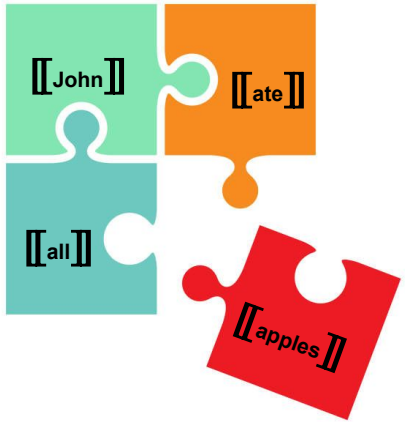
http://pmb.let.rug.nl

Compositional Semantics in The Parallel Meaning Bank



Lasha Abzianidze
(joint work with Johan Bos)

August 29, LACompLing 2018

Parallel Meaning Bank

university of groningen

How do they differ?

x1 x2 e1 t1
male.n.02(x1)
Name(x1, alfred~nobel)
invent.v.01(e1)
Time(e1, t1)
Result(e1, x2)
Agent(e1, x1)
time.n.08(t1)
YearOfCentury(t1, 1866)
t1 ≺ now
dynamite.n.01(x2)

x1 x2 e1 t1
male.n.02(x1)
Name(x1, alfred~nobel)
invent.v.01(e1)
Time(e1, t1)
Result(e1, x2)
Agent(e1, x1)
time.n.08(t1)
YearOfCentury(t1, 1866)
t1 ≺ now
dynamite.n.01(x2)

How do they differ?

Alfred Nobel invented dynamite in 1866.

x1 x2 e1 t1
male.n.02(x1)
Name(x1, alfred~nobel)
invent.v.01(e1)
Time(e1, t1)
Result(e1, x2)
Agent(e1, x1)
time.n.08(t1)
YearOfCentury(t1, 1866)
t1 < now
dynamite.n.01(x2)

x1 x2 e1 t1
male.n.02(x1)
Name(x1, alfred~nobel)
invent.v.01(e1)
Time(e1, t1)
Result(e1, x2)
Agent(e1, x1)
time.n.08(t1)
YearOfCentury(t1, 1866)
t1 < now
dynamite.n.01(x2)

Alfred Nobel vond in 1866 het dynamiet uit.

They don't differ

Alfred Nobel erfand 1866 das Dynamit.

Alfred Nobel invented dynamite in 1866.

x1 x2 e1 t1

male.n.02(x1)
  Name(x1, alfred~nobel)
invent.v.01(e1)
  Time(e1, t1)
  Result(e1, x2)
  Agent(e1, x1)
time.n.08(t1)
  YearOfCentury(t1, 1866)
  t1 ≺ now
dynamite.n.01(x2)

x1 x2 e1 t1

male.n.02(x1)
  Name(x1, alfred~nobel)
invent.v.01(e1)
  Time(e1, t1)
  Result(e1, x2)
  Agent(e1, x1)
time.n.08(t1)
  YearOfCentury(t1, 1866)
  t1 ≺ now
dynamite.n.01(x2)

x1 x2 e1 t1

male.n.02(x1)
  Name(x1, alfred~nobel)
invent.v.01(e1)
  Time(e1, t1)
  Result(e1, x2)
  Agent(e1, x1)
time.n.08(t1)
  YearOfCentury(t1, 1866)
  t1 ≺ now
dynamite.n.01(x2)

x1 x2 e1 t1

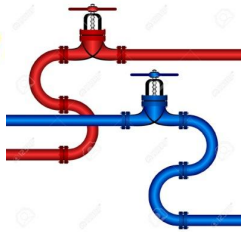
male.n.02(x1)
  Name(x1, alfred~nobel)
invent.v.01(e1)
  Time(e1, t1)
  Result(e1, x2)
  Agent(e1, x1)
time.n.08(t1)
  YearOfCentury(t1, 1866)
  t1 ≺ now
dynamite.n.01(x2)

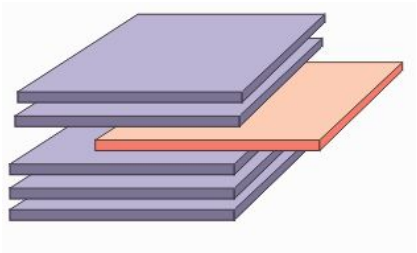
Alfred Nobel vond in 1866 het dynamiet uit.

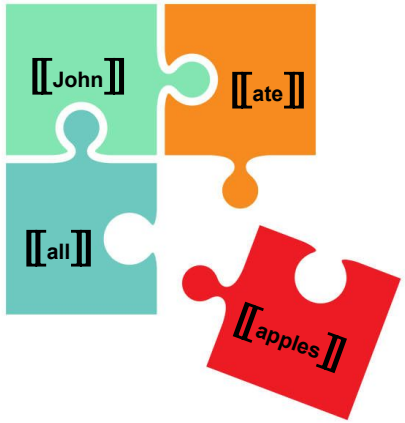
Alfred Nobel inventò la dinamite nel 1866.

Plan

Universal
Semantic Tagging

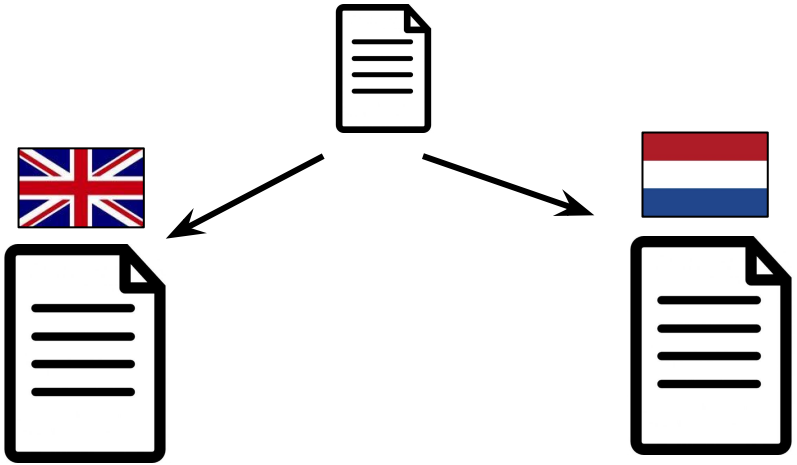




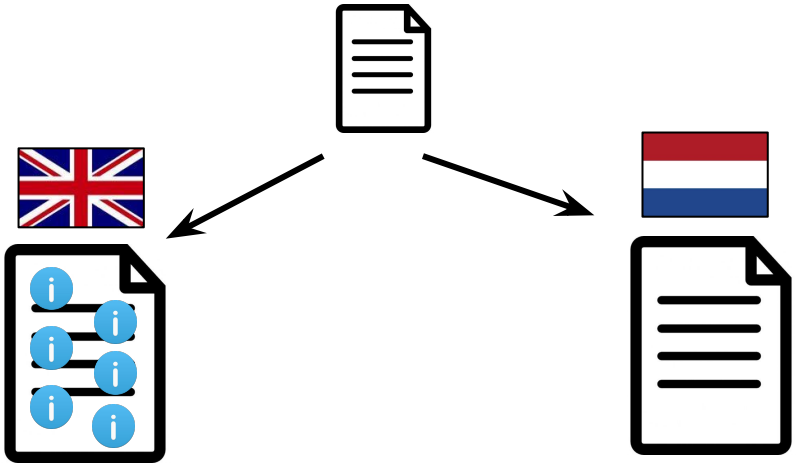


Basic idea behind the PMB

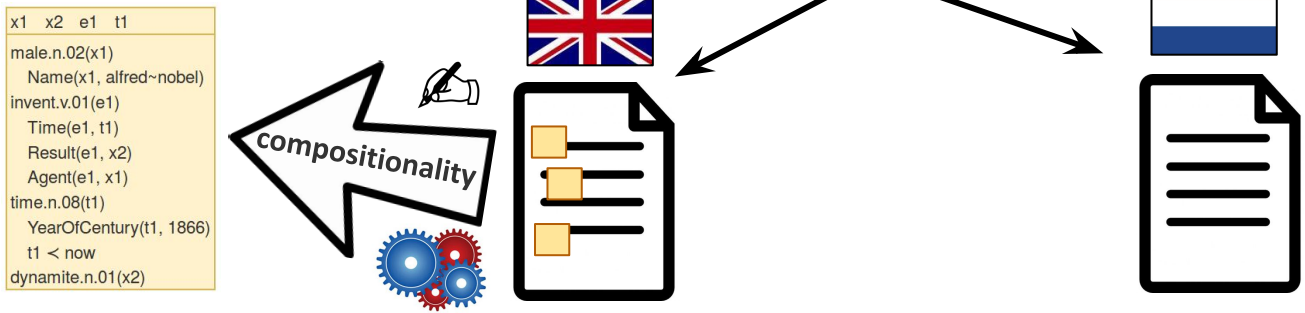
Basic idea behind the PMB



Basic idea behind the PMB

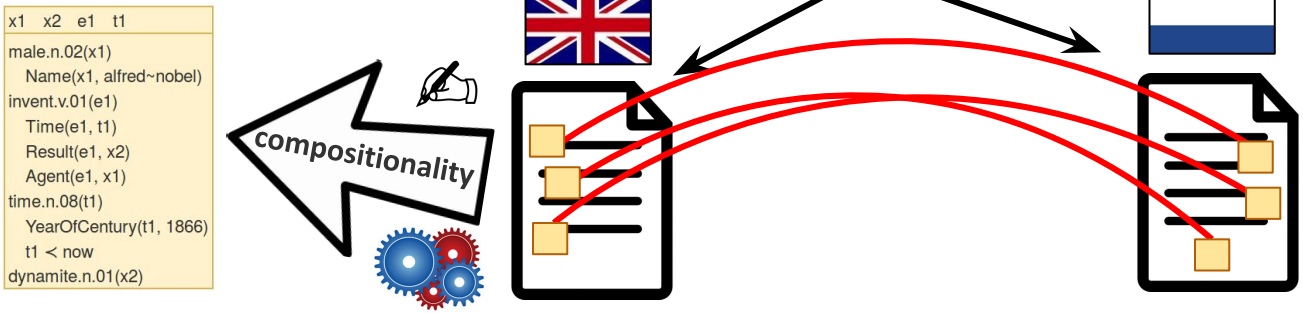


Basic idea behind the PMB

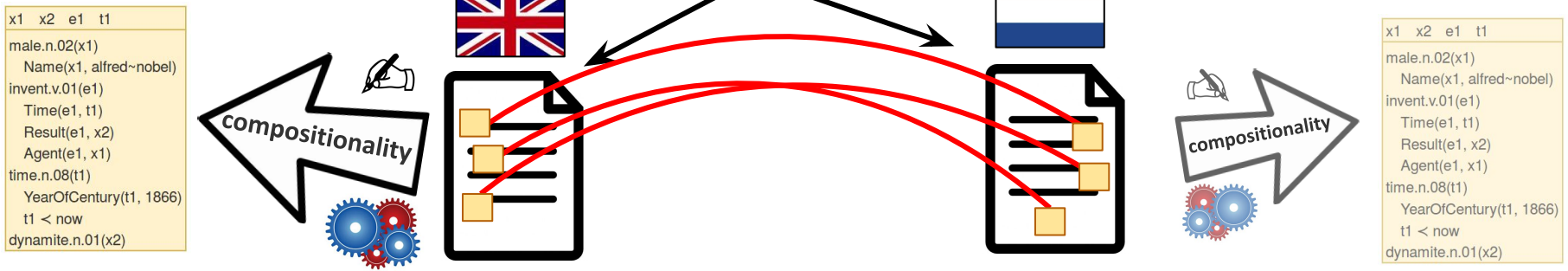


Groningen MEANING BANK

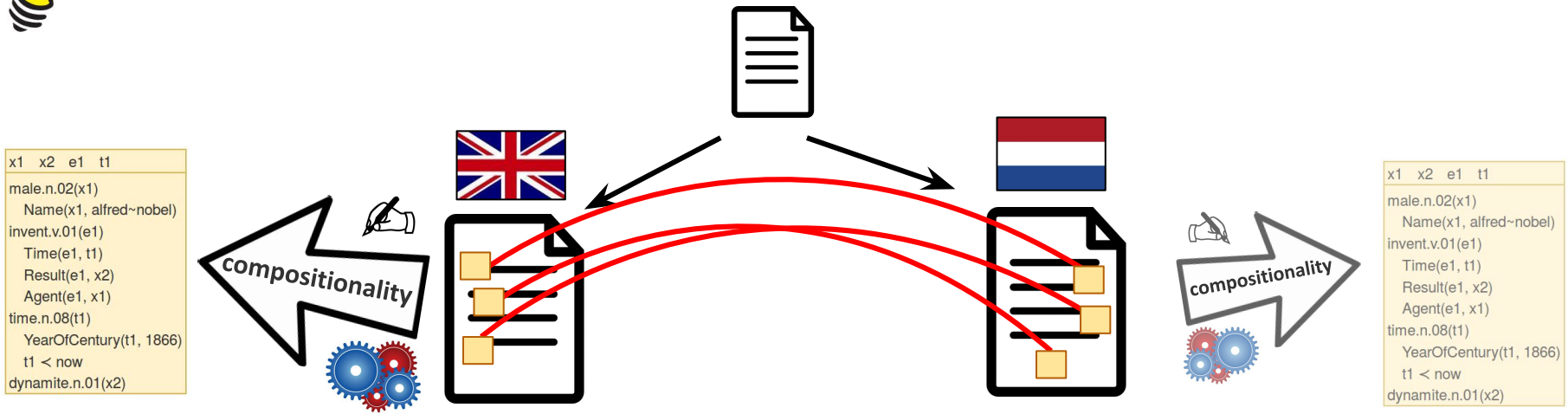
Basic idea behind the PMB



Basic idea behind the PMB



Basic idea behind the PMB

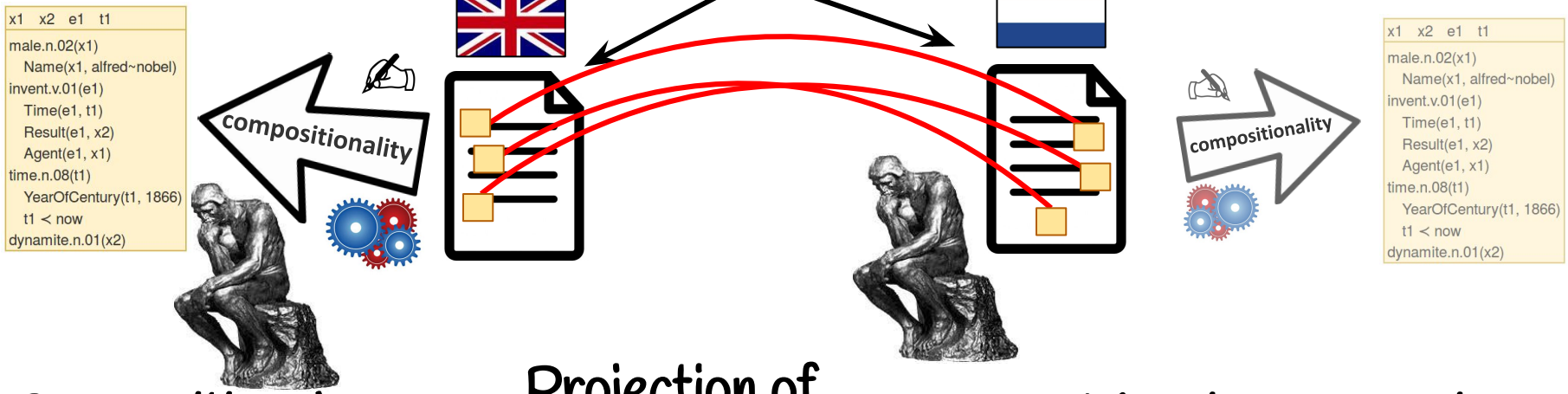


Compositional Semantics

Projection of lexical semantics

Meaning preserving translations

Basic idea behind the PMB



Compositional Semantics

Projection of lexical semantics

Meaning preserving translations

Objectives of the PMB project

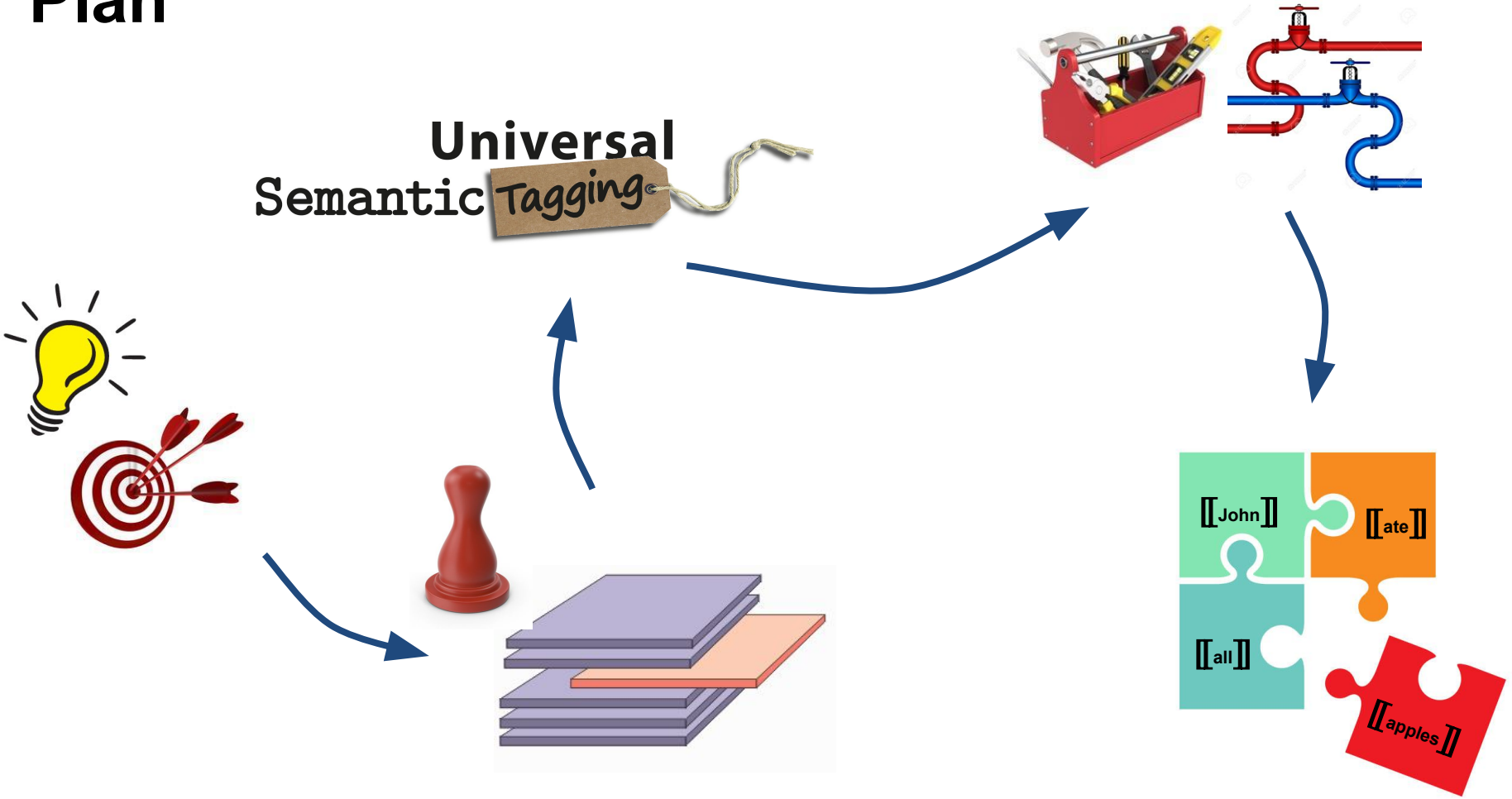
- Parallel corpus annotated with formal meaning representations
- Semantic tools for languages other than English
  - One system with language-specific models
- Learn more about (cross-lingual) compositional semantics

- Study translations form a formal semantic perspective
- Improve translations

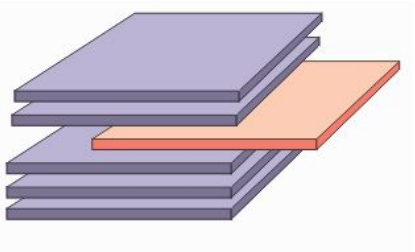
Design of the PMB

- Several (but not too many) annotation layers
- Collaborative annotation:
  - Machines
  - Experts
  - The crowd
- Distinguish Bronze/Silver/Gold annotation layers

Plan



Language-neutral annotation layers



- Segmentation
- Semantic tagging (~~part-of-speech tagging~~)
- Symbolization (~~lemmatization~~)
- Word sense disambiguation (Wordnet 3.0; Miller, 1994)
- Semantic role labeling (Verbnet roles; Bonial et al, 2011)
- Syntactic parsing (Combinatory Categorial Grammar)
- Semantic parsing (Discourse representation theory)

Segmentation (challenges)

- Split texts into sentences
  - John said “I won’t go. I am lazy”.
- Split sentences into “meaningful atoms/words”
  - San Diego, Secretary of State, Royal Bank of Scotland, …
  - baseball club, knitting needles, pair of scissors …
  - unhappy, impossible, disagree, …
  - ten-year-old, data-driven, New York-based …
  - as well as, instead of, again and again, …

Symbolization

Mapping words to non-logical symbols

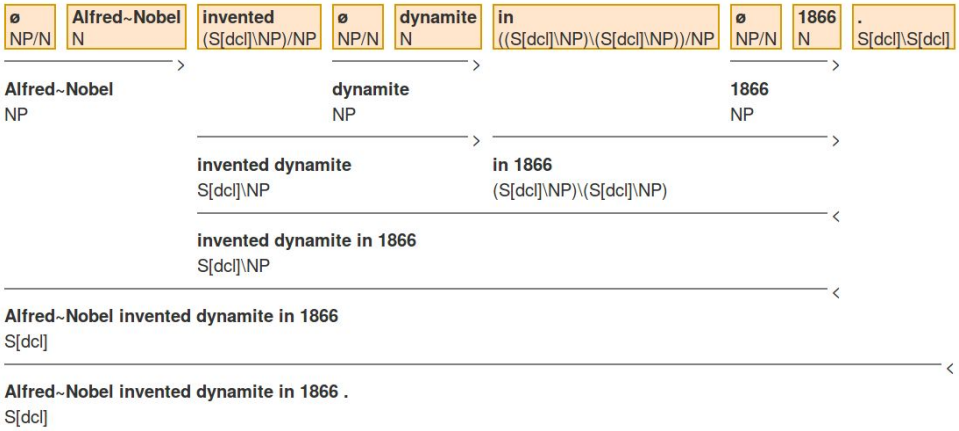
- Lemmatization
- Normalization

token	symbol
third	3
men	man
played	play
2:30 pm	14:30
2,5 million	2500000
km	kilometer

Syntactic parsing

with Combinatory categorial grammar (CCG; Steedman, 2000)

- Goes well with compositional semantics
- Efficient and wide-coverage parsers are available
  - C&C (Clark & Curran, 2007)
  - EasyCCG (Lewis & Steedmam, 2014)
  - DepCCG (Yoshikawa et al, 2017)
  - EasySRL (Lewis et al., 2016)
- Lexicalized grammar



Semantic parsing

with Discourse Representation Theory (DRT; Kamp & Reyle, 1993)

Difference with Kamp's DRT:

- Projective DRT (Venhuizen, 2015)
- Neo-Davidsonian event analysis
- No analysis of plurals yet
- Just three tenses (no aspect)
- WordNet synsets as concepts
- Presupposition as anaphora (van der Sandt, 1993)

x1 x2 e1 t1
male.n.02(x1)
Name(x1, alfred~nobel)
invent.v.01(e1)
Time(e1, t1)
Result(e1, x2)
Agent(e1, x1)
time.n.08(t1)
YearOfCentury(t1, 1866)
t1 ≺ now
dynamite.n.01(x2)

Discourse representation structures has:

- Logical symbols: comparison & Boolean operators, and variables
- Non-logical symbols: concepts, relations, and normalizations

Semantic parsing

with Discourse Representation Theory (DRT; Kamp & Reyle, 1993)

Difference with Kamp's DRT:

- Projective DRT (Venhuizen, 2015)
- Neo-Davidsonian event analysis
- No analysis of plurals yet
- Just three tenses (no aspect)
- WordNet synsets as concepts
- Presupposition as anaphora (van der Sandt, 19

b2

b1 ← x1 b2 ← x2 b2 ← e1 b3 ← t1
b1 ← male(x1)
b1 ← Name(x1, alfred~nobel)
b2 ← invent(e1)
b2 ← Time(e1, t1)
b2 ← Result(e1, x2)
b2 ← Agent(e1, x1)
b3 ← time(t1)
b3 ← YearOfCentury(t1, 1866)
b3 ← t1 < now
b2 ← dynamite(x2)

Discourse representation structures has:

- Logical symbols: comparison & Boolean operators, and variables
- Non-logical symbols: concepts, relations, and normalizations

Language-neutral annotation tools

- Segmentation: elephant (Evang et al., 2013)
- Symbolization: **baseline** (conditioned unigrams)
- Word sense disambiguation (Wordnet 3.0): no
- Semantic role labeling (Verbnet roles): no
- Syntactic parsing: **EasyCCG** (Lewis & Steedman, 2014)
- Semantic parsing with DRT: **Boxer** (Bos, 2008; 2015)

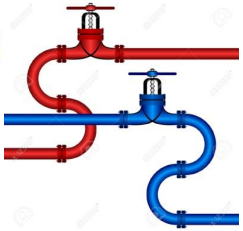
Language-neutral annotation tools

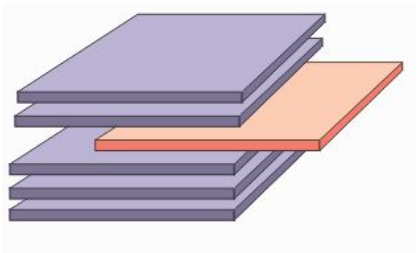
- Segmentation: elephant (Evang et al., 2013)
- Symbolization: **baseline** (conditioned unigrams)
- Word sense disambiguation (Wordnet 3.0): no
- Semantic role labeling (Verbnet roles): no
- Syntactic parsing: **EasyCCG** (Lewis & Steedman, 2014)
- Semantic parsing with DRT: **Boxer** (Bos, 2008; 2015)

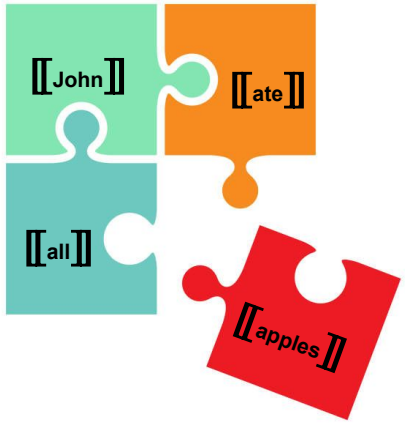
Plan

Universal

Semantic Tagging

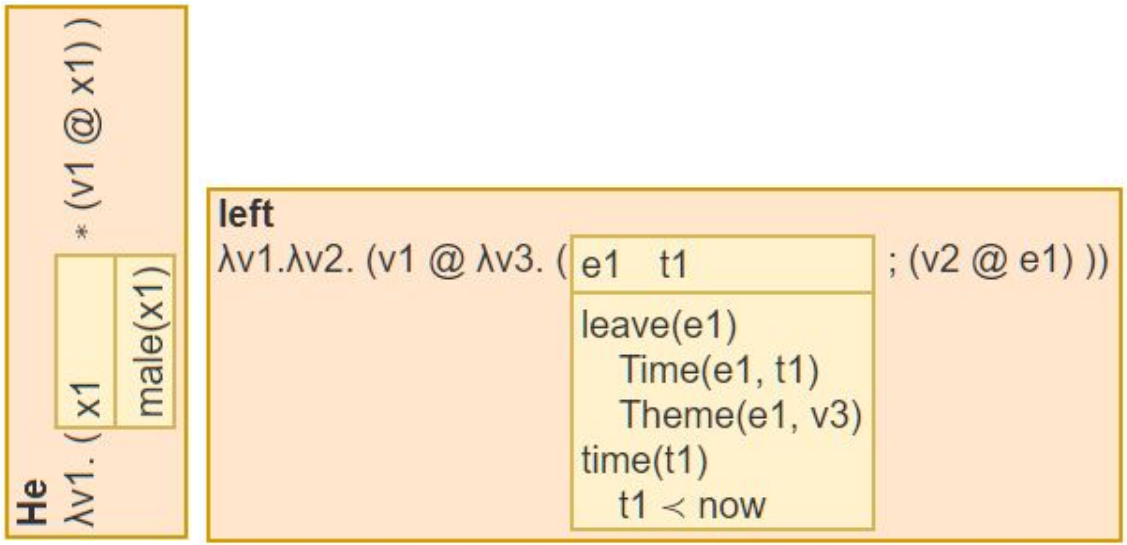




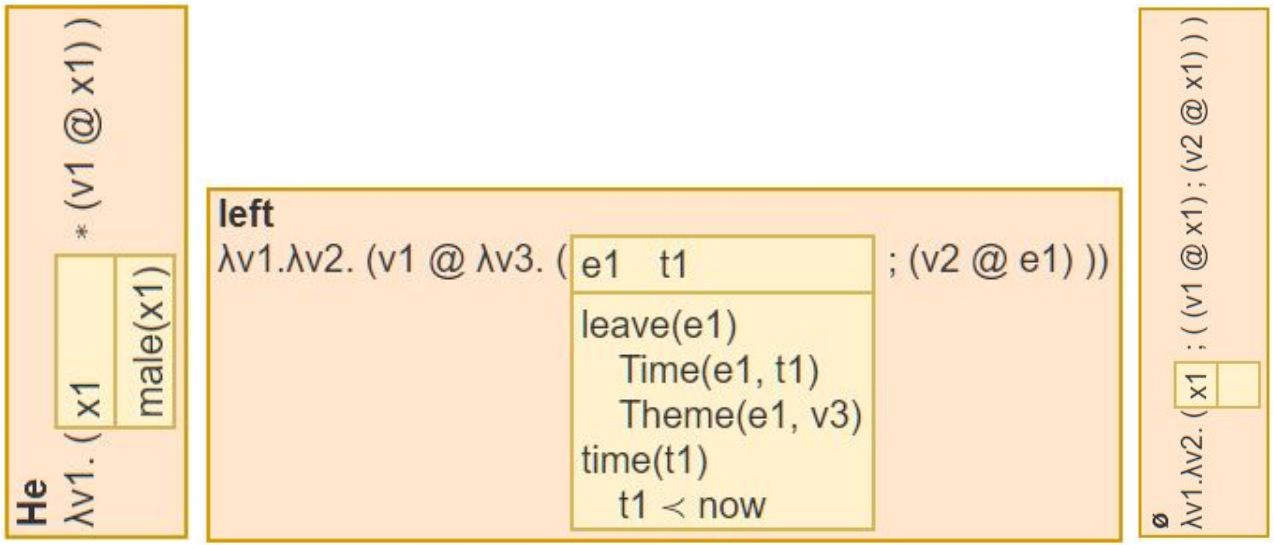


Idea behind Universal Semantic Tagging

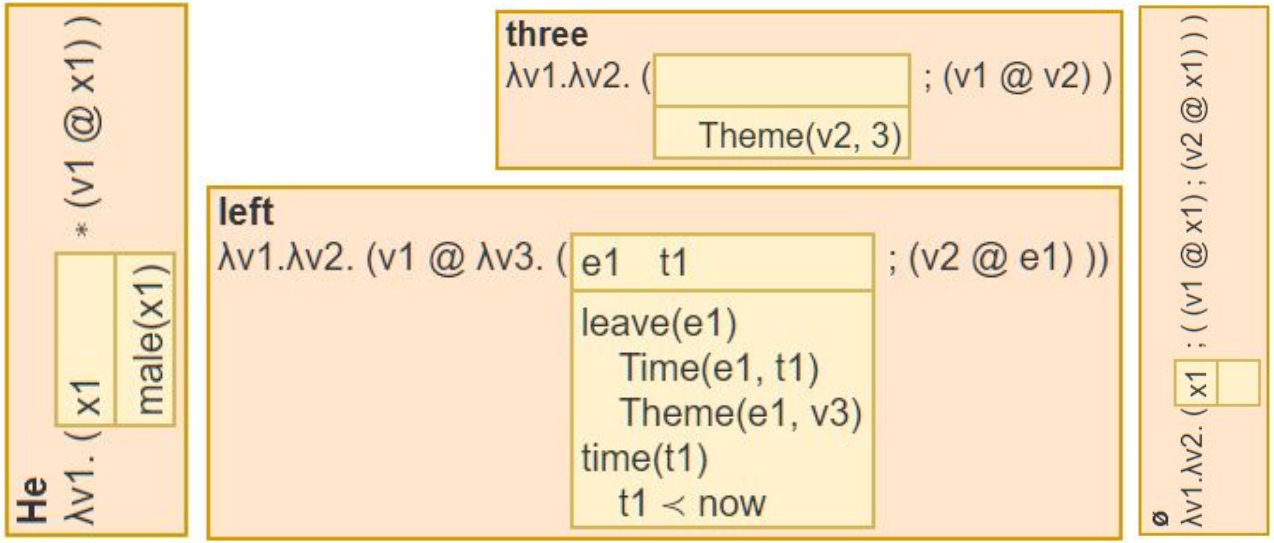
Idea behind Universal Semantic Tagging



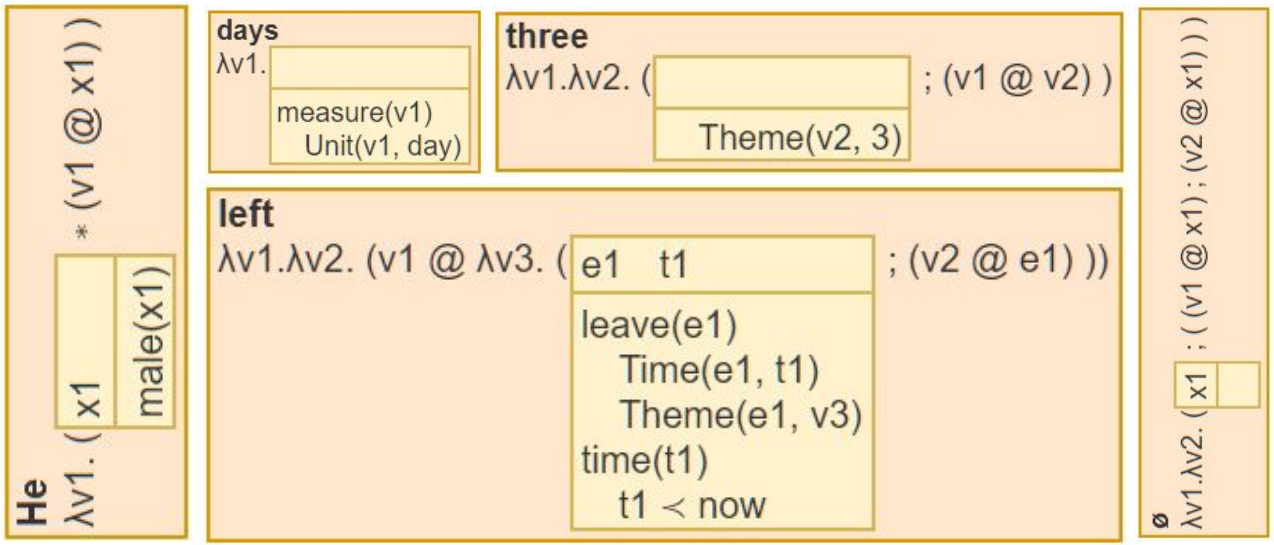
Idea behind Universal Semantic Tagging



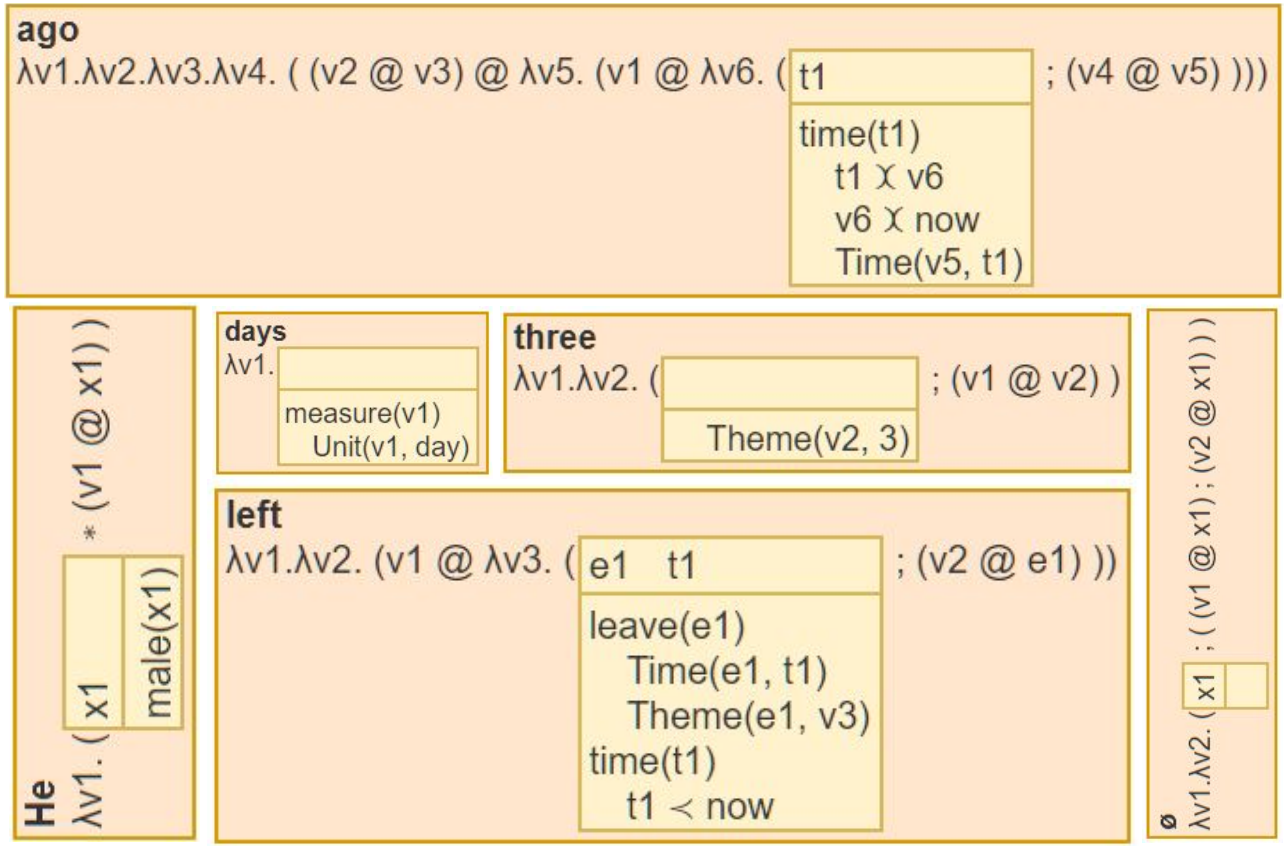
Idea behind Universal Semantic Tagging



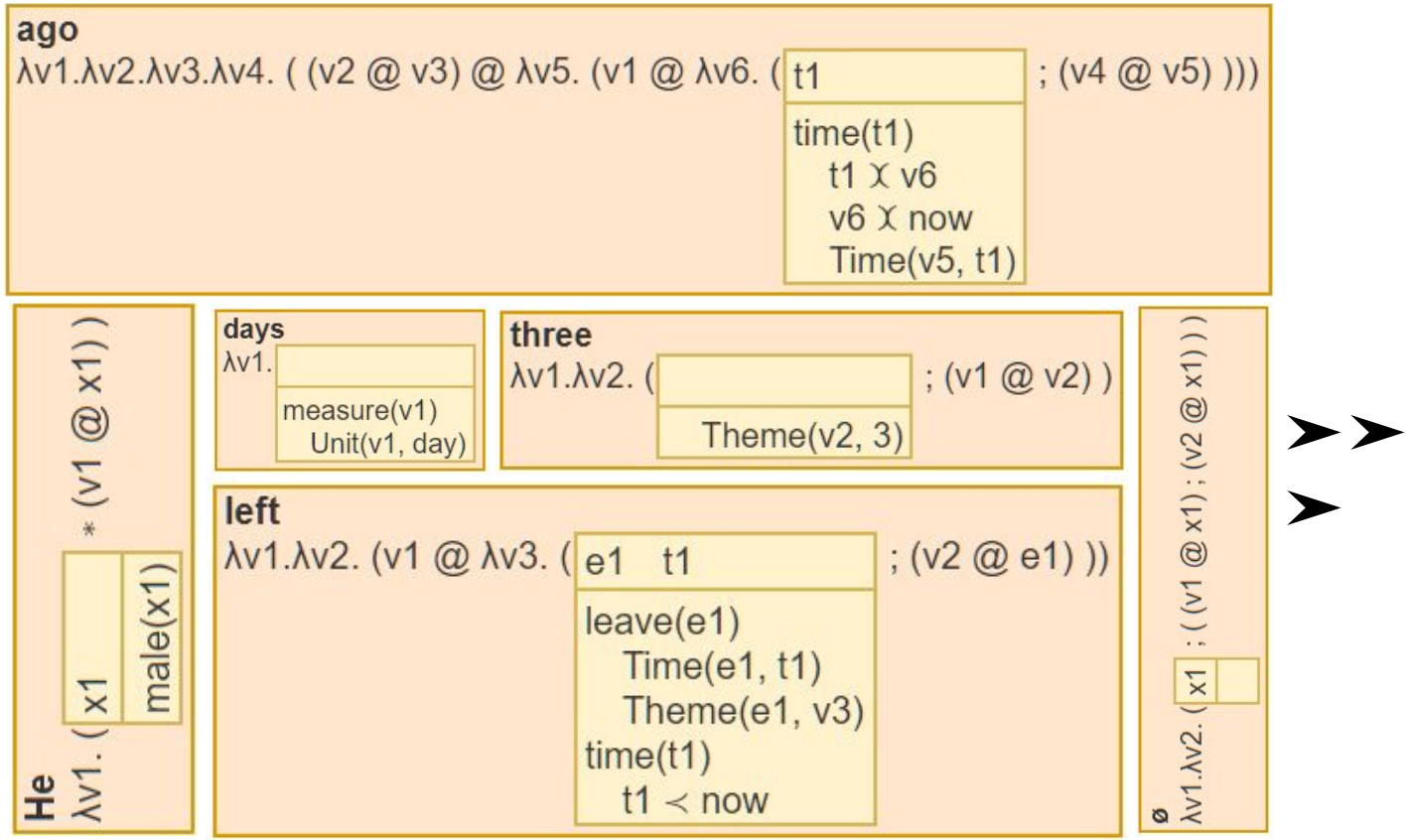
Idea behind Universal Semantic Tagging



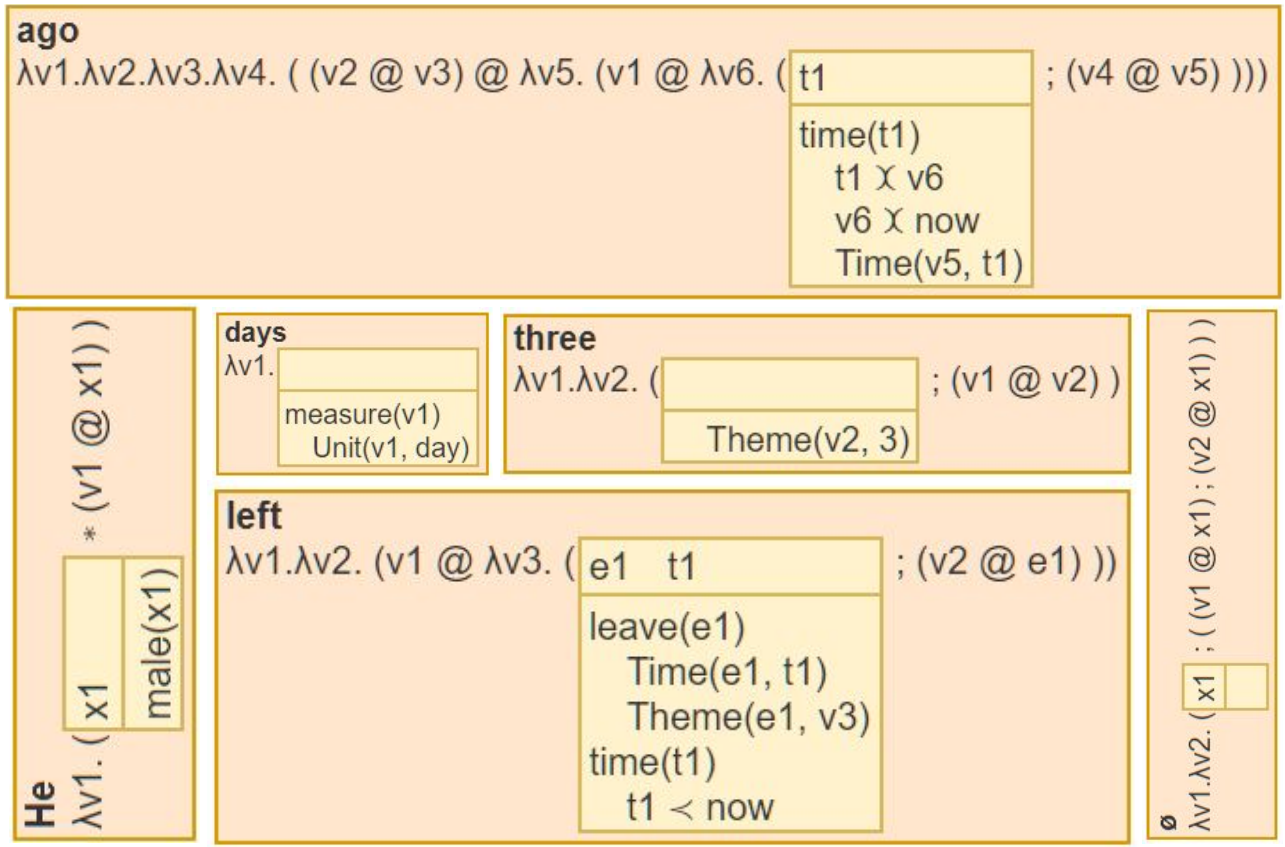
Idea behind Universal Semantic Tagging



Idea behind Universal Semantic Tagging

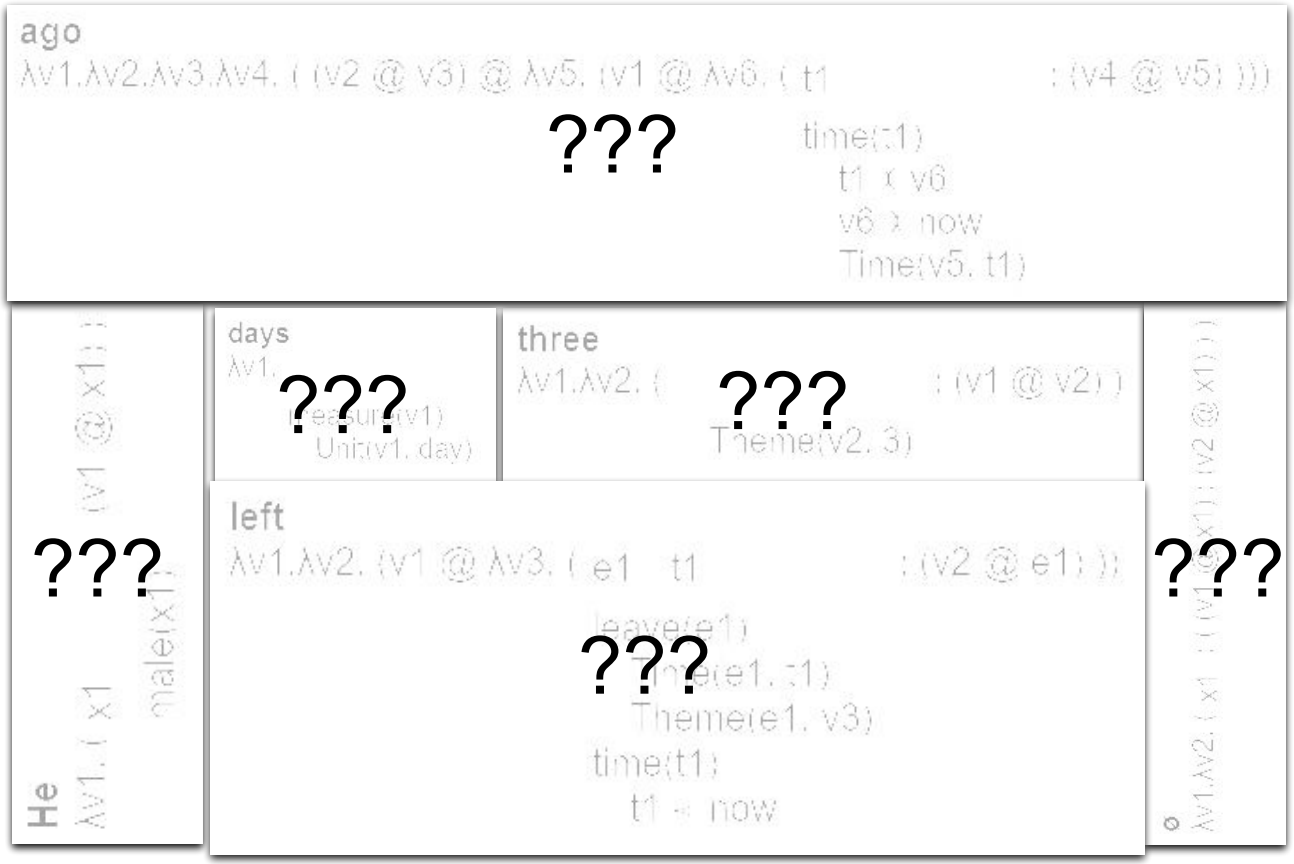


Idea behind Universal Semantic Tagging



x1	e1	t1	t2
male.n.02(x1)			
leave.v.01(e1)			
	Time(e1, t1)		
	Theme(e1, x1)		
time.n.08(t1)			
	t1 ⊀ t2		
	t1 ≺ now		
measure.n.02(t2)			
	t2 ⊀ now		
	Unit(t2, day)		
	Theme(t2, 3)		

Idea behind Universal Semantic Tagging



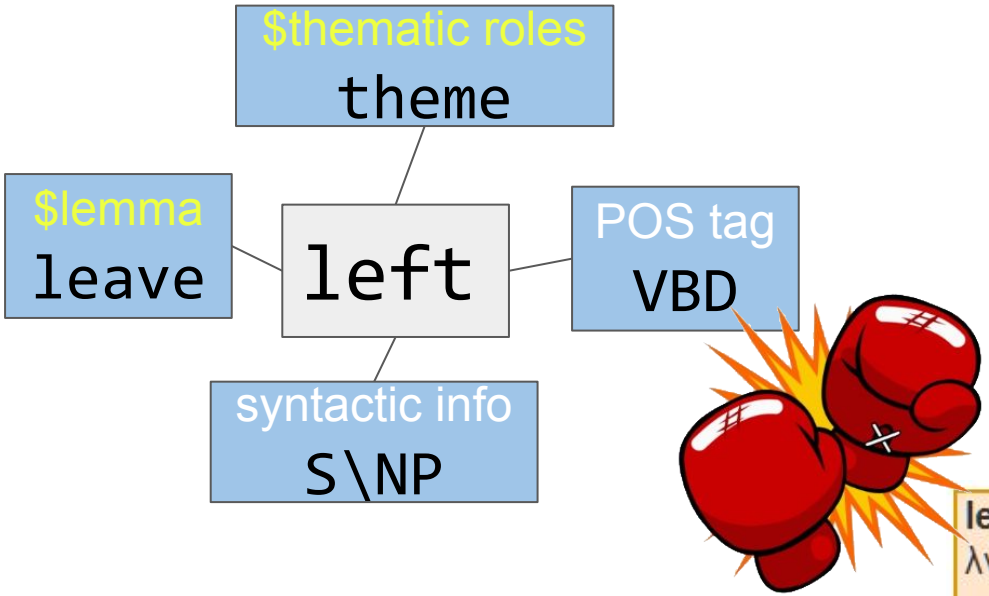
➤➤➤

Goal

x1	e1	t1	t2

male.n.02(x1)
leave.v.01(e1)
Time(e1, t1)
Theme(e1, x1)
time.n.08(t1)
t1 ⊼ t2
t1 ≺ now
measure.n.02(t2)
t2 ⊼ now
Unit(t2, day)
Theme(t2, 3)

Before semantic tags

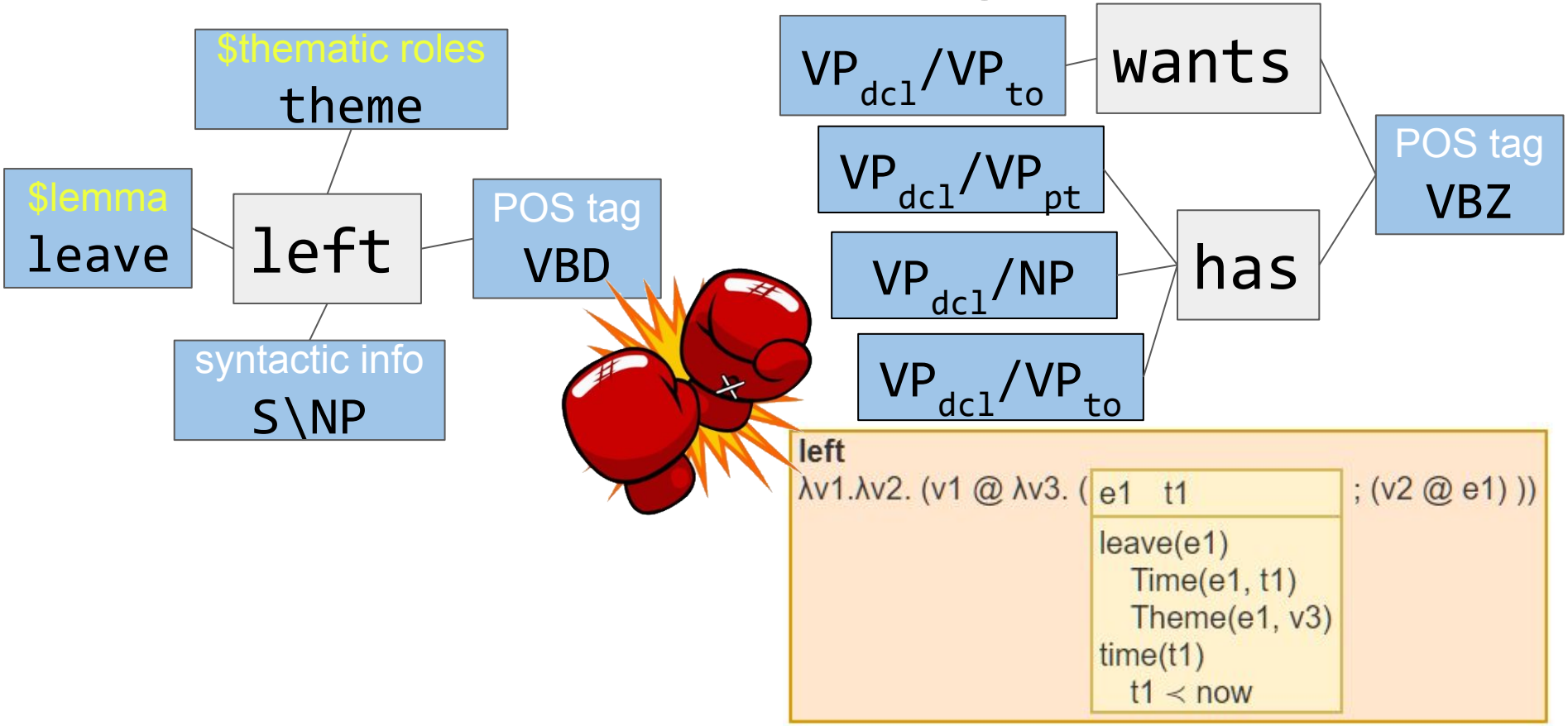


left

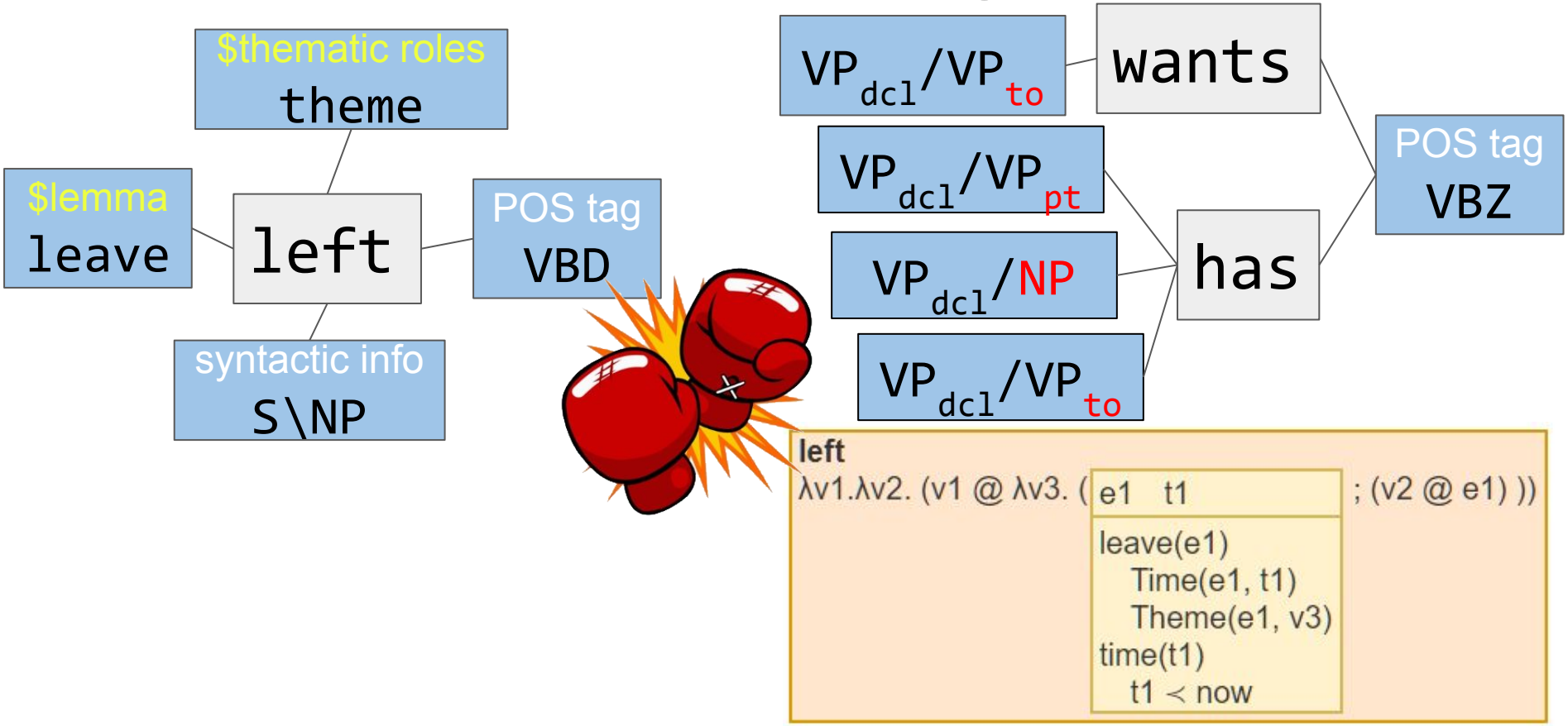
λv1.λv2. (v1 @ λv3. (e1 t1 ; (v2 @ e1)))

leave(e1)
Time(e1, t1)
Theme(e1, v3)
time(t1)
t1 ≺ now

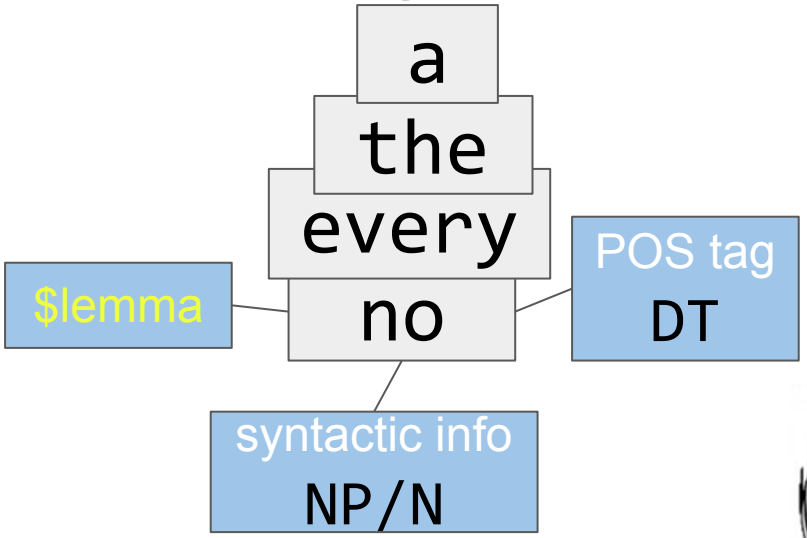
Before semantic tags

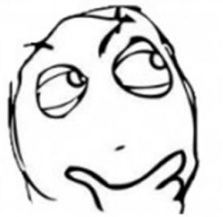


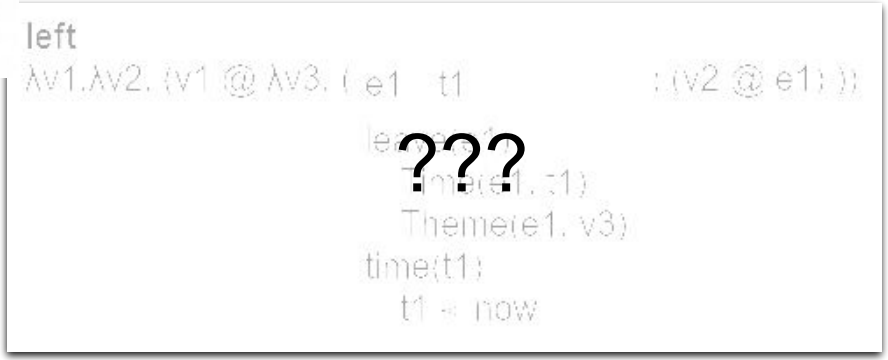
Before semantic tags



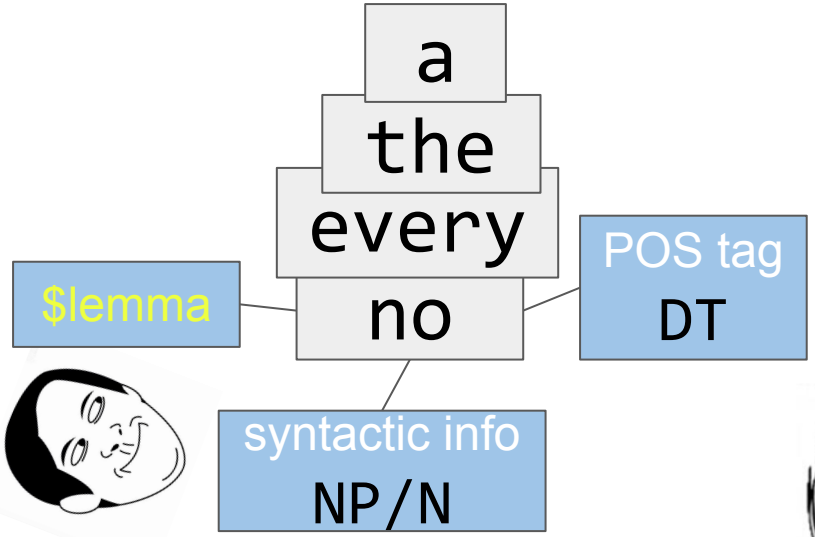
POS tags are not enough

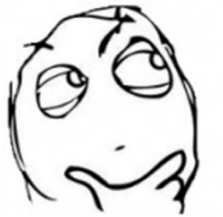






POS tags are not enough



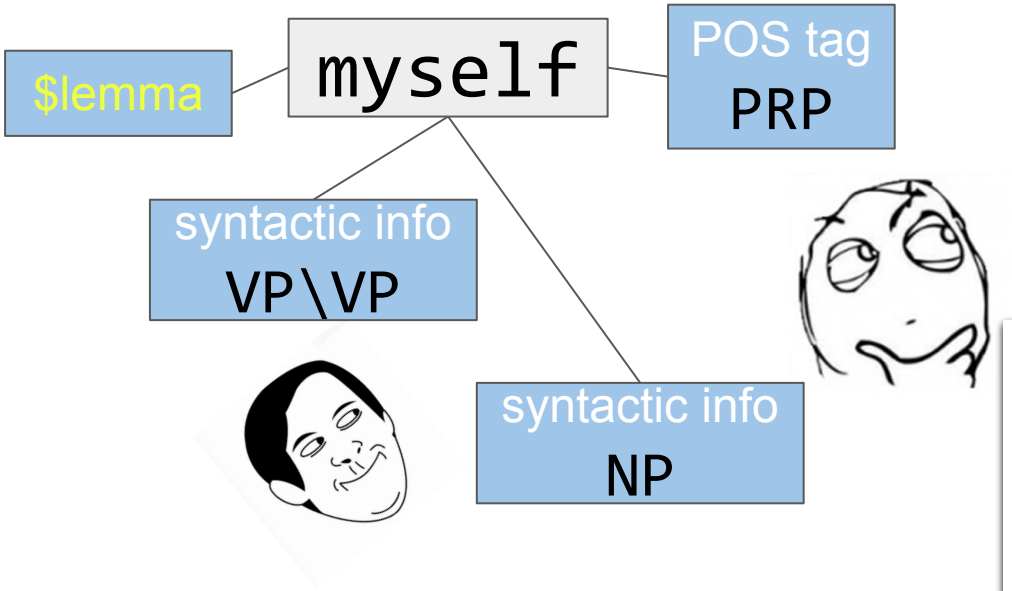


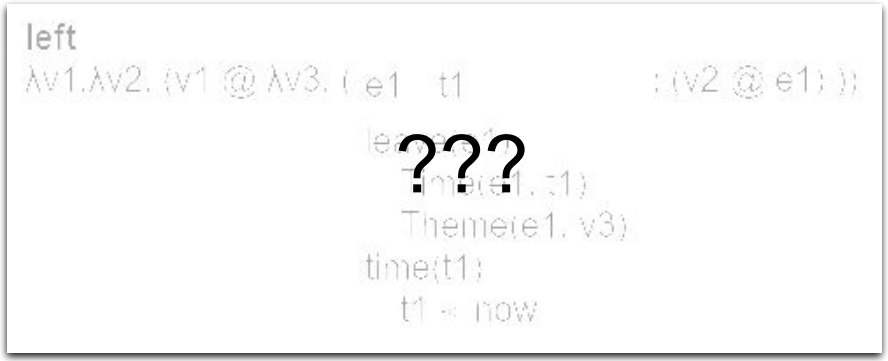
left
λv1.λv2. (v1 @ λv3. (e1 t1 : (v2 @ e1)))

???

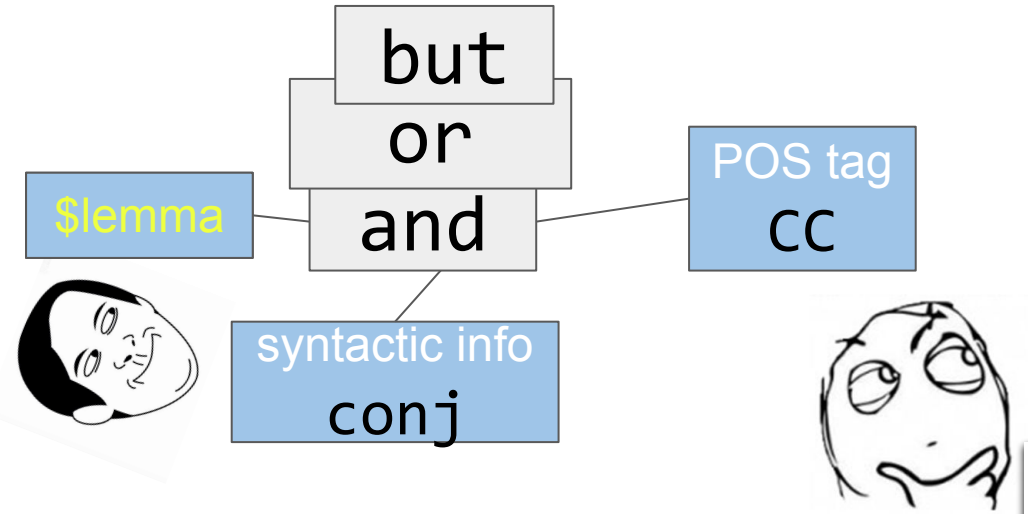
Theme(e1, t1)
Theme(e1, v3)
time(t1)
t1 = now

POS tags are not enough





POS tags are not enough

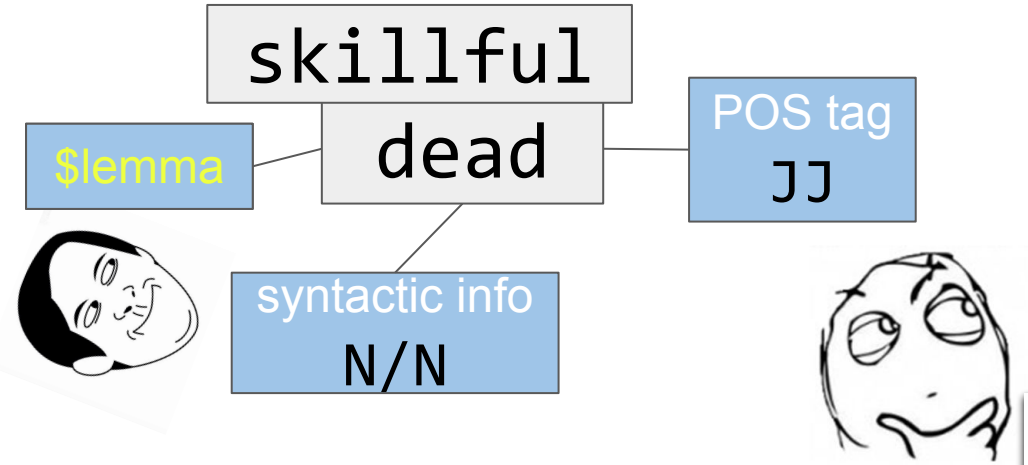


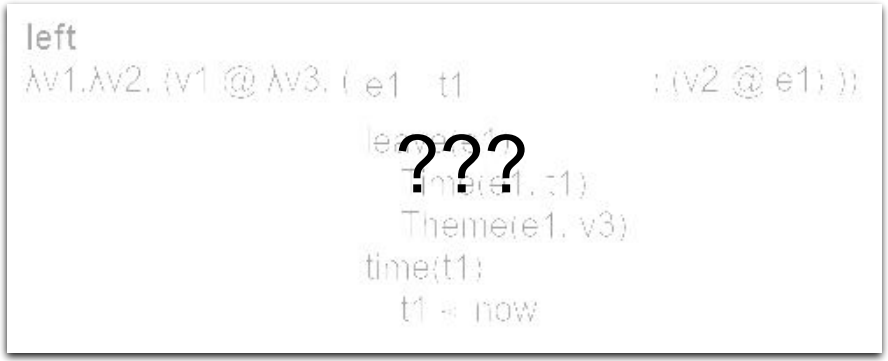
left
λv1.λv2. (v1 @ λv3. (e1 t1 : (v2 @ e1)))

???

Theme(e1, t1)
Theme(e1, v3)
time(t1)
t1 = now

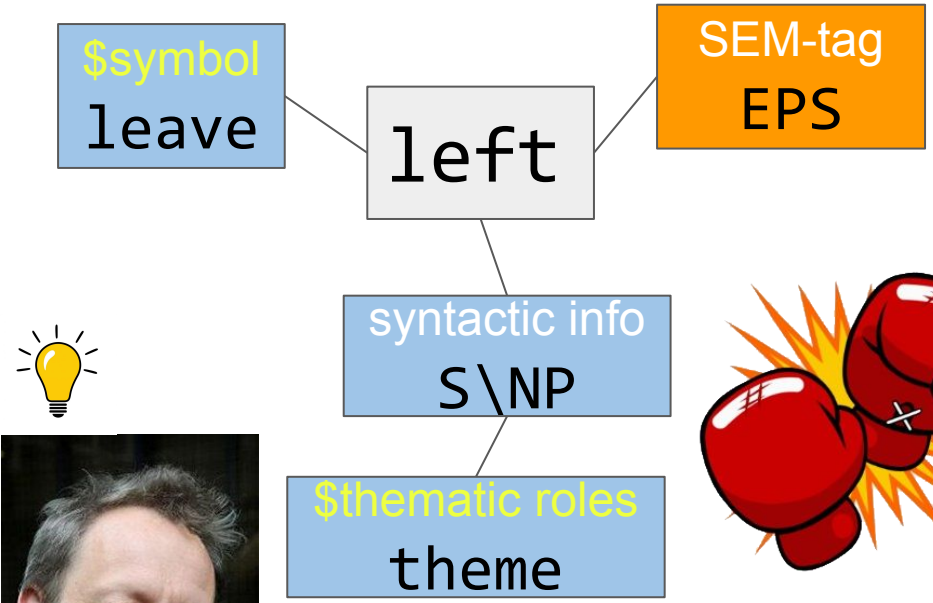
POS tags are not enough





Universal semantic tags

Universal semantic tags

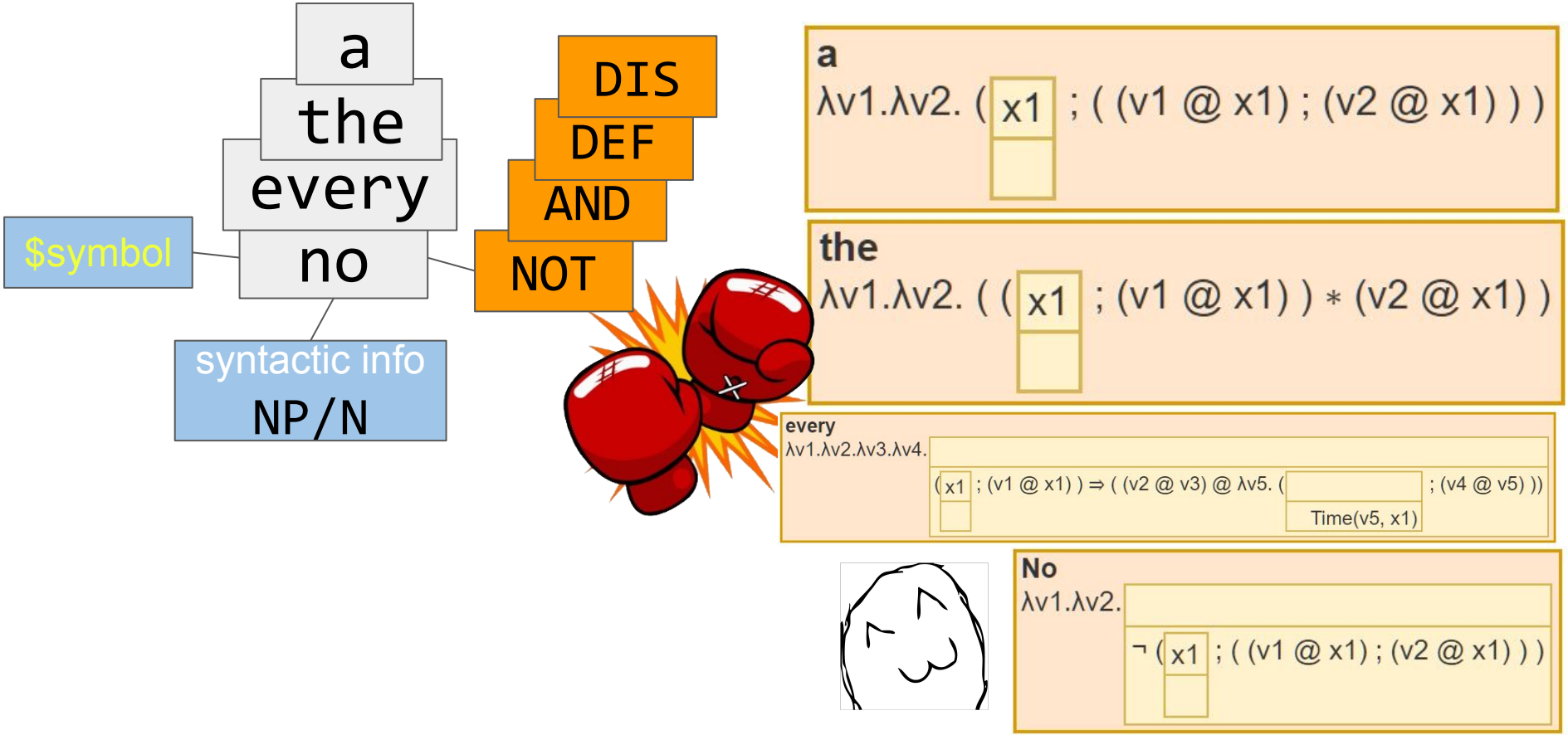


left

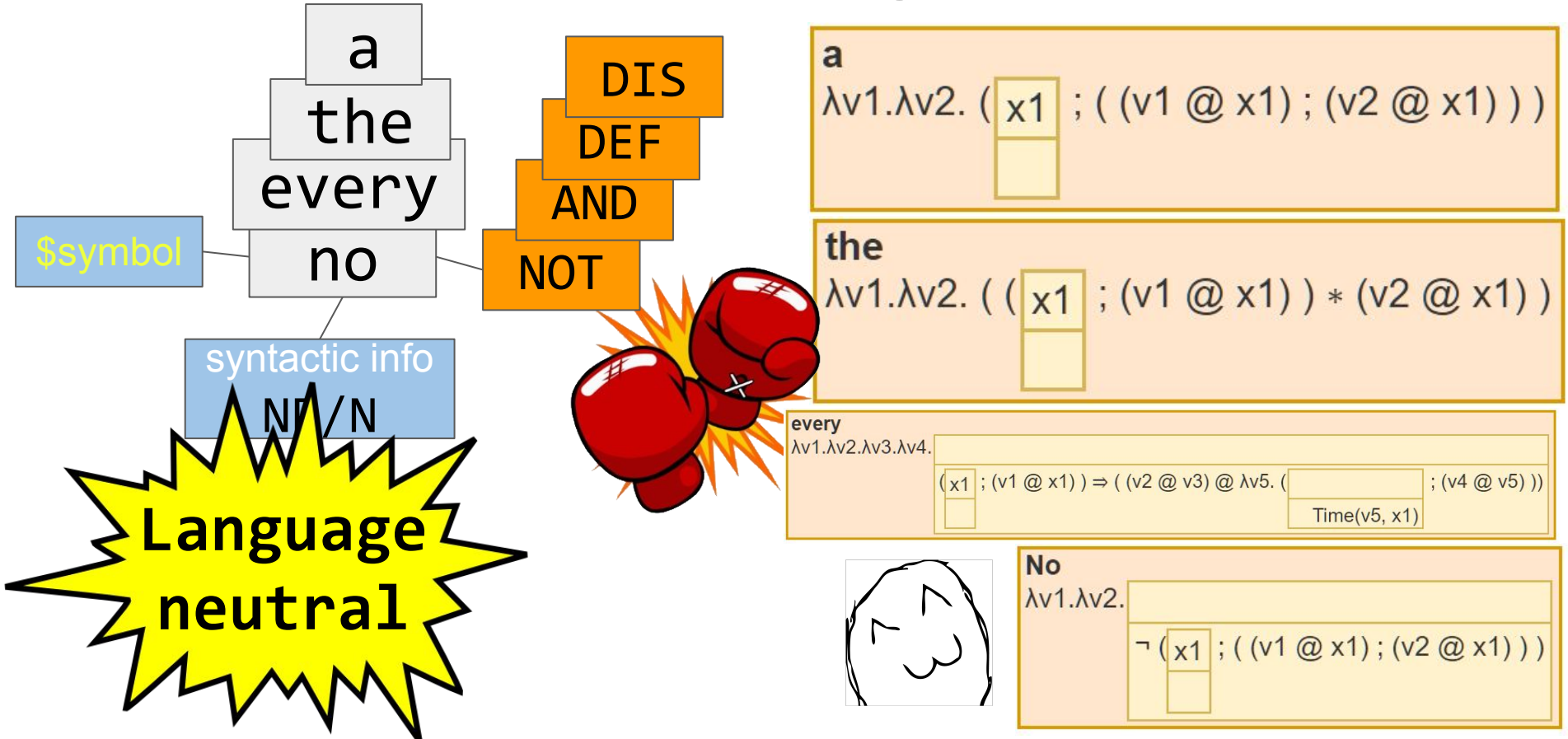
λv1.λv2. (v1 @ λv3. (e1 t1 ; (v2 @ e1)))

leave(e1)
Time(e1, t1)
Theme(e1, v3)
time(t1)
t1 ≺ now

Universal semantic tags



Universal semantic tags



SEM tags for semantics

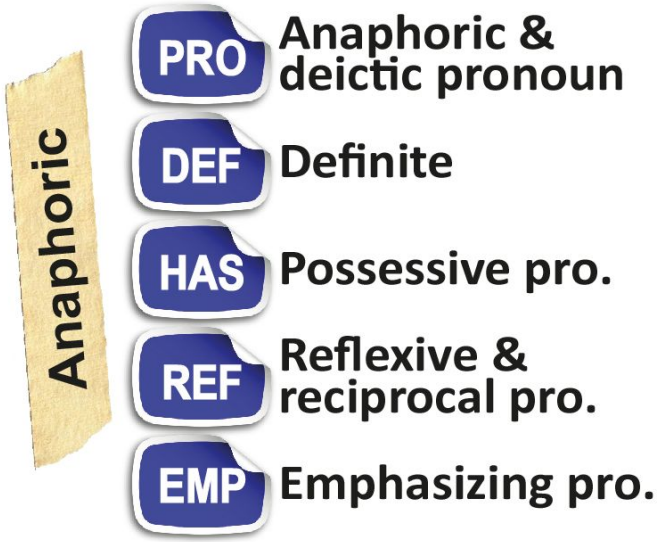
- Schema of lexical semantics is determined by a sem-tag and a syntactic category⟨SEM, CAT⟩
- Less sensitive to syntax (vs POS tags)
- Semantic info complements thematic roles, syntax and *lemma*.
- Generalizes over POS tags and Named Entity classes

UNIvesral SEemantic Tagset

- 73 sem-tags divided into 13 classes
- Under development (v0.7)
- Designed in a data-driven fashion (EN, NL, IT, DE)

Attribute

QUC	Concrete quantity	two, six million, twice, 5
QUV	Vague quantity	millions, many, enough
COL	Colour	red, crimson, light_blue
IST	Intersective	open, vegetarian, quickly
SST	Subsective	skillful surgeon, tall kid
PRI	Privative	former, fake
DEG	Degree	2 meters tall, 20 years old
INT	Intensifier	very, much, too, rather
REL	Relation	in, on, of, after
SCO	Score	3-0, grade A



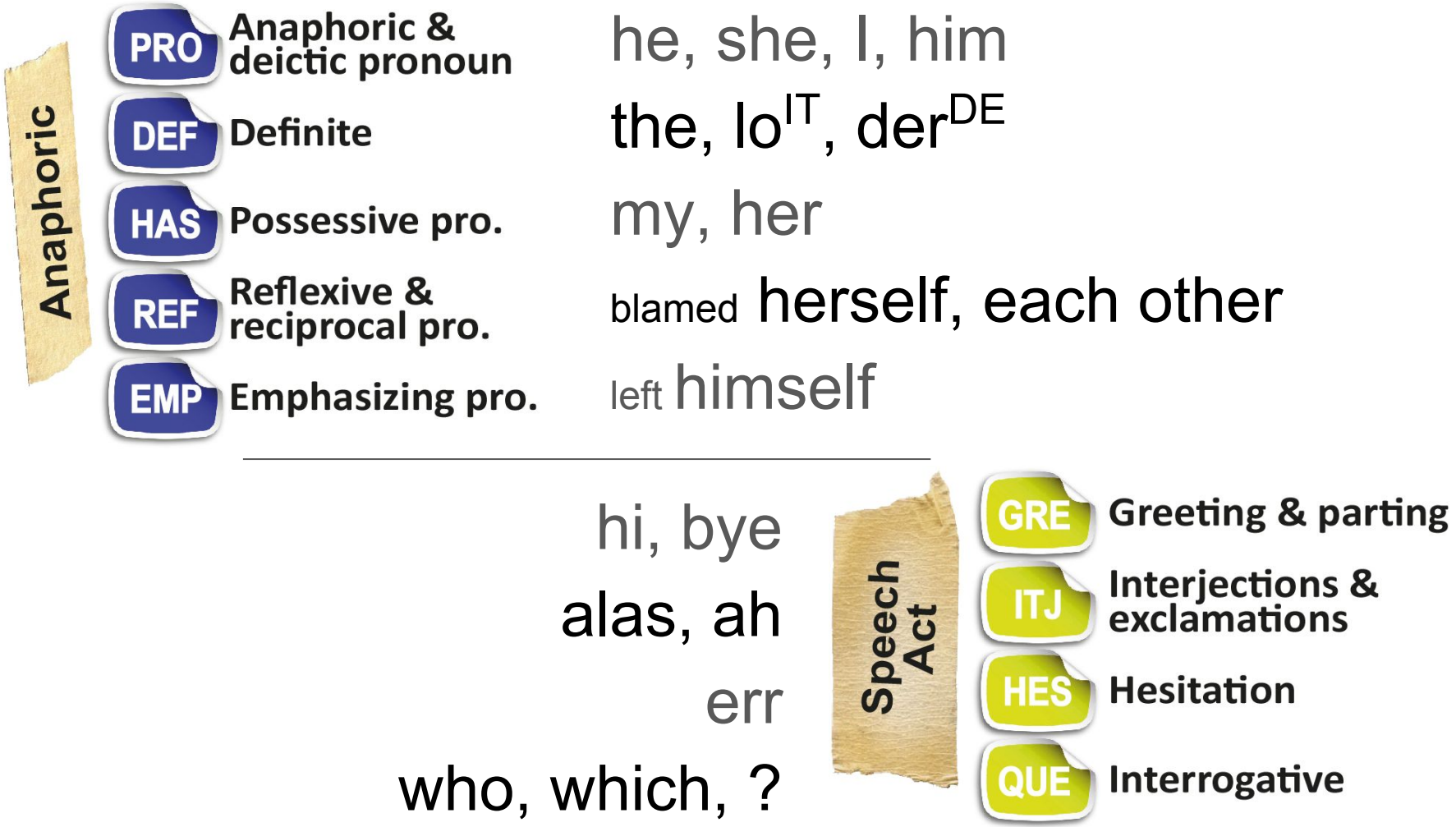
he, she, I, him

the, lo[IT], der[DE]

my, her

blamed herself, each other

left himself


Anaphoric
PRO
Anaphoric & deictic pronoun
he, she, I, him
DEF
Definite
the, lo^IT, der^DE
HAS
Possessive pro.
my, her
REF
Reflexive & reciprocal pro.
blamed herself, each other
EMP
Emphasizing pro.
left himself
hi, bye
alas, ah
err
who, which, ?
Speech Act
GRE
Greeting & parting
ITJ
Interjections & exclamations
HES
Hesitation
QUE
Interrogative

NOW **Present tense**

PST **Past tense**

FUT **Future tense**

PRG **Progressive**

PFT **Perfect**

is skiing, do ski, has skied, now

was baked, had gone, did go

will, shall

has been being treated, aan_het[NL]

has been going/done

NOW **Present tense** — is skiing, do ski, has skied, now

PST **Past tense** — was baked, had gone, did go

FUT **Future tense** — will, shall

PRG **Progressive** — has been being treated, aan_het[NL]

PFT **Perfect** — has been going/done

---

to walk, is eaten, destruction

we walk, he walks

ate, went

is running

has eaten

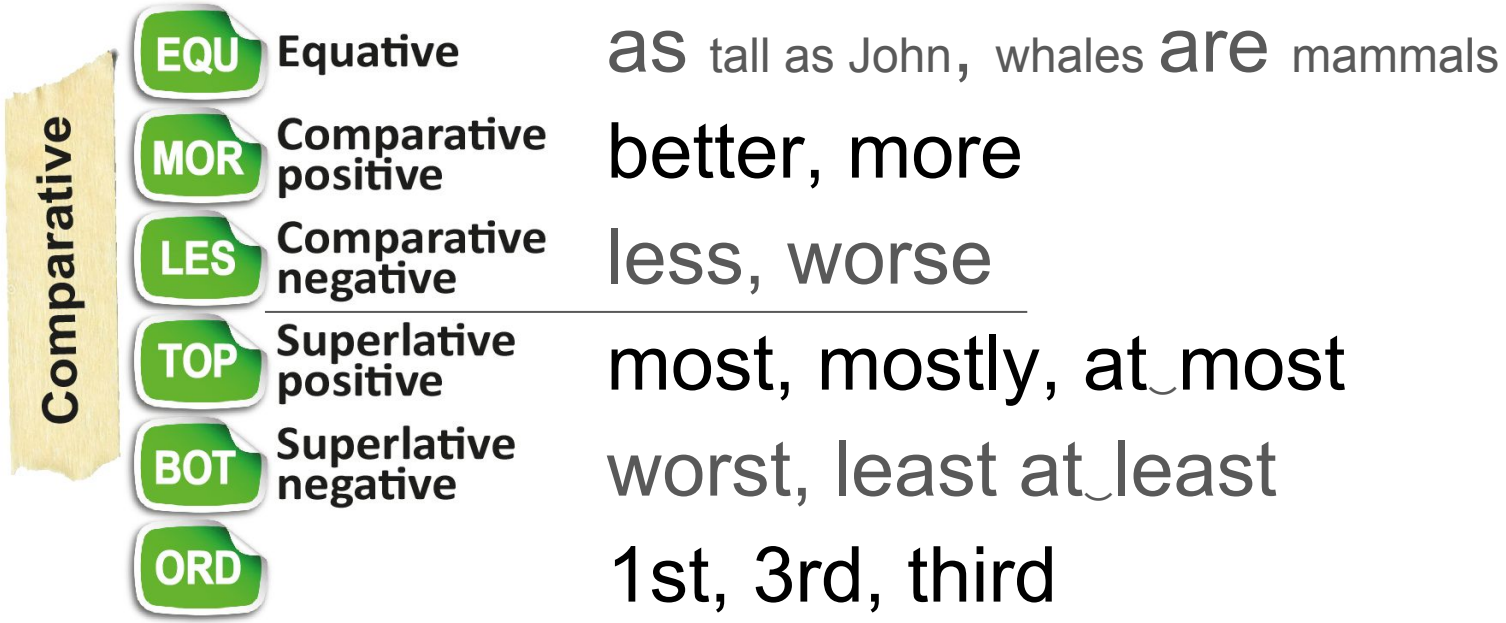
EXS **Untensed simple**

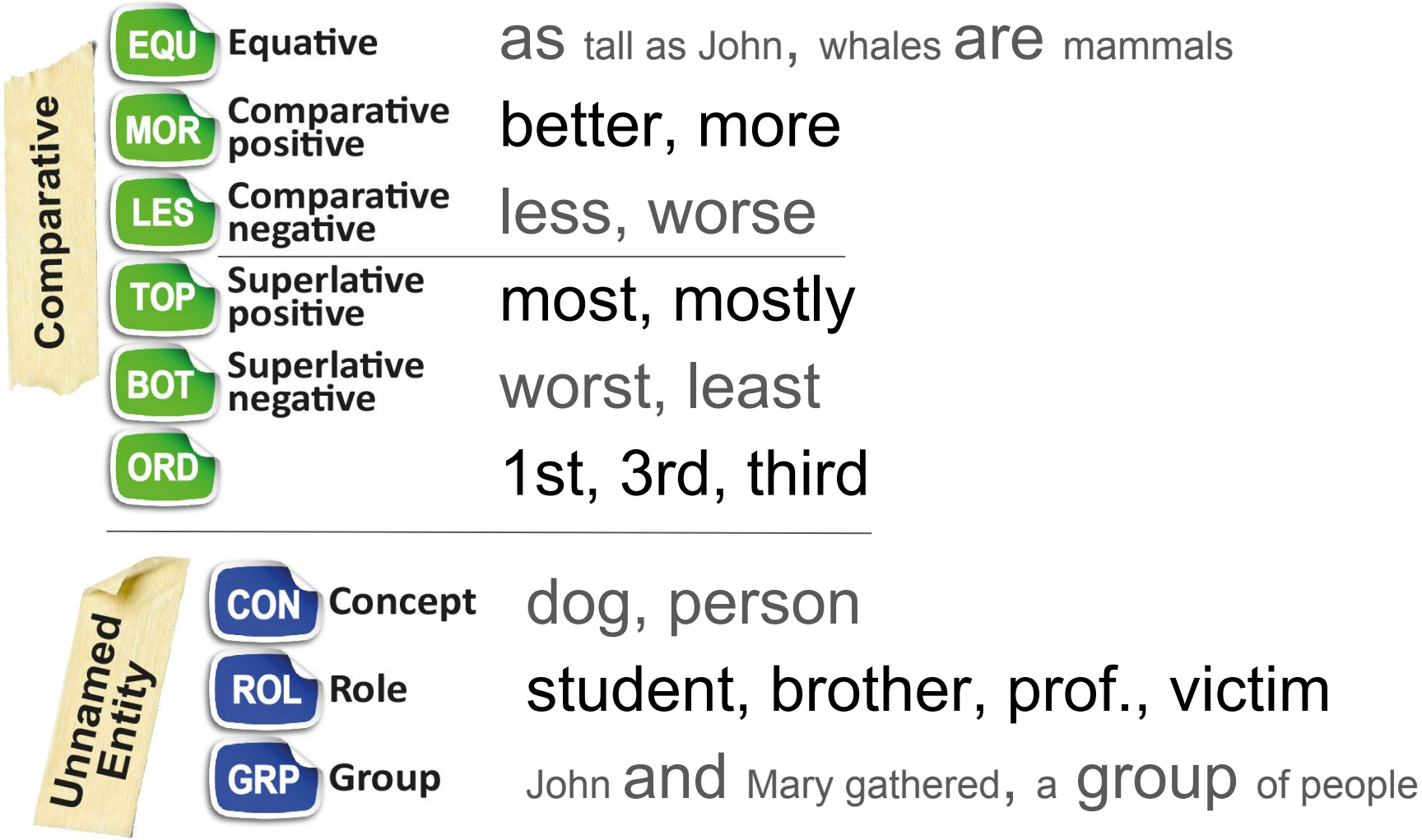
ENS **Present simple**

EPS **Past simple**

EXG **Untensed progressive**

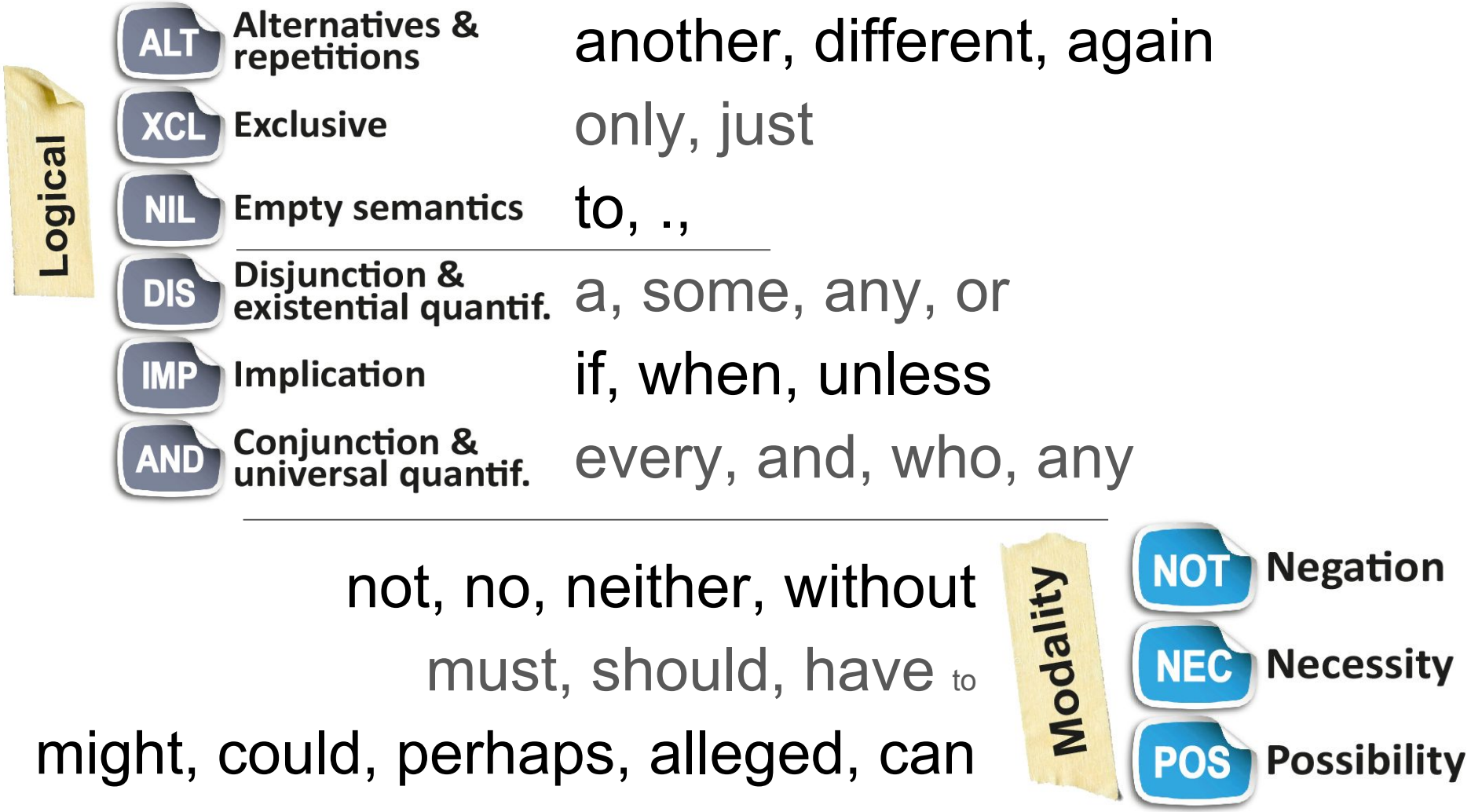
EXT **Untensed perfect**


Comparative
EQU
Equative
as tall as John, whales are mammals
MOR
Comparative positive
better, more
LES
Comparative negative
less, worse
TOP
Superlative positive
most, mostly, at_most
BOT
Superlative negative
worst, least at_least
ORD
1st, 3rd, third

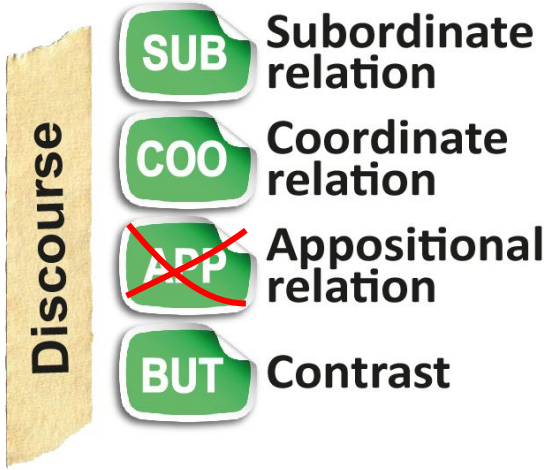

Comparative
EQU
Equative
as tall as John, whales are mammals
MOR
Comparative positive
better, more
LES
Comparative negative
less, worse
TOP
Superlative positive
most, mostly
BOT
Superlative negative
worst, least
ORD
1st, 3rd, third
Unnamed Entity
CON
Concept
dog, person
ROL
Role
student, brother, prof., victim
GRP
Group
John and Mary gathered, a group of people

Logical

ALT	Alternatives & repetitions	another, different, again
XCL	Exclusive	only, just
NIL	Empty semantics	to, .,
DIS	Disjunction & existential quantif.	a, some, any, or
IMP	Implication	if, when, unless
AND	Conjunction & universal quantif.	every, and, who, any


Logical
ALT
Alternatives & repetitions
another, different, again
XCL
Exclusive
only, just
NIL
Empty semantics
to, .,
DIS
Disjunction & existential quantif.
a, some, any, or
IMP
Implication
if, when, unless
AND
Conjunction & universal quantif.
every, and, who, any
not, no, neither, without
must, should, have to
might, could, perhaps, alleged, can
Modality
NOT
Negation
NEC
Necessity
POS
Possibility

DAT	Full date	27.04.2017, 27/04/17
DOM	Day of Month	27th December
YOC	Year of century	2017
DOW	Day of week	Thursday
MOY	Month of year	April
DEC	Decade	80s, 1990s
CLO	Clocktime	8:45‿pm, 10‿o'clock, noon

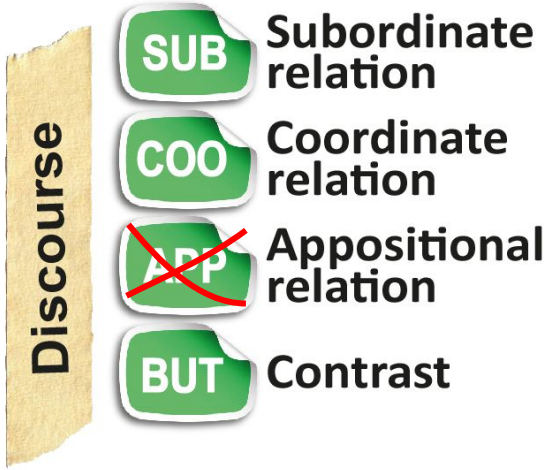


that, while, because

so, ;, and

which, —

but, yet



that, while, because

so, ;, and

which, —

but, yet

---

here, this, above

just, later, tomorrow

latter, former, above

Named Entity

PER	Person	Axl Rose, Sherlock Holmes
GPE	Geo-political entity	Paris, Japan
GPO	Geo-political origin	German, French
LOC	Geographical location	Alps, Nile
ORG	Organization	IKEA, EU
ART	Artifact	iOS_7
HAP	Happening	Eurovision_2017
UOM	Unit of measurement	meter, $, %, degree Celsius
CTC	Contact info	112, info@mail.com
~~URL~~	URL	http://pmb.let.rug.nl
LIT	Literal use of names	his name is John
NTH	Other names	table 1a, equation (1)

Tagging & Semantics

Formal compositional semantics are less favoured:

- Semantics problems
- Difficult to scale up

Make a study of formal semantics modular

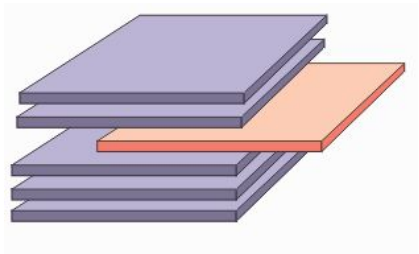
NLP community loves tagging/labeling tasks

- Conceptually a simple task
- Create an annotated data
- Employ ML techniques for learning

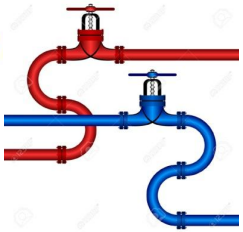
Language-neutral annotation tools

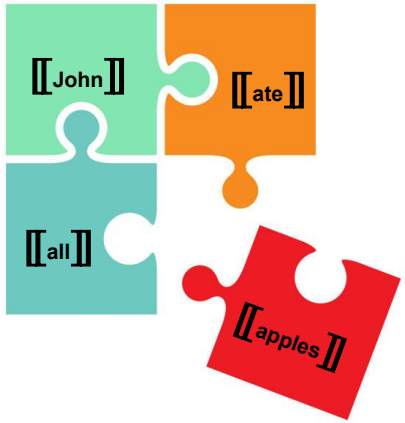
- Segmentation: elephant (Evang et al., 2013)
- Symbolization: **baseline** (conditioned unigrams)
- Word sense disambiguation (Wordnet 3.0): no
- Semantic role labeling (Verbnet roles): no
- Syntactic parsing: **EasyCCG** (Lewis & Steedman, 2014)
- Semantic parsing with DRT: **Boxer** (Bos, 2008; 2015)
- Semantic tagging: **TnT tagger** (Brants, 2000) with 93% accuracy for EN

Plan

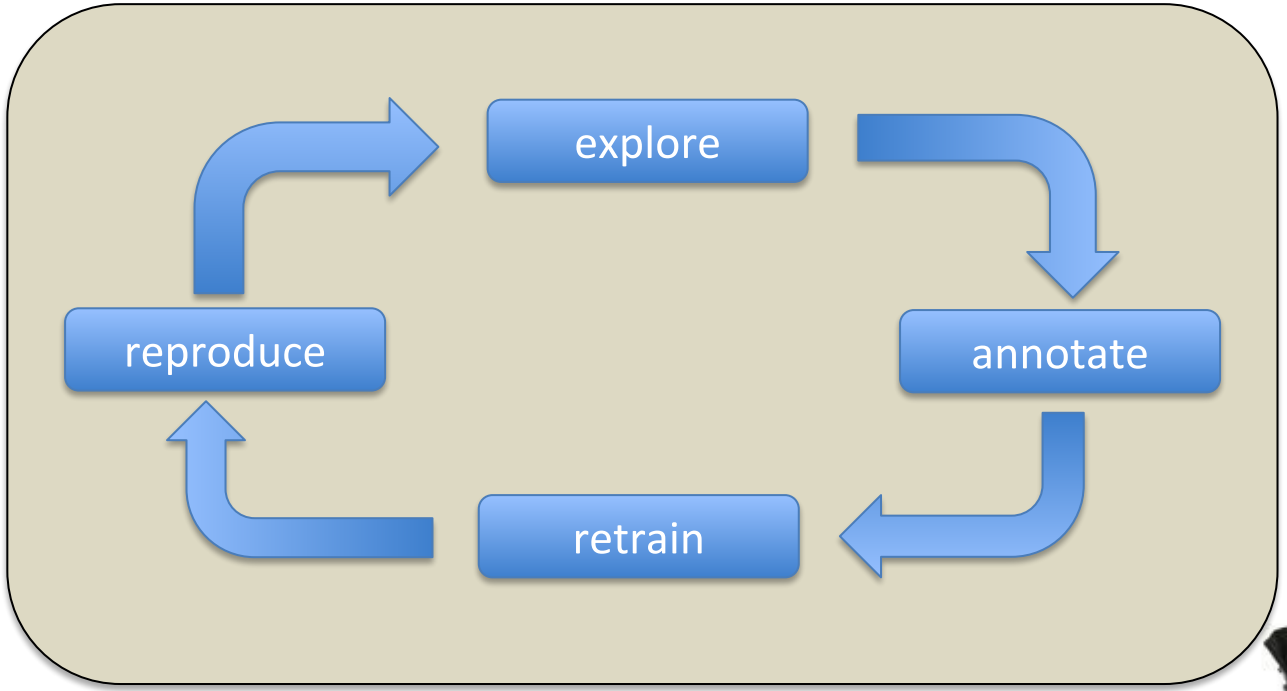


Universal
Semantic Tagging

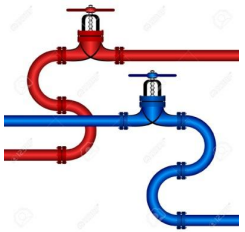


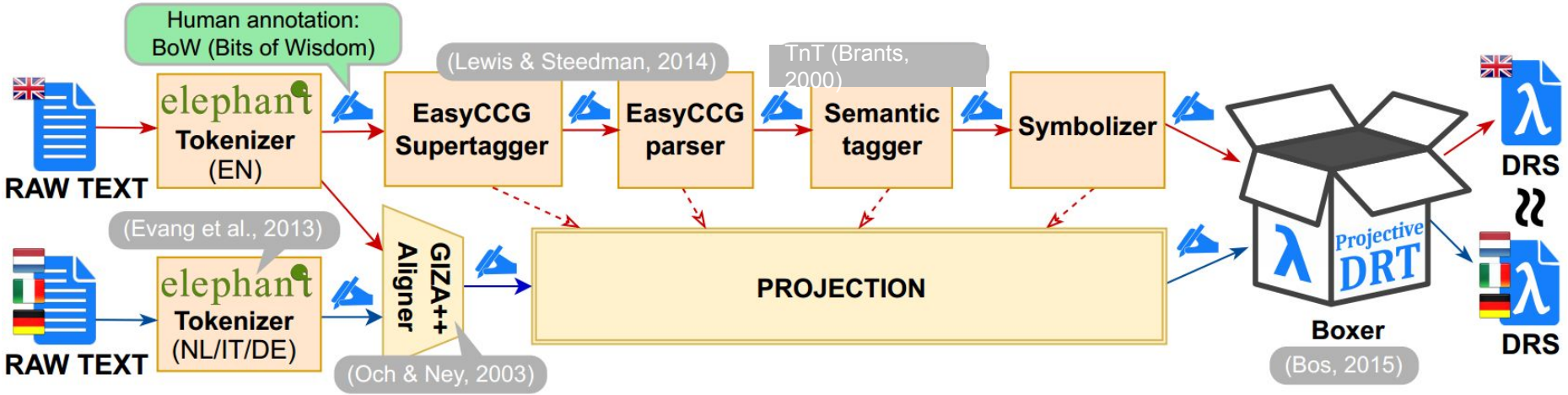


Semantic Annotation: the REAR cycle

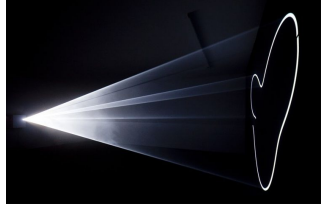


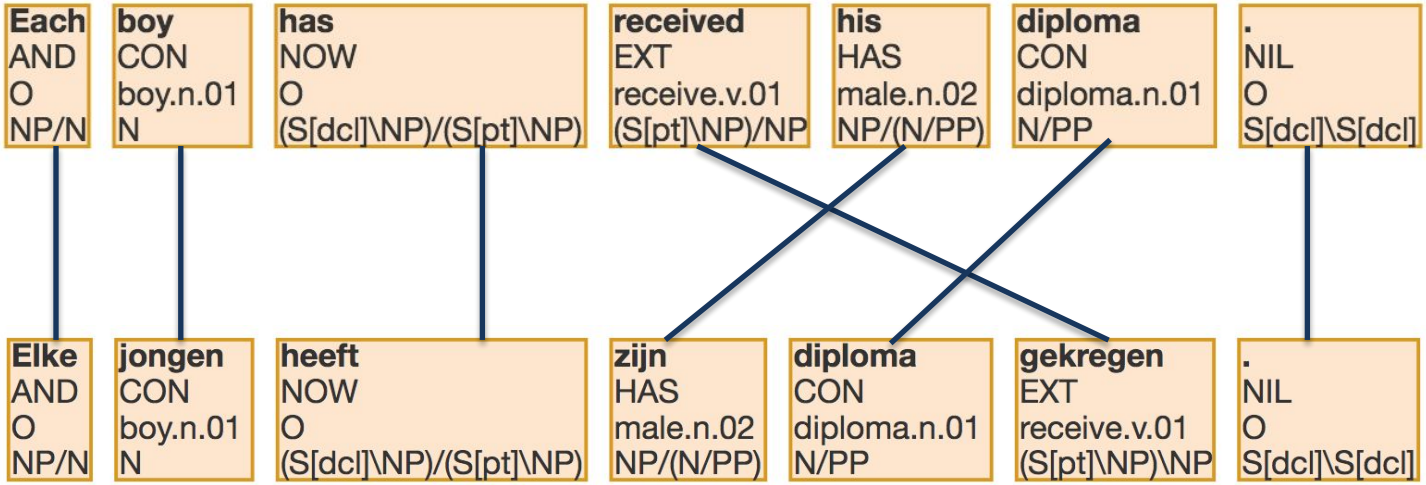
The PMB pipeline





Projection: EN → NL (PMB 19/0830)

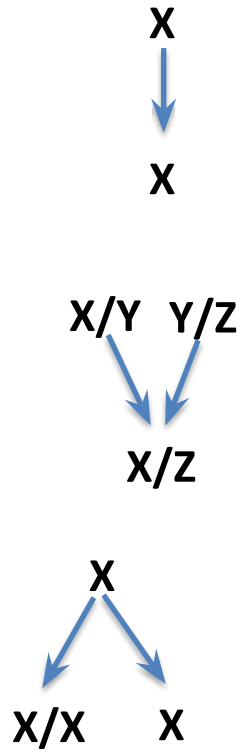




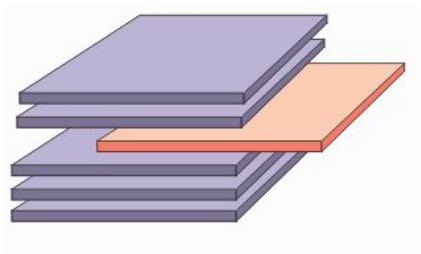
see: **Evang & Bos (COLING 2016)**

Copy, Merge, Split

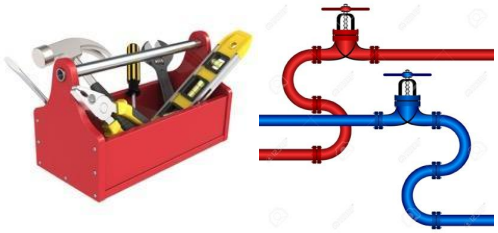
- **Copy**:
  transfer of category from source to target
- **Merge**:
  two source categories merge into
  one target category (composition)
- **Split**:
  one source category into two target
  categories (de-composition)

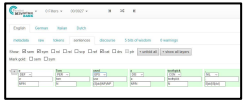


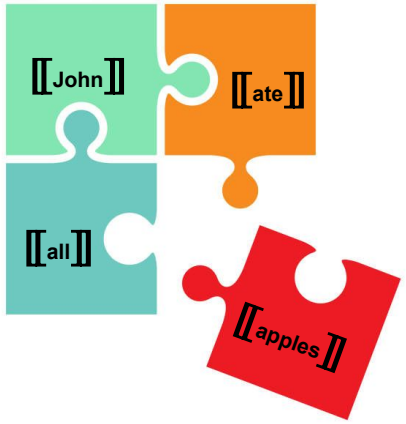
Plan



Universal
Semantic Tagging



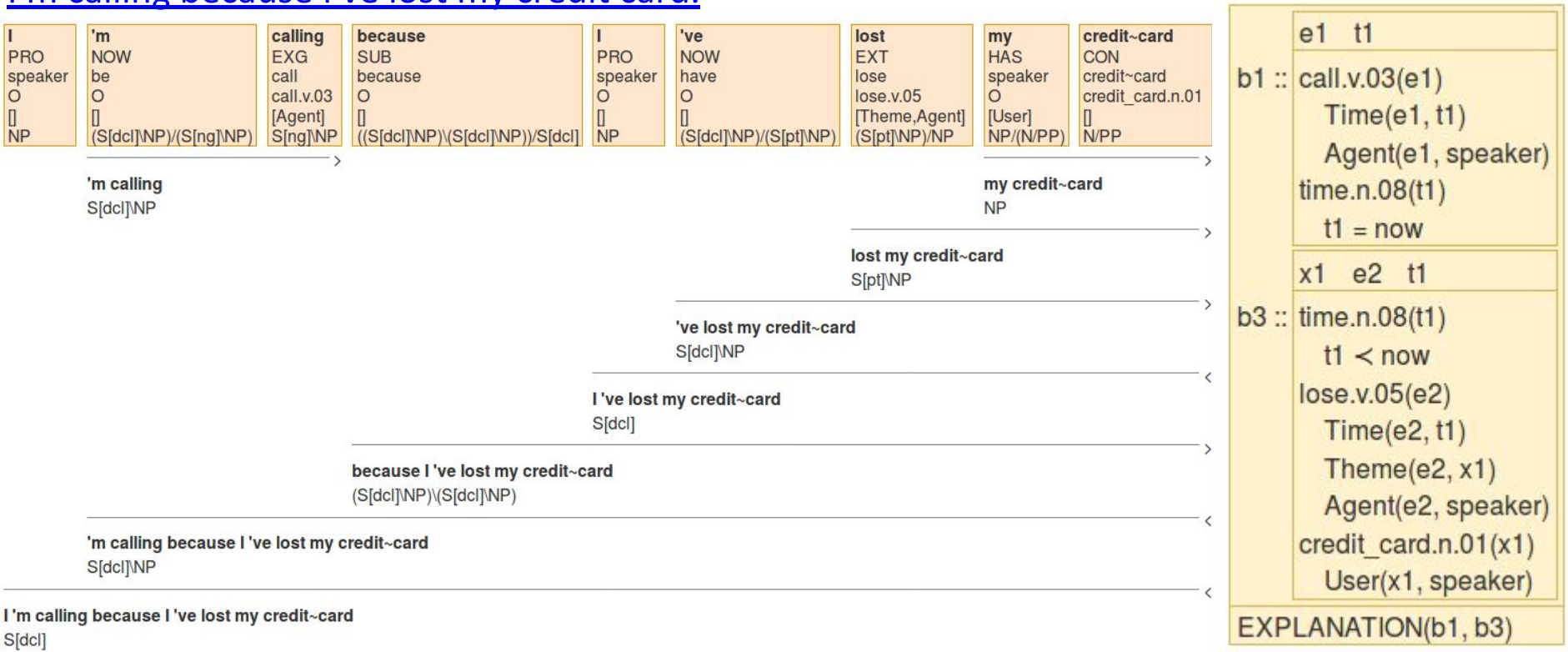




Compositional semantics

Discourse relations

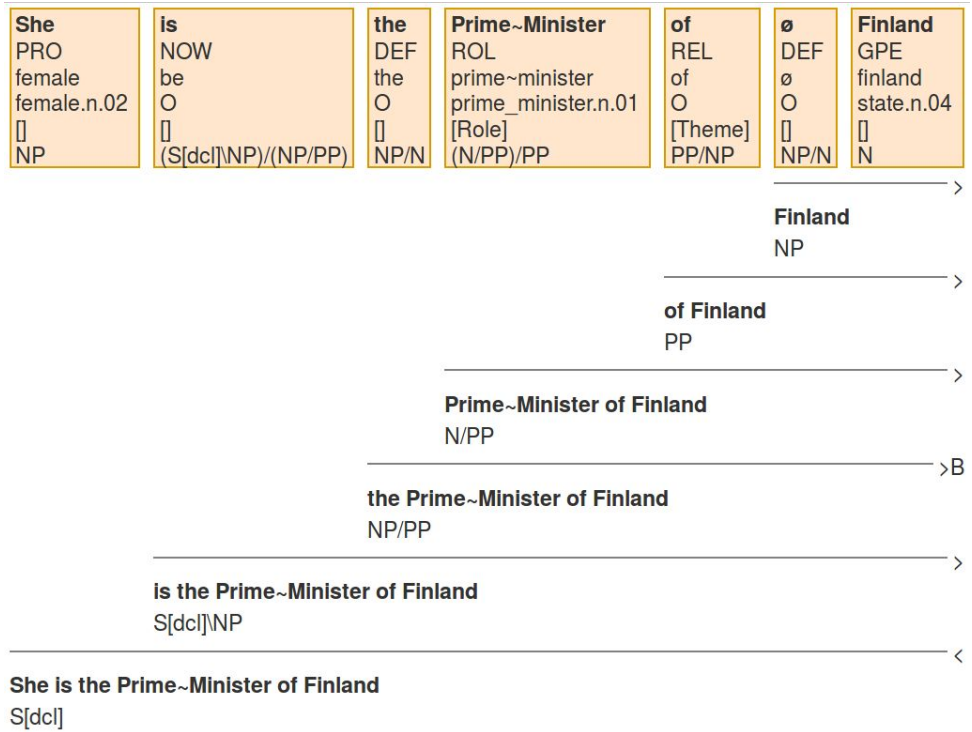
I'm calling because I've lost my credit card.



Compositional semantics

Agent nouns

[She is the Prime Minister of Finland.]

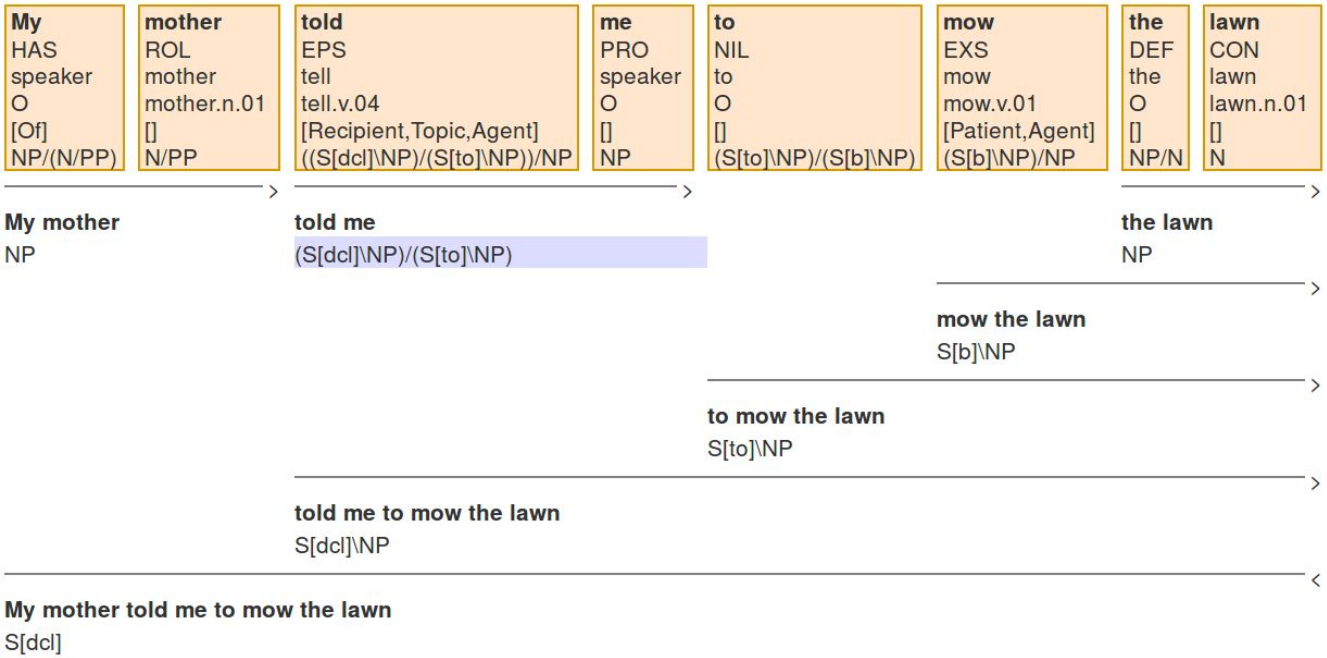



```
x1  x2  x3  x4  t1
female.n.02(x1)
   x1 = x2
time.n.08(t1)
   t1 = now
prime_minister.n.01(x3)
   Time(x3, t1)
   Theme(x3, x4)
person.n.01(x2)
   Role(x2, x3)
state.n.04(x4)
   Name(x4, finland)
```

Compositional semantics

Embedded clauses and relational nouns

My mother told me to mow the lawn.



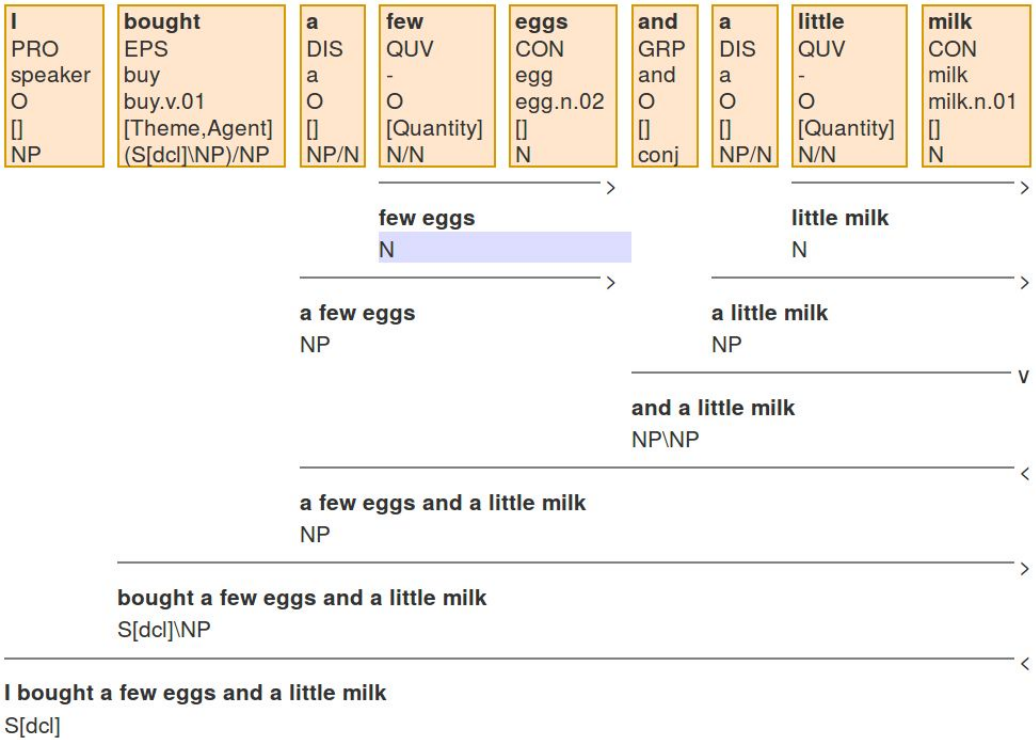
x1 x2 e1 p1 t1

mother.n.01(x2)
Of(x2, speaker)
person.n.01(x1)
Role(x1, x2)
p1: x3 e2
mow.v.01(e2)
Patient(e2, x3)
Agent(e2, speaker)
lawn.n.01(x3)
tell.v.04(e1)
Time(e1, t1)
Topic(e1, p1)
Agent(e1, x1)
Recipient(e1, speaker)
time.n.08(t1)
t1 < now

Compositional semantics

Coordination and vague quantities

I bought a few eggs and a little milk.



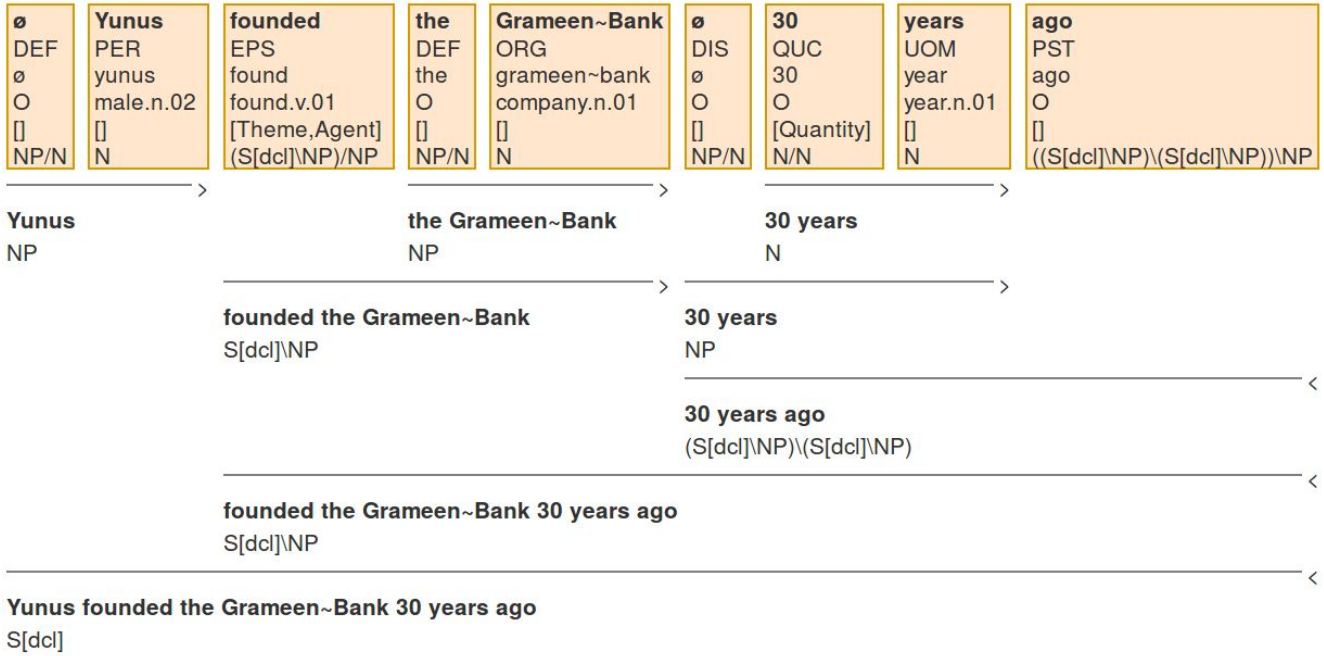
x1 x2 x3 e1 t1

buy.v.01(e1)
  Time(e1, t1)
  Theme(e1, x1)
  Agent(e1, speaker)
time.n.08(t1)
  t1 < now
egg.n.02(x2)
  Quantity(x2, -)
group.n.01(x1)
  Part(x1, x3)
  Part(x1, x2)
milk.n.01(x3)
  Quantity(x3, -)

Compositional semantics

Temporal reference

Yunus founded the Grameen Bank 30 years ago.



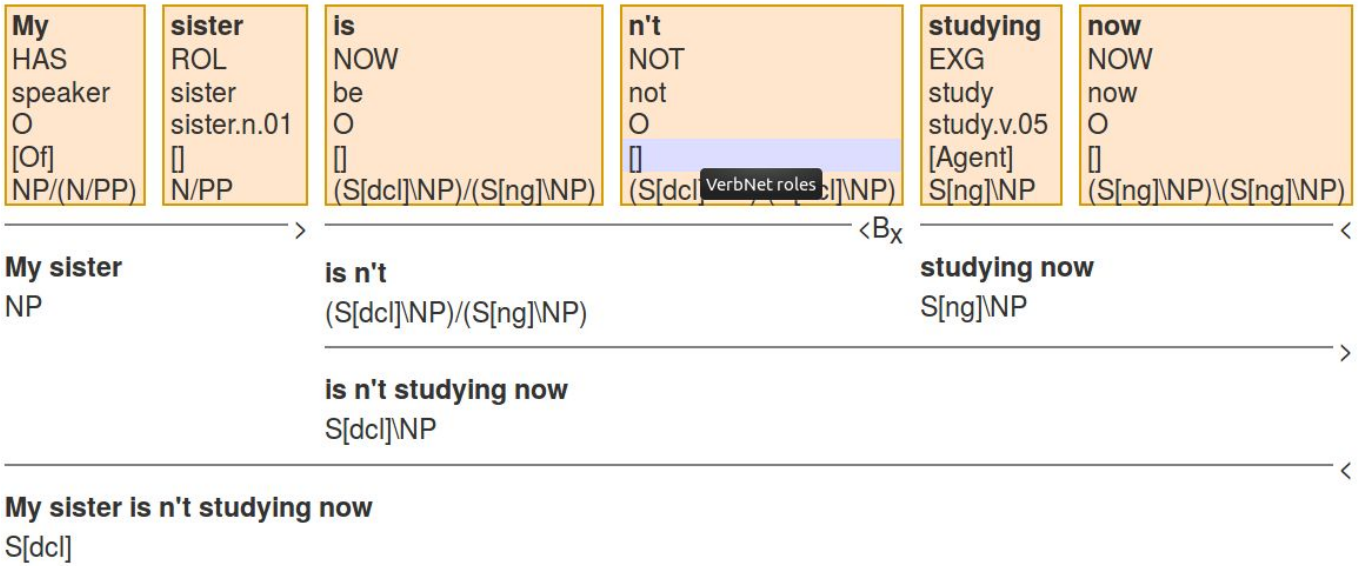
x1	x2	e1	t1	t2

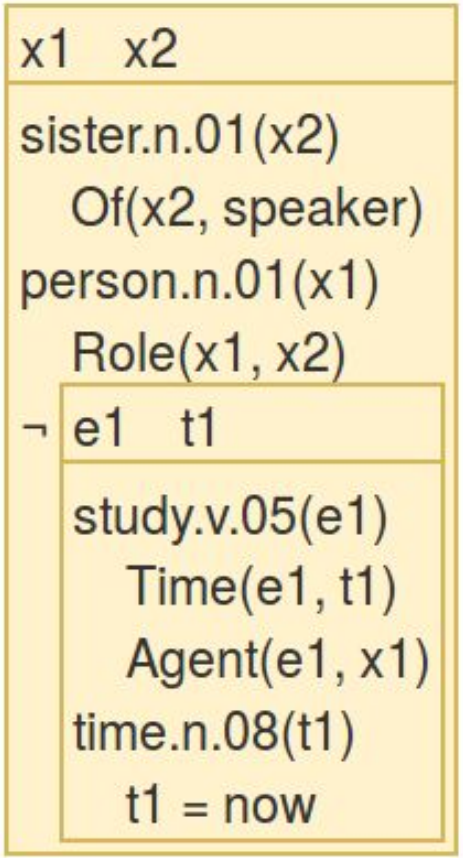
male.n.02(x1)
  Name(x1, yunus)
found.v.01(e1)
  Time(e1, t1)
  Theme(e1, x2)
  Agent(e1, x1)
time.n.08(t1)
  t1 ⋈ t2
  t1 ≺ now
company.n.01(x2)
  Name(x2, grameen~bank)
measure.n.02(t2)
  t2 ⋈ now
  Unit(t2, year)
  Quantity(t2, 30)

Compositional semantics

Negation & tense

My sister isn't studying now.

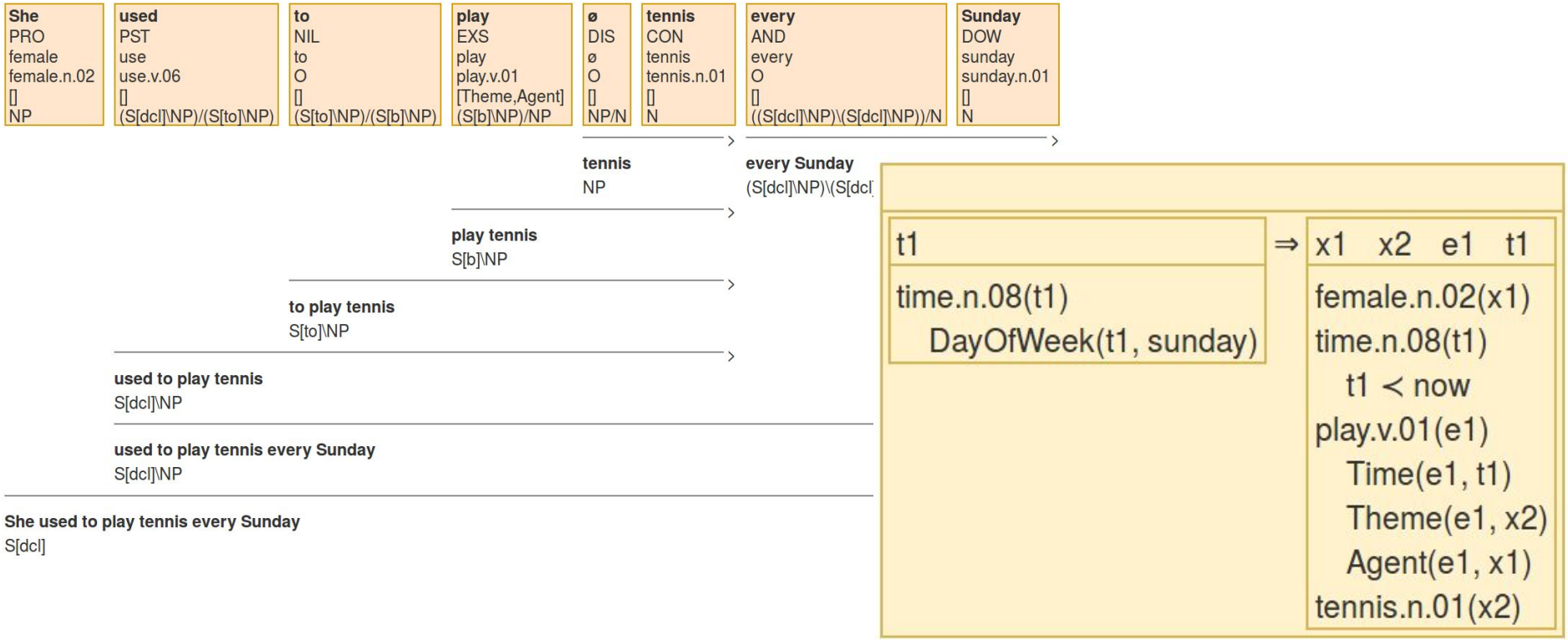




Compositional semantics

Temporal reference

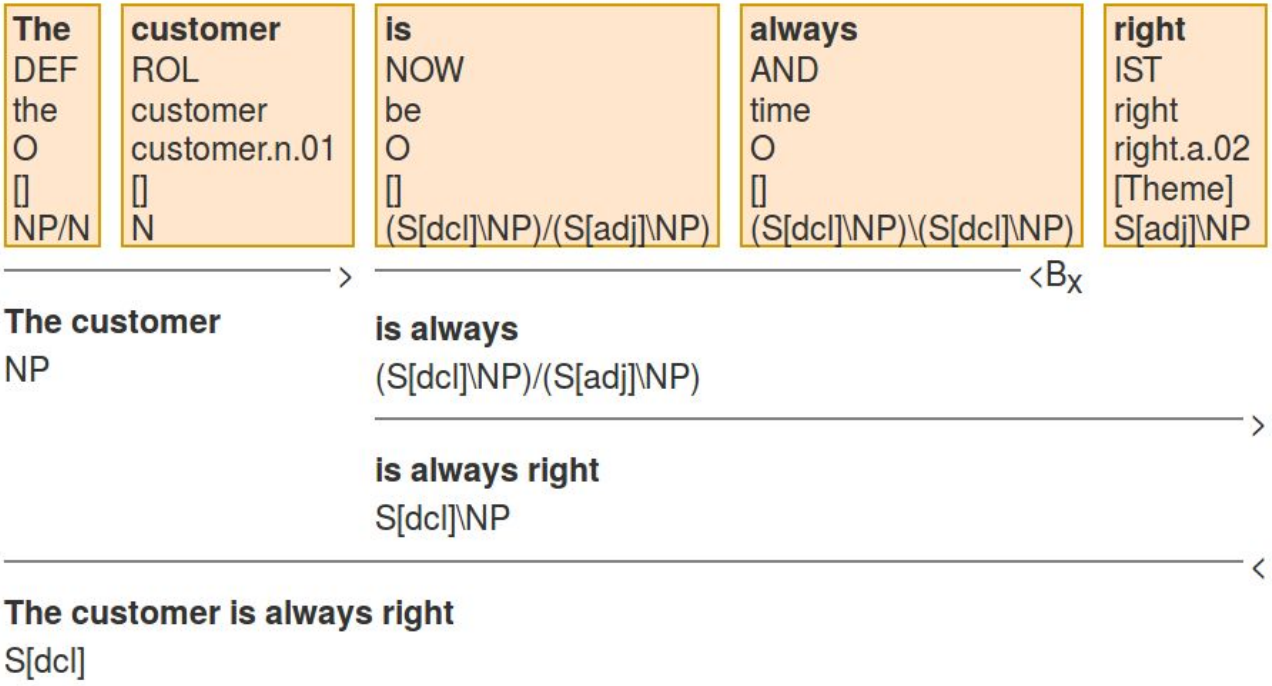
She used to play tennis every Sunday.

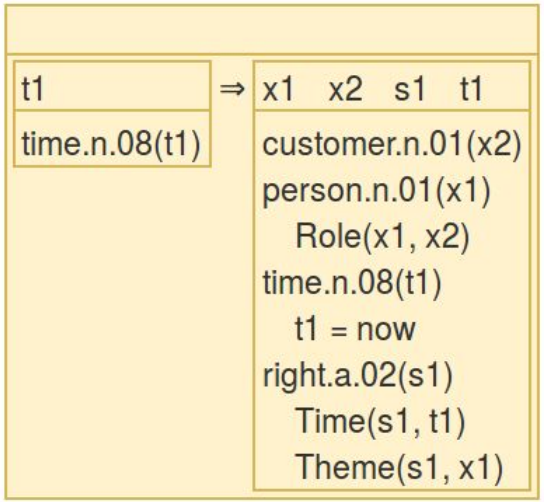


Compositional semantics

Temporal reference & generics

The customer is always right.





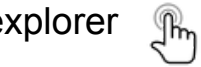
Results

- Released datasets (DRSs and Sem-tagged) http://pmb.let.rug.nl/data.php
- Partial matching of DRSs (CLFs) with Counter (van Noord et al., 2018)
- Automatically learning CLFs (van Noord et al., 2018)
- Semantic tagging for NLP tasks: UPOS, UDEP, NLI (Abdou et al., 2018)

```
x1   x2   e1   t1
male.n.02(x1)
   Name(x1, alfred~nobel)
invent.v.01(e1)
   Time(e1, t1)
   Result(e1, x2)
   Agent(e1, x1)
time.n.08(t1)
   YearOfCentury(t1, 1866)
   t1 < now
dynamite.n.01(x2)
```

```
b1 REF x1                    % Alfred~Nobel [0...12]
b1 Name x1 "alfred~nobel"    % Alfred~Nobel [0...12]
b1 male "n.02" x1            % Alfred~Nobel [0...12]
b2 REF e1                    % invented [13...21]
b2 Agent e1 x1               % invented [13...21]
b2 Result e1 x2              % invented [13...21]
b2 invent "v.01" e1          % invented [13...21]
b3 TPR t1 "now"              % invented [13...21]
b2 Time e1 t1                % invented [13...21] in [31...33]
b3 REF t1                    % invented [13...21]
b3 time "n.08" t1            % invented [13...21] 1866 [34...38]
b2 REF x2                    % dynamite [22...30]
b2 dynamite "n.01" x2        % dynamite [22...30]
b3 YearOfCentury t1 "1866"   % 1866 [34...38]
                             % . [38...39]
```

Conclusion

- Parallel corpus annotated with shared formal compositional MRs
- Online explorer for collaborative annotation http://pmb.let.rug.nl/explorer 
- Universal semantic tags
- Language neutral NLP systems

Conclusion

- Parallel corpus annotated with shared formal compositional MRs
- Online explorer for collaborative annotation http://pmb.let.rug.nl/explorer
- Universal semantic tags
- Language neutral NLP systems

Coming shared task @ IWCS-19

```
b1 REF x1                    % Alfred~Nobel [0...12]
b1 Name x1 "alfred~nobel"    % Alfred~Nobel [0...12]
b1 male "n.02" x1            % Alfred~Nobel [0...12]
b2 REF e1                    % invented [13...21]
b2 Agent e1 x1               % invented [13...21]
b2 Result e1 x2              % invented [13...21]
b2 invent "v.01" e1          % invented [13...21]
b3 TPR t1 "now"              % invented [13...21]
b2 Time e1 t1                % invented [13...21] in [31...33]
b3 REF t1                    % invented [13...21]
b3 time "n.08" t1            % invented [13...21] 1866 [34...38]
b2 REF x2                    % dynamite [22...30]
b2 dynamite "n.01" x2        % dynamite [22...30]
b3 YearOfCentury t1 "1866"   % 1866 [34...38]
                             % . [38...39]
```

```
x1   x2   e1   t1
male.n.02(x1)
   Name(x1, alfred~nobel)
invent.v.01(e1)
   Time(e1, t1)
   Result(e1, x2)
   Agent(e1, x1)
time.n.08(t1)
   YearOfCentury(t1, 1866)
   t1 < now
dynamite.n.01(x2)
```

```
Alfred Nobel invented dynamite in 1866.
```

Lost in Translation – Found in Meaning

NWO vici project: 2016-2020

Kilian Evang
(Postdoc)

Johan Bos
(Project Leader)

Talita Anthonio
(Student Assistant)

Johannes Bjerva
(PhD Student)

Lasha Abzianidze
(Postdoc)

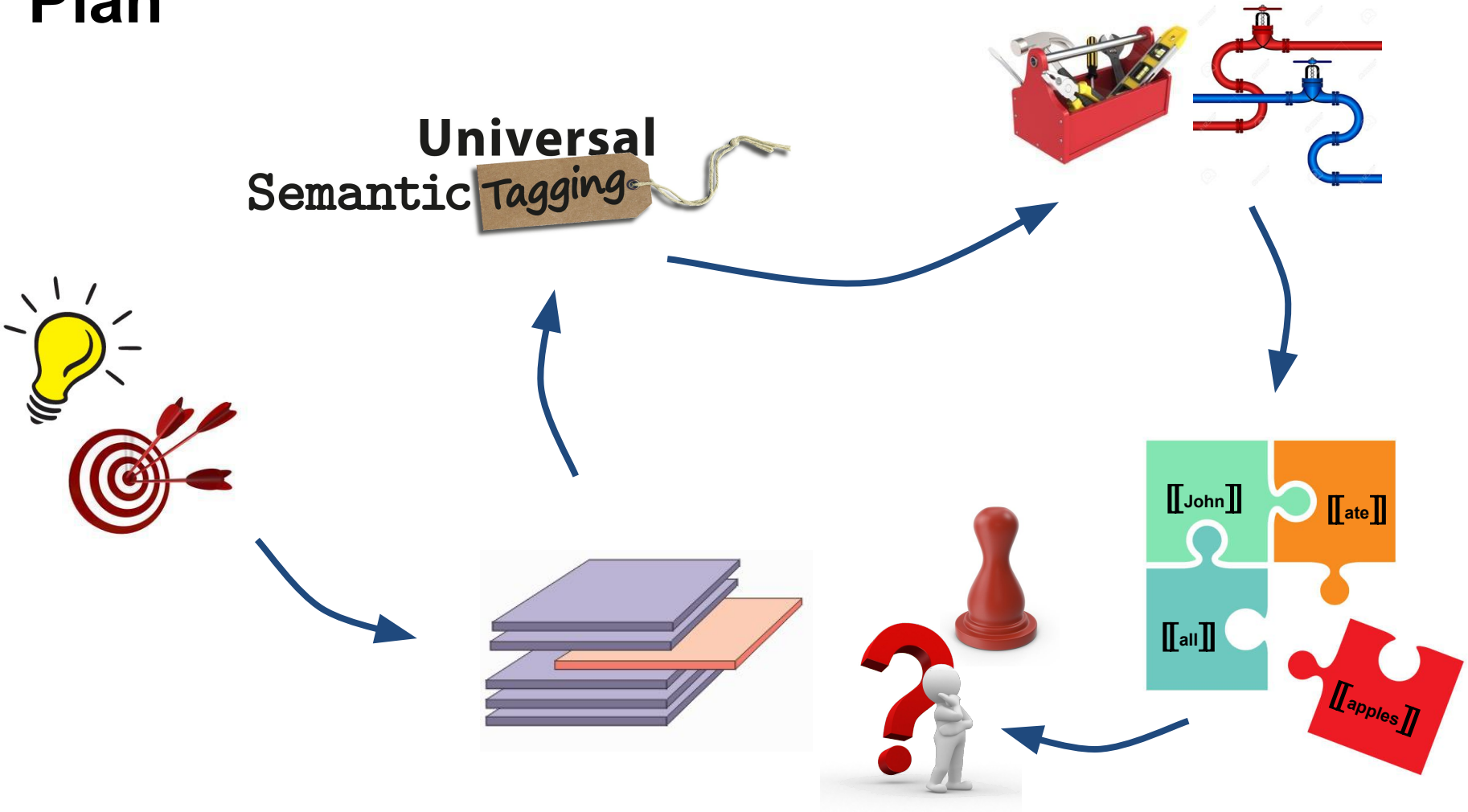
Rik van Noord
(PhD Student)

Hessel Haagsma
(PhD Student)

Jaap Nanninga
(Student Assistant)

Parallel Meaning Bank

university of groningen

NWO
Netherlands Organisation for Scientific Research

Plan