kar som fugle
45	fustage_1	sb.	beholder	lukket, ofte cylinderforme
46	fyrværkeri_1	sb.	beholder	beholdere med krudt og f
47	førstehjælpsl	sb.	beholder	beholder med førstehjælp
48	gasbeholder_	sb.	beholder	beholder til opbevaring af
49	gasflaske_1	sb.	beholder	stor flaskelignende behol
50	gaskedel_1	sb.	beholder	stor beholder hvor vand op
51	glas_2	sb.	beholder	beholder af dette el. et an
52	gryde_1	sb.	beholder	åben, ofte cylindrisk behol
53	hylster_1	sb.	beholder	(aflang) beholder der er be
54	hæk,2_2	sb.	beholder	beholder med tremmer hv
55	iltflaske_1	sb.	beholder	beholder med ilt til teknisk
56	ionbytter_1	sb.	beholder	beholder indeholdende m



Qualia relations from definitions

Correlation between definitions given in traditional dictionaries and Qualia Structure

Meaning dimensions in terms of Qualia Structure:

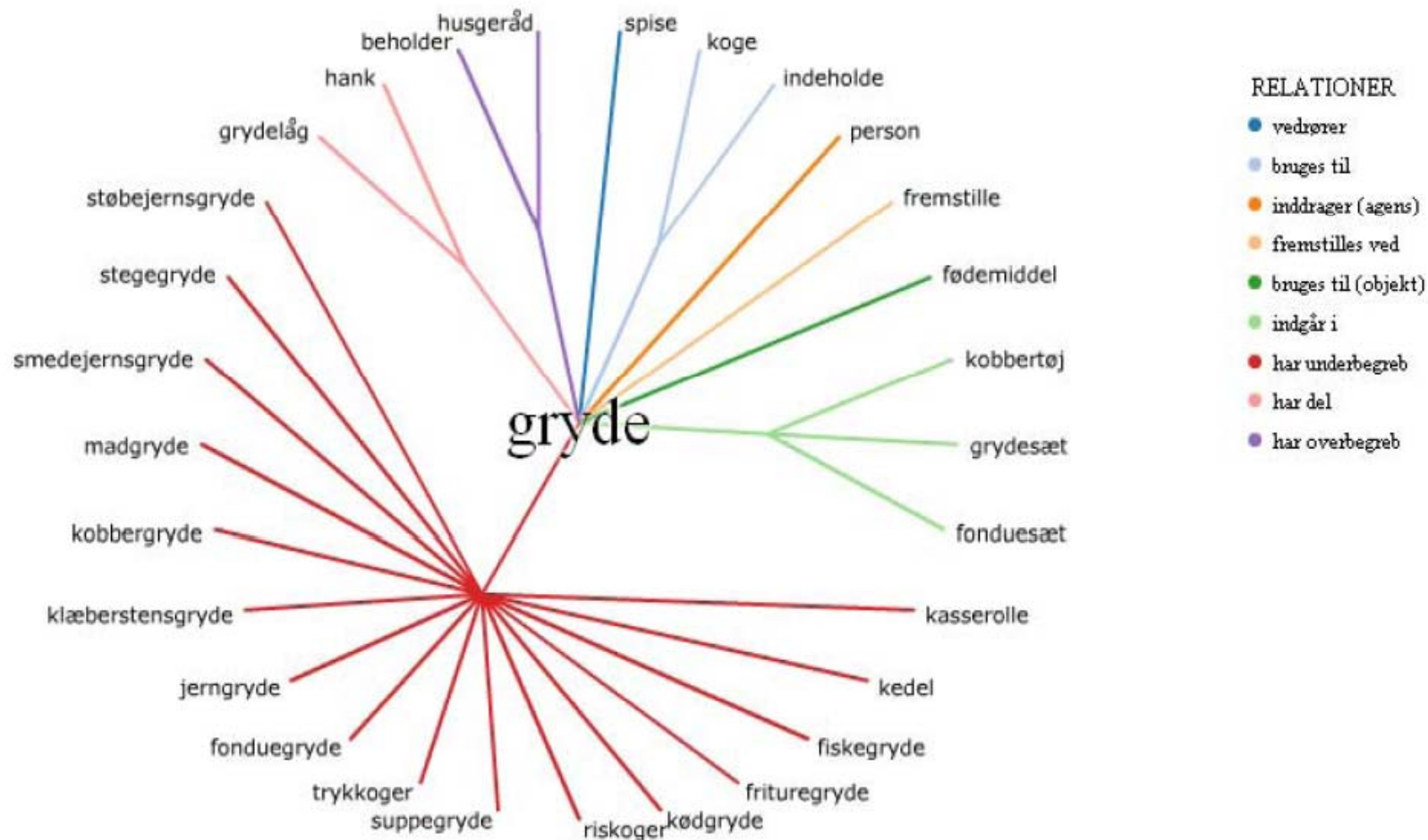
- **formal role** (is_a - genus proximum)
- **telic role** (used_for ..)
- **constitutive role** (has_as_parts..)
- **agentive role** (made_by)

Definition of "pot" in DDO:

A pot is a container (FORMAL), usually with two handles and a lid (CONSTITUTVE), used for cooking food (TELIC)

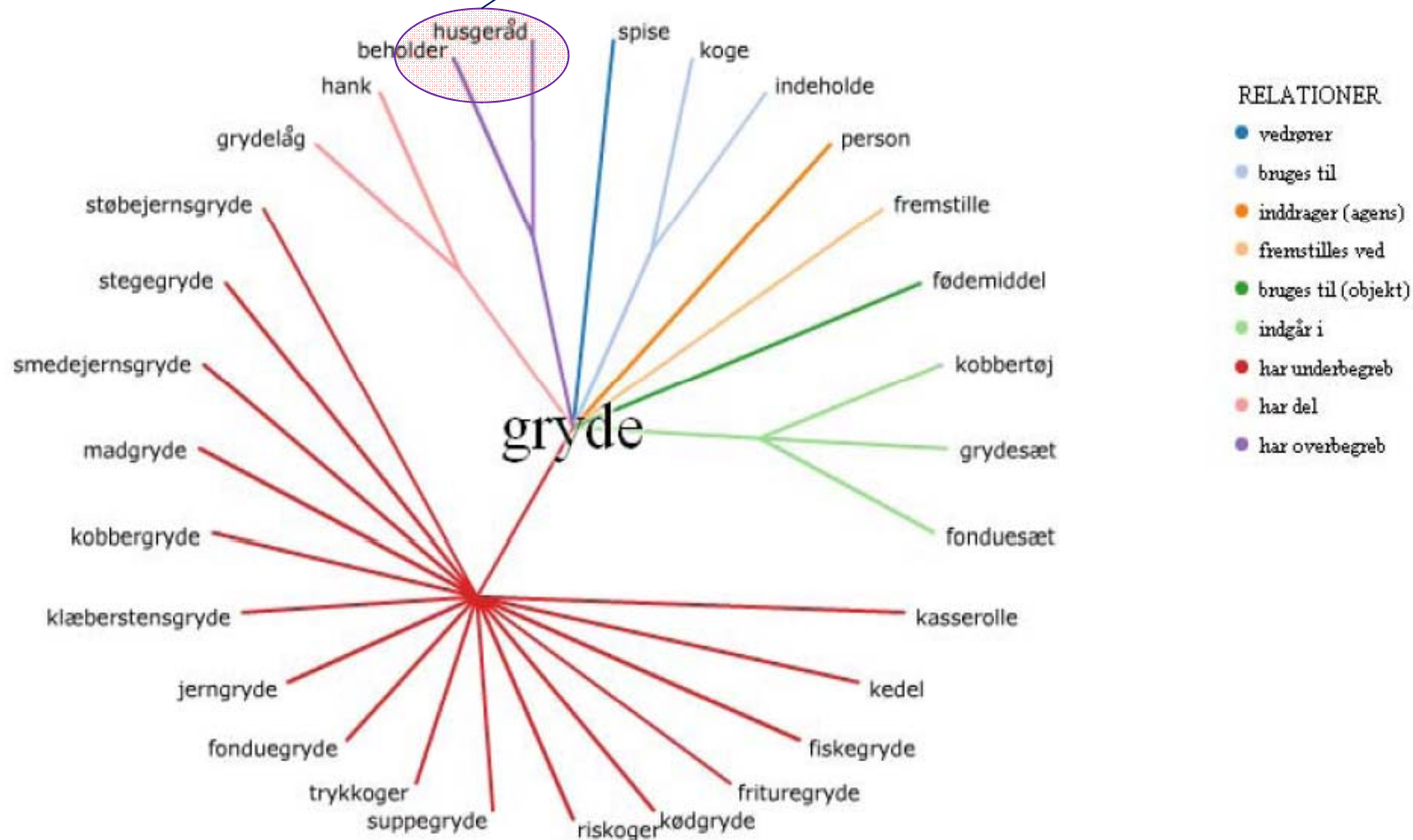
ii) Qualia relations from definitions and examples

Definition: Container (FORMAL), usually with two handles and a lid (CONSTITUTVE), used for cooking food (TELIC)



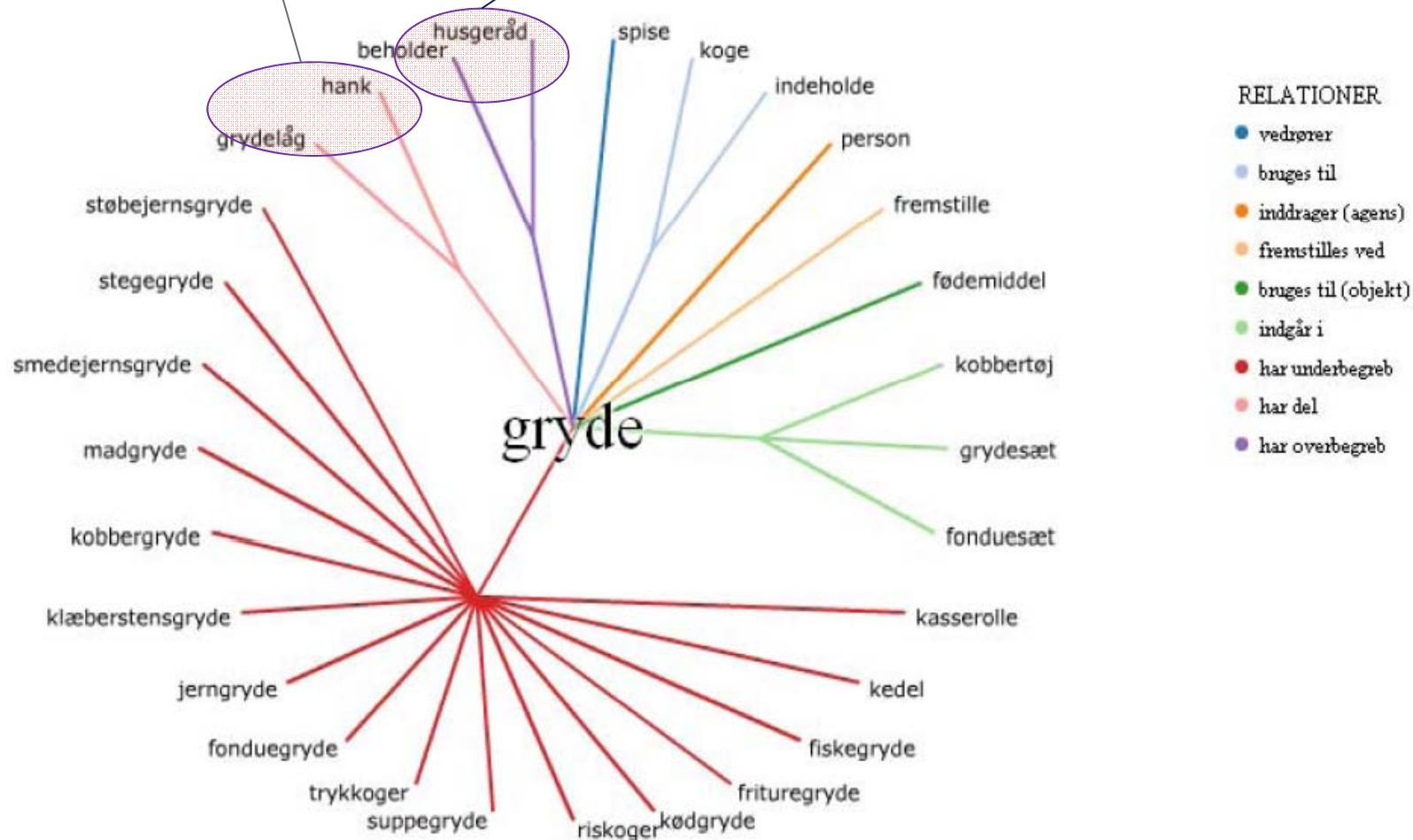
ii) Qualia relations from definitions and examples

Definition: Container (FORMAL), usually with two handles and a lid (CONSTITUTVE), used for cooking food (TELIC)



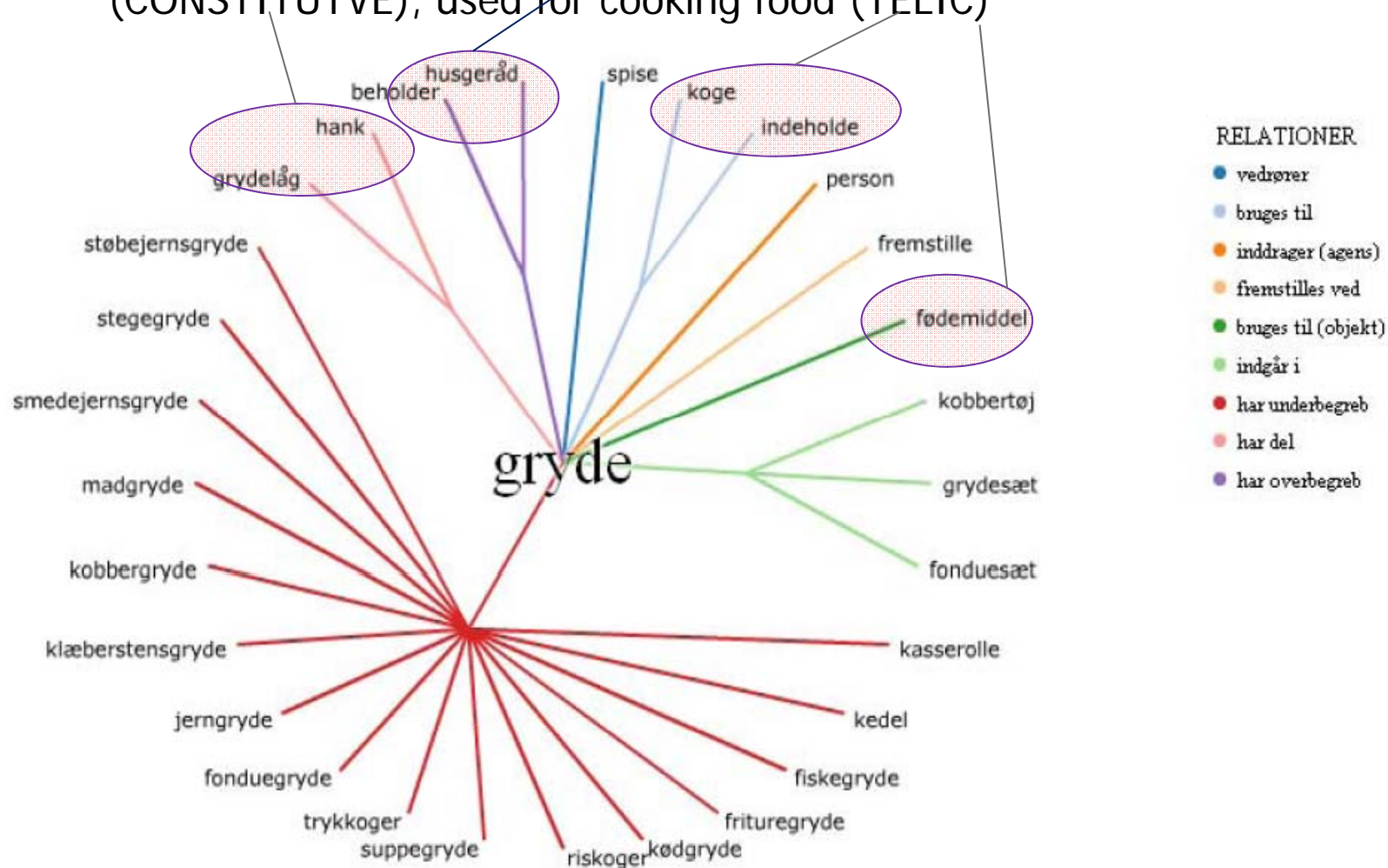
ii) Qualia relations from definitions and examples

Definition: Container (FORMAL), usually with two handles and a lid (CONSTITUTVE), used for cooking food (TELIC)



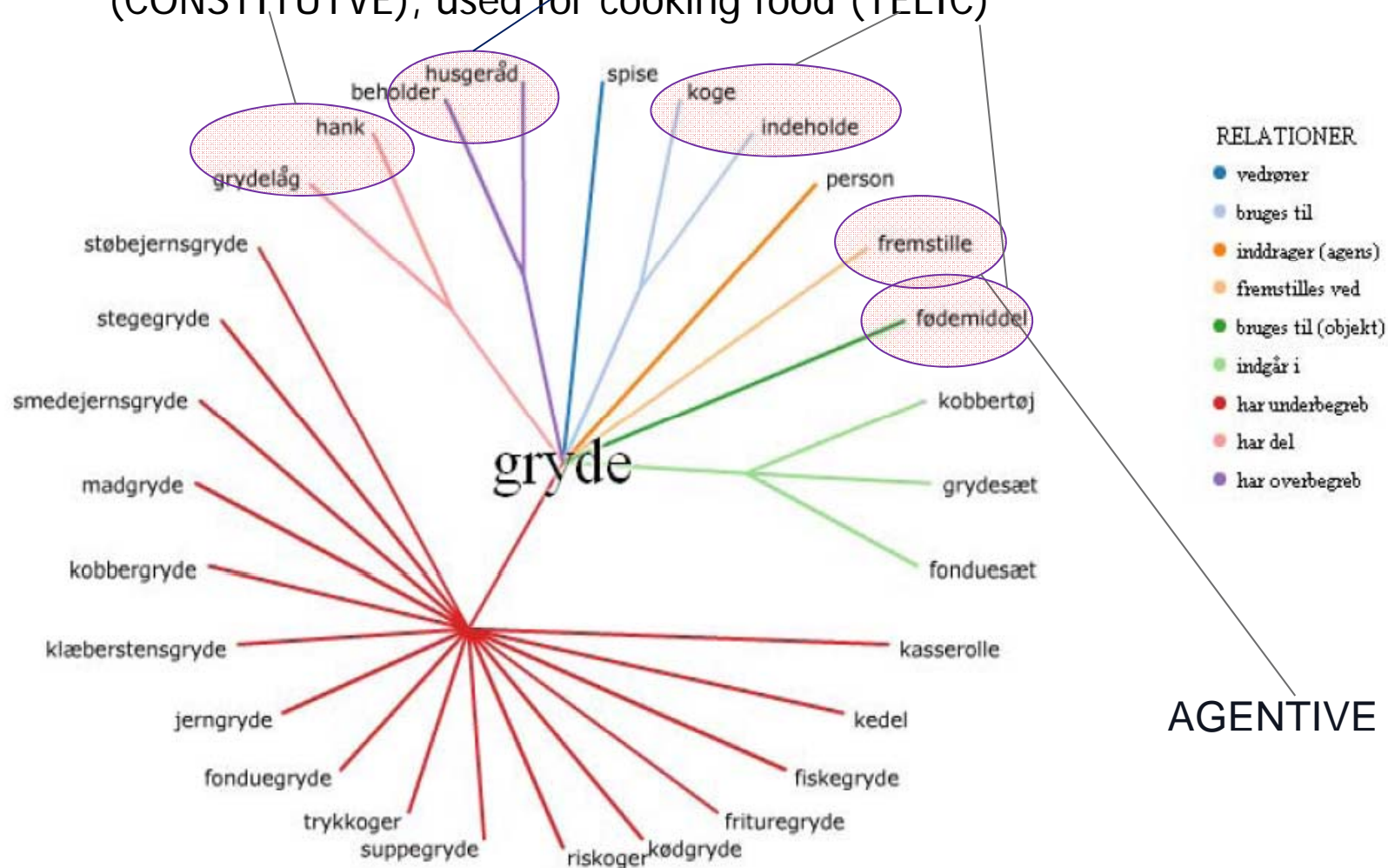
ii) Qualia relations from definitions and examples

Definition: Container (FORMAL), usually with two handles and a lid (CONSTITUTVE), used for cooking food (TELIC)



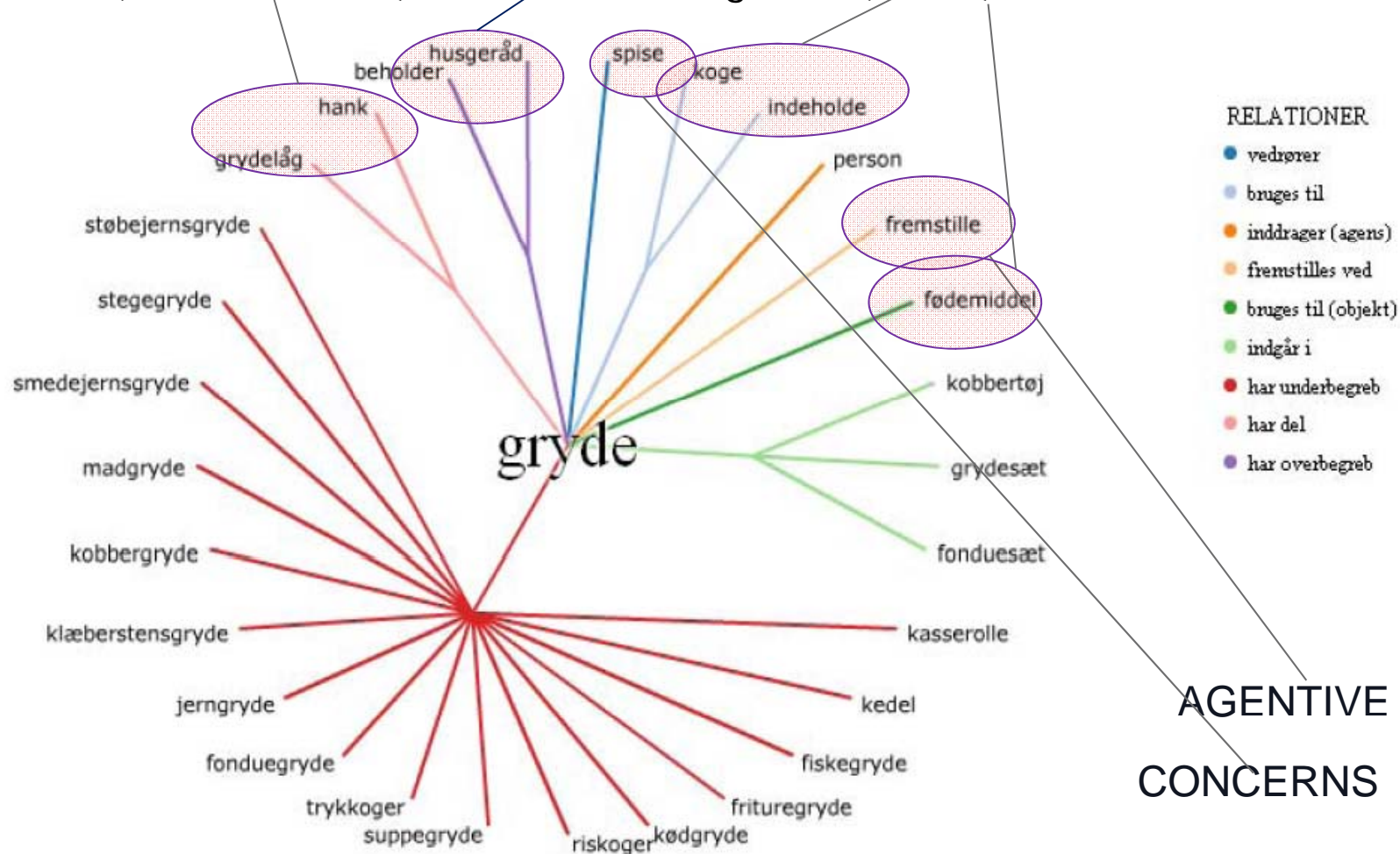
ii) Qualia relations from definitions and examples

Definition: Container (FORMAL), usually with two handles and a lid (CONSTITUTVE), used for cooking food (TELIC)



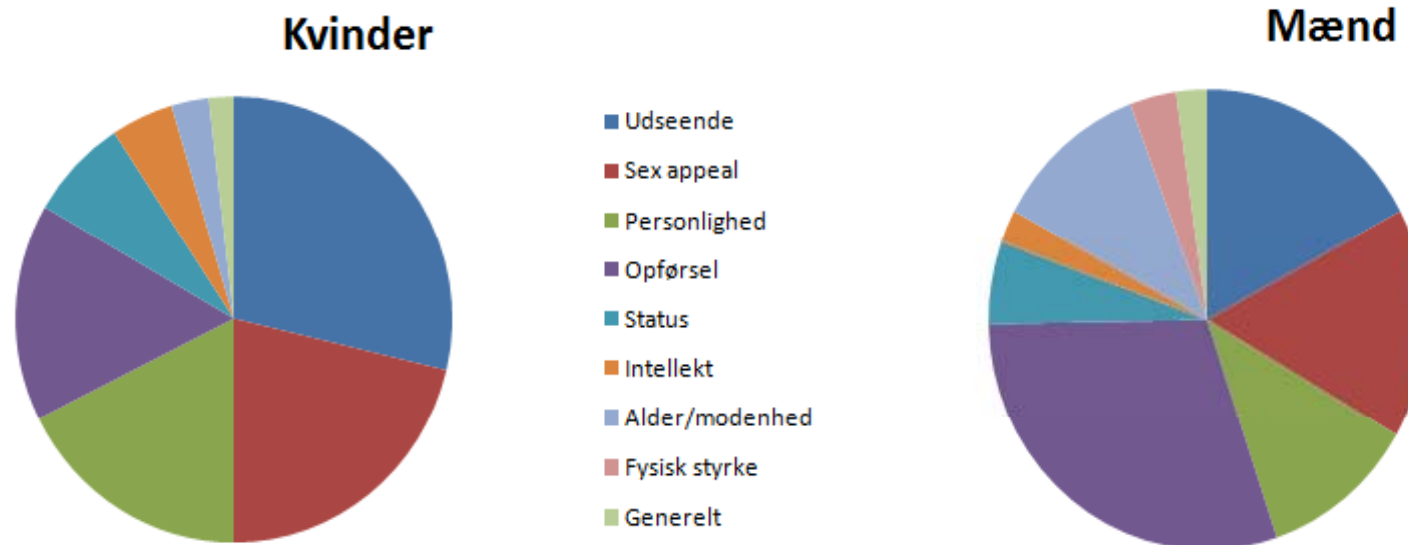
ii) Qualia relations from definitions and examples

Definition: Container (FORMAL), usually with two handles and a lid (CONSTITUTVE), used for cooking food (TELIC)



Extracting new knowledge about the vocabulary

Eks: Positive/negative connotation on 16% of all person references i DanNet:
rappenskralde, sild, skår, skude, vatnisse, fløs, kuvøseguf, laps, strandløve, træmand ... (Braasch & Pedersen 2010) (Pedersen & Braasch 2009)





Ambiguity and granularity – and eventual of merging senses

Homonymy: (pande (pan), pande (forehead)):
two headwords in the dictionary, each with
their own definition

Polysemy: (mus (mouse, animal), mus
(mouse, computer device)
one headword, two definitions

Regular polysemy: (*country* as geographical,
country as human group)

frequent cases: one headword, two
definitions

Infrequent cases: not listed

Eventually in same definition



Ambiguity and granularity – and eventual of merging senses

Different strategies:

1. Use most frequent senses
2. Collapse polysemous senses belonging to same ontological type (ex. *card* (Semiotic - Artifact))
3. Collapse senses with the same purpose (ex. *needle* (sewing, injecting, measuring))
4. Collapse subsenses
5. Collapse differently for different word classes (i.e. verbs tend to be very finegrained in lexicons)

Conclusions



- Consider which resource fits your annotation needs
- Consider strategy for eventual merging of senses
- Be prepared to accept and overcome discrepancies in coverage

Discussion



- Have you decided on lexical resource?
- How many choices do you have in your own language?
- Have you considered a strategy for eventual merging of senses?
- Are you encountering coverage problems?
- Other problems?