Towards Universal Semantic Tagging

Lasha Abzianidze
(joint work with Johan Bos)
29.05.2018

university of groningen

Parallel Meaning Bank

Lexical semantics

From *lexical* semantics to phrasal semantics

Semantic lexicon is usually large

Lexical semantics

From *lexical* semantics to phrasal semantics

Semantic lexicon is usually large

Which lexical semantics to assign to word tokens?

Lexical semantics

From *lexical* semantics to phrasal semantics

Semantic lexicon is usually large

Which lexical semantics to assign to word tokens?

- 07/1937 I **have** gone to the cinema
- 00/1564 I **have** a big dog
- 00/2206 I **have** to warn him

Lexical semantics

From *lexical* semantics to phrasal semantics

Semantic lexicon is usually large

Which lexical semantics to assign to word tokens?

- 07/1937 I **have** gone to the cinema
- 00/1564 I **have** a big dog
- 00/2206 I **have** to warn him

Can POS tags help?

Lexical semantics

From *lexical* semantics to phrasal semantics

Semantic lexicon is usually large

Which lexical semantics to assign to word tokens?

- 07/1937 I **have** gone to the cinema
- 00/1564 I **have** a big dog
- 00/2206 I **have** to warn him

Can POS tags help? **NO** as all the three gets VBP

More examples

- He **himself**^{PRP} tried it, Tom cut **himself**^{PRP} while shaving
- **and**^{CC}, **or**^{CC}, **but**^{CC}
- … **to**^{TO} write …, … **to**^{TO} cinema …
- does not like **any**^{DT} X. Give me **any**^{DT} X
- **a(n)**/**every**/**no**/**the**/**some**/**each**/**that**/**these**/**(n)either...**^{DT}
- **ill**^{JJ} / **skillful**^{JJ} / **fake**^{JJ} professor
- Google**,** New York; Ann**,** Bill and Mary; Ann**,** a director**,**...

Outline

- Groningen/Parallel Meaning bank
- UNIversal SEmantic Tagset
- Results & Challenges
- Conclusion

Formal compositional semantics in Parallel Meaning Bank

- Heavy lexical units: DRSs
- Few combining rules: Rules of CCG
- λ-calculus for computation: λ-DRS

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)

Formal compositional semantics in Parallel Meaning Bank

- **Heavy lexical units**: DRSs
- Few combining rules: Rules of CCG
- λ-calculus for computation: λ-DRS

left

λv1.λv2. (v1 @ λv3. (

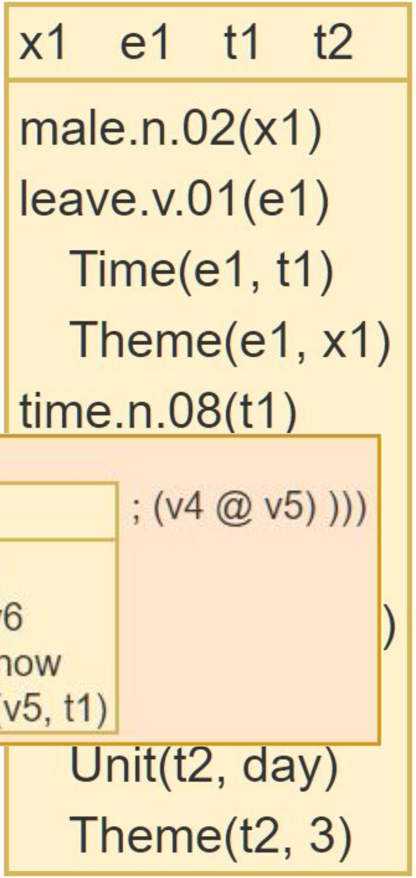
e1 t1
leave(e1) Time(e1, t1) Theme(e1, v3) time(t1) t1 ≺ now

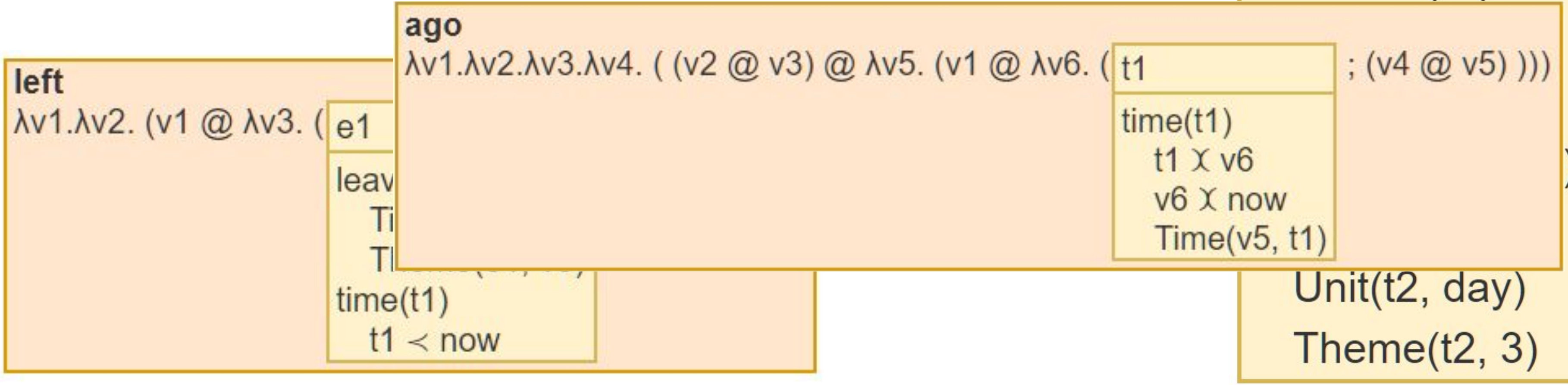
; (v2 @ e1)))

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)

Formal compositional semantics in Parallel Meaning Bank

- **Heavy lexical units**: DRSs
- Few combining rules: Rules of CCG
- λ-calculus for computation: λ-DRS





Compositionality  Projection

He | left | ∅ | three | days | ago

He
λv1. ([x1 | male(x1)] ⋆ (v1 @ x1))

left
λv1.λv2. (v1 @ λv3. ([e1 t1 | leave(e1), Time(e1, t1), Theme(e1, v3), time(t1), t1 ≺ now] ; (v2 @ e1)))

ø
λv1.λv2. ([x1 |] ; ((v1 @ x1) ; (v2 @ x1)))

three
λv1.λv2. ([| Theme(v2, 3)] ; (v1 @ v2))

days
λv1. [| measure(v1), Unit(v1, day)]

ago
λv1.λv2.λv3.λv4. ((v2 @ v3) @ λv5. (v1 @ λv6. ([t1 | time(t1), t1 ⊃⊂ v6, v6 ⊃⊂ now, Time(v5, t1)] ; (v4 @ v5))))

He | is | ∅ | drie | dagen | geleden | vertrokken

Hij
λv1. ([x1 | male(x1)] ⋆ (v1 @ x1))

is
λv1.v1

ø
λv1.λv2. ([x1 |] ; ((v1 @ x1) ; (v2 @ x1)))

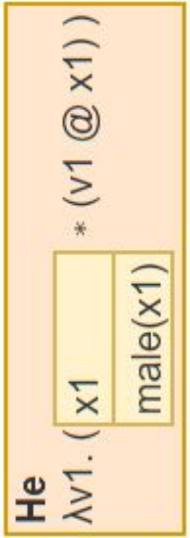
drie
λv1.λv2. ([| Theme(v2, 3)] ; (v1 @ v2))

dagen
λv1. [| measure(v1), Unit(v1, day)]

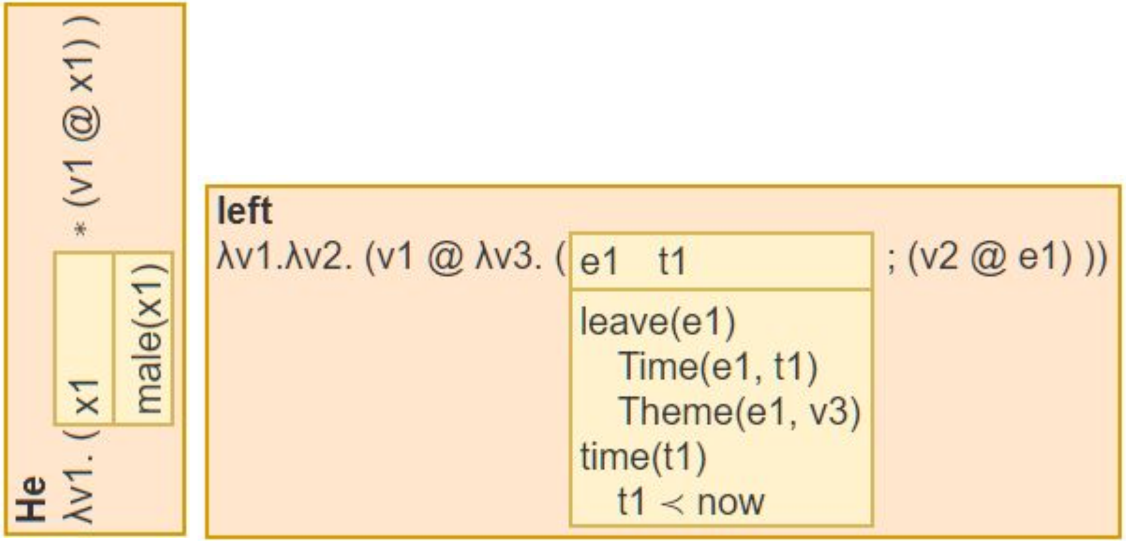
geleden
λv1.λv2.λv3.λv4. ((v2 @ v3) @ λv5. (v1 @ λv6. ([t1 | time(t1), t1 ⊃⊂ v6, v6 ⊃⊂ now, Time(v5, t1)] ; (v4 @ v5))))

vertrokken
λv1.λv2. (v1 @ λv3. ([e1 t1 | leave(e1), Time(e1, t1), Theme(e1, v3), time(t1), t1 ≺ now] ; (v2 @ e1)))

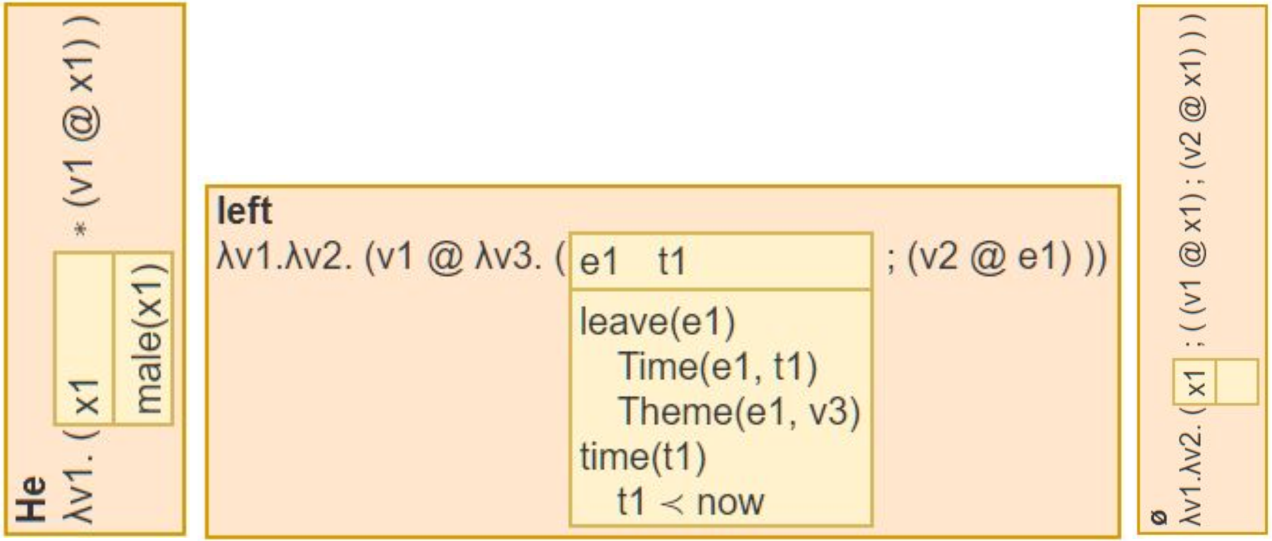
Semantic building blocks



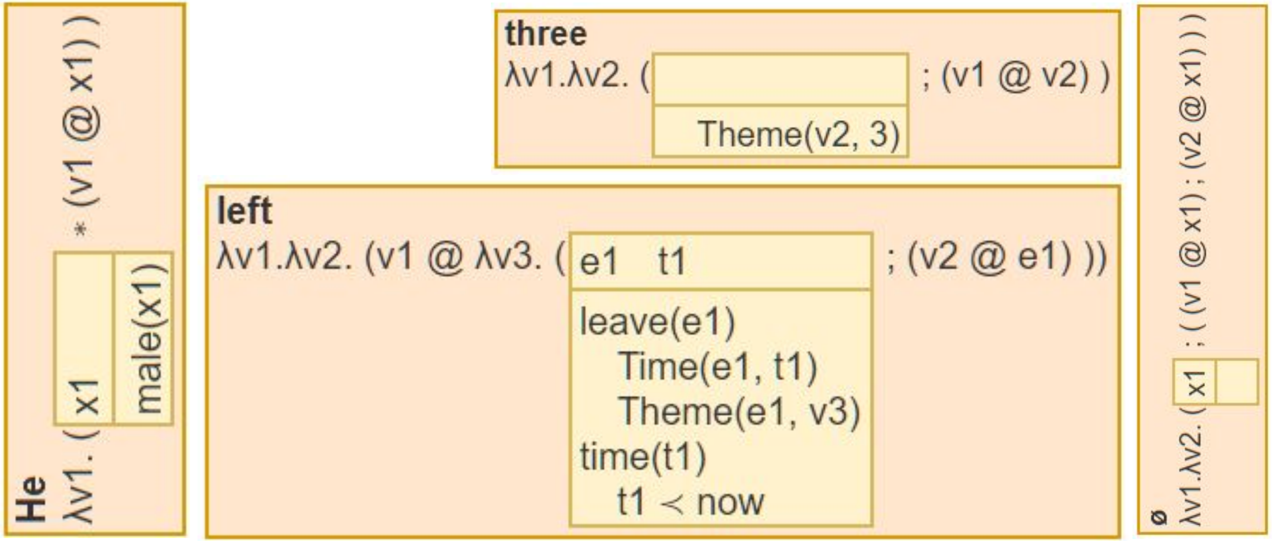
Semantic building blocks



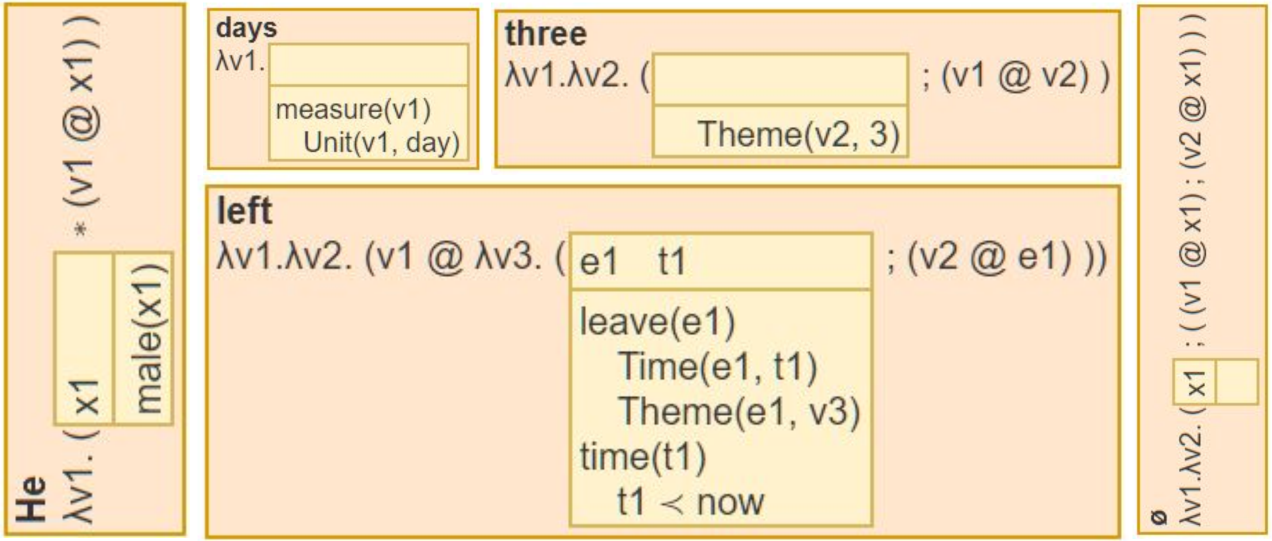
Semantic building blocks



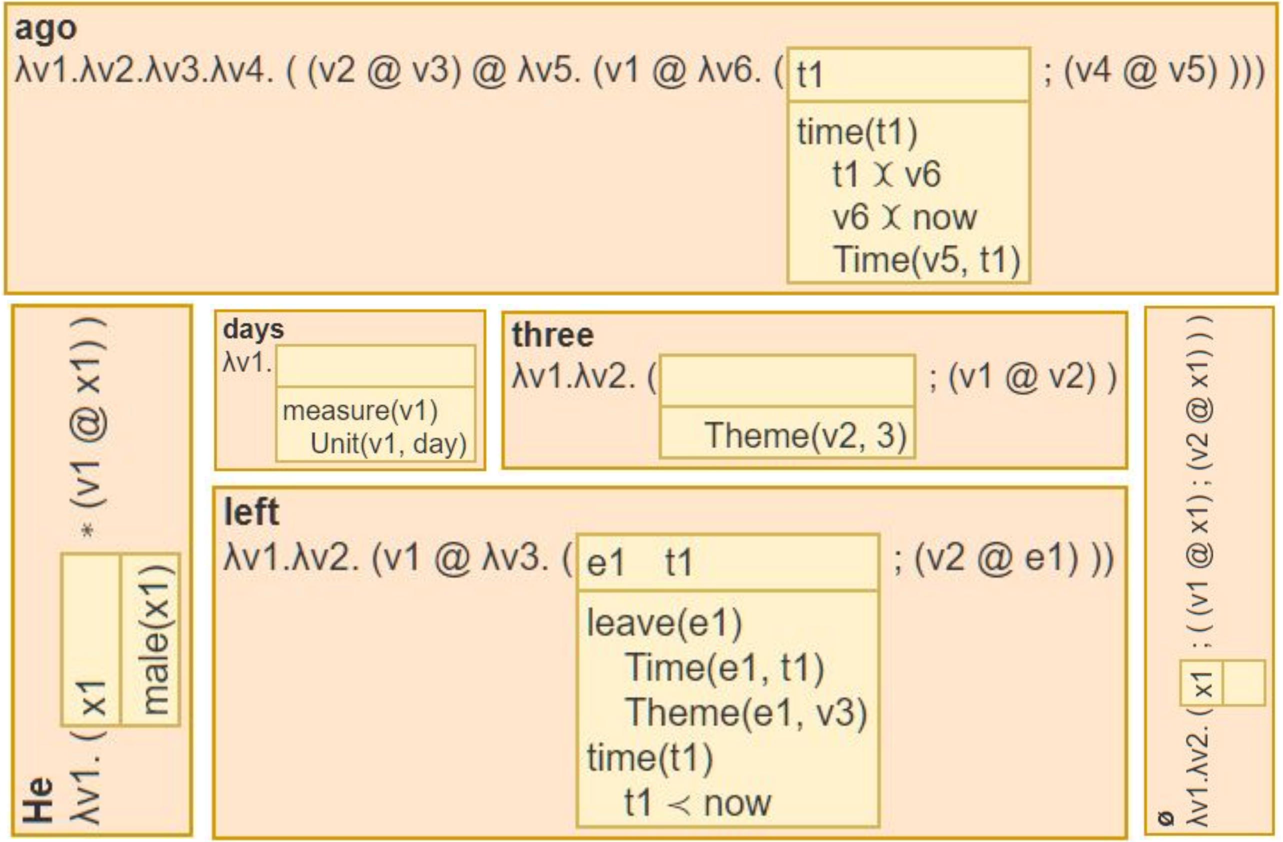
Semantic building blocks



Semantic building blocks



Semantic building blocks



Semantic building blocks

ago
λv1.λv2.λv3.λv4. ((v2 @ v3) @ λv5. (v1 @ λv6. ([t1 | time(t1), t1 X v6, v6 X now, Time(v5, t1)] ; (v4 @ v5))))

He
λv1. ([x1 | male(x1)] ∗ (v1 @ x1))

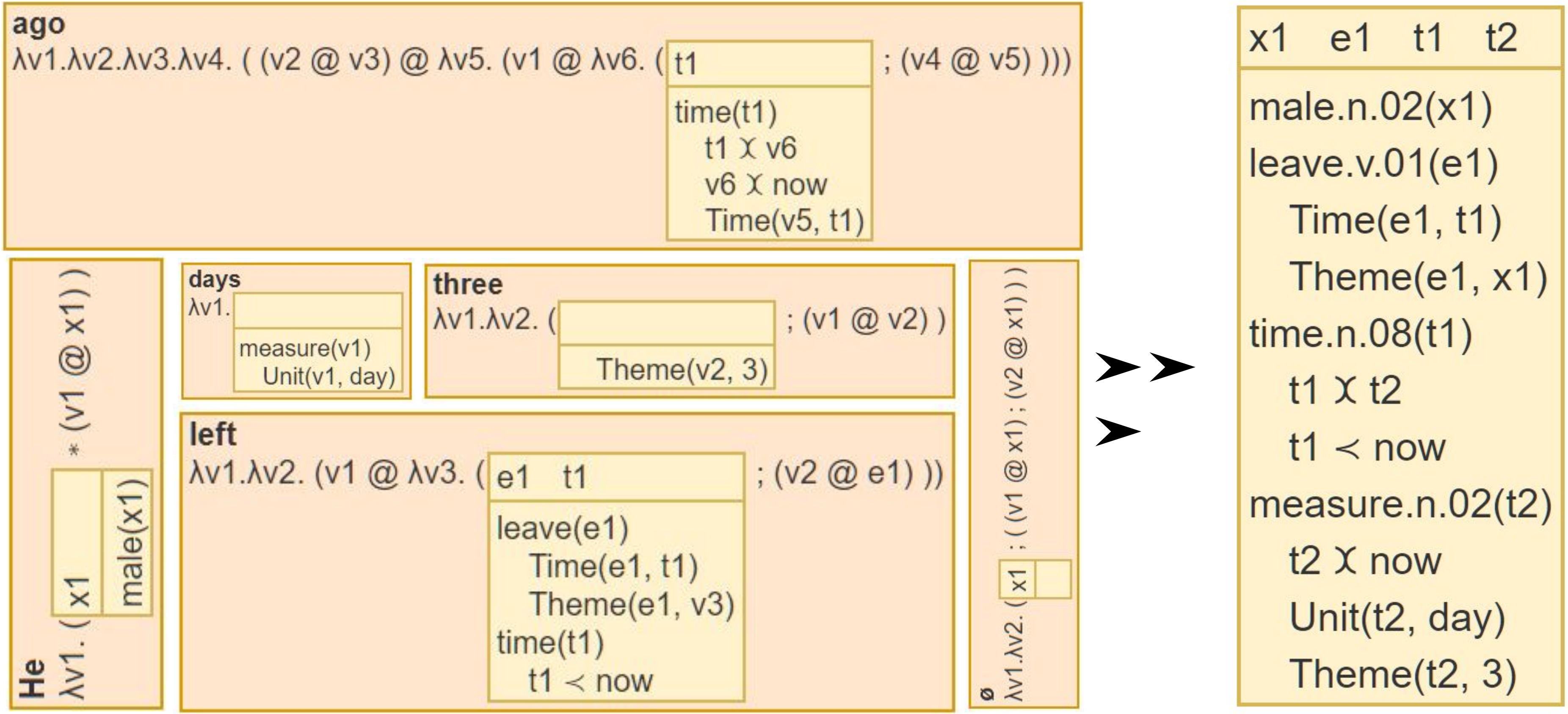
days
λv1. [| measure(v1), Unit(v1, day)]

three
λv1.λv2. ([| Theme(v2, 3)] ; (v1 @ v2))

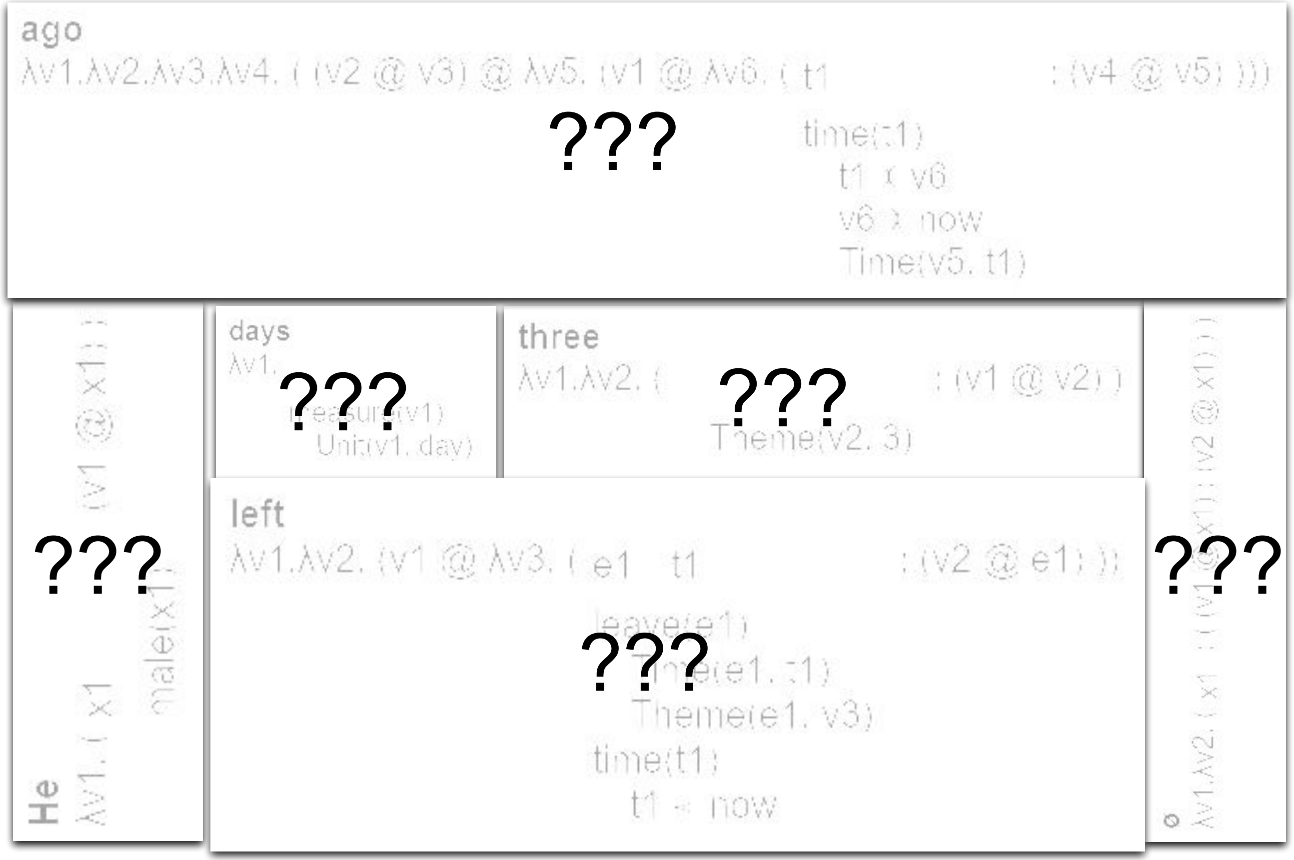
left
λv1.λv2. (v1 @ λv3. ([e1 t1 | leave(e1), Time(e1, t1), Theme(e1, v3), time(t1), t1 ≺ now] ; (v2 @ e1)))

ø
λv1.λv2. ([x1 |] ; ((v1 @ x1) ; (v2 @ x1)))

Semantic building blocks



Semantic building blocks

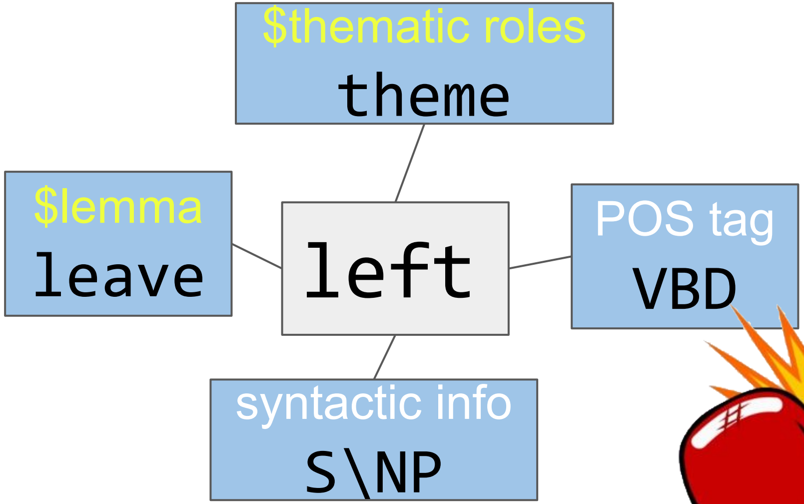


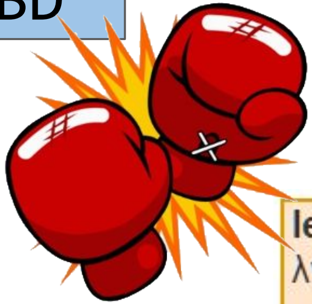
Goal

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)

GMB: Before UNISET

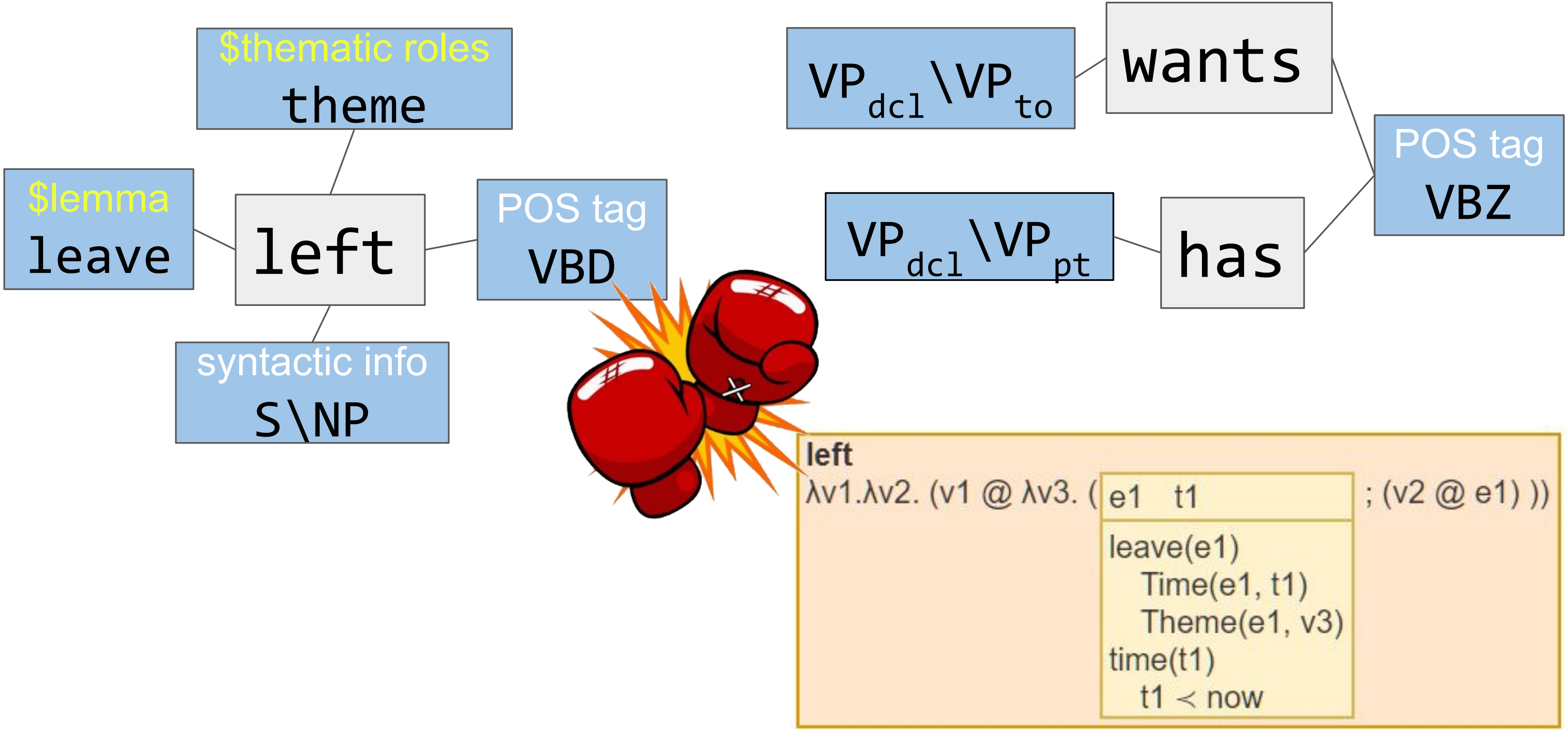




left

$$\lambda v1.\lambda v2.\ (v1\ @\ \lambda v3.\ (\begin{array}{|l|}\hline e1 \quad t1 \\ \hline leave(e1) \\ \quad Time(e1, t1) \\ \quad Theme(e1, v3) \\ time(t1) \\ \quad t1 \prec now \\ \hline \end{array}\ ;\ (v2\ @\ e1)\ ))$$

GMB: Before UNISET



GMB: Before UNISET

$thematic roles
theme

$lemma
leave

left

POS tag
VBD

syntactic info
S\NP

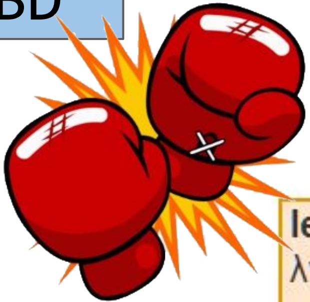
$VP_{dcl}\backslash VP_{to}$

wants

POS tag
VBZ

$VP_{dcl}\backslash VP_{pt}$

has



left

λv1.λv2. (v1 @ λv3. ($\boxed{\begin{array}{l} e1 \quad t1 \\ \hline leave(e1) \\ \quad Time(e1, t1) \\ \quad Theme(e1, v3) \\ time(t1) \\ \quad t1 \prec now \end{array}}$; (v2 @ e1)))

POS tags are not enough

a

the

every

no

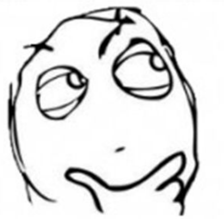
$lemma

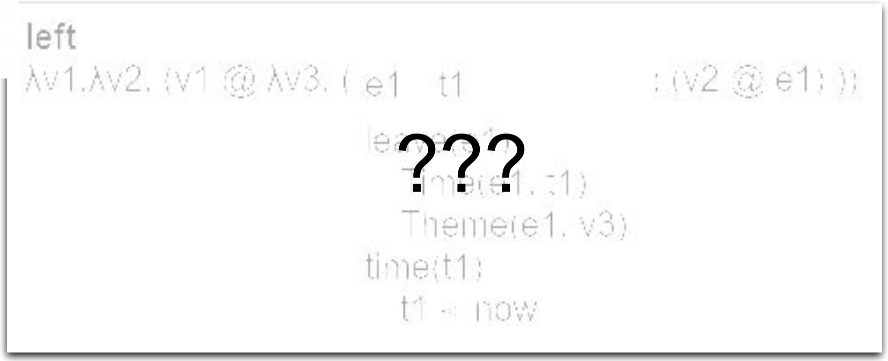
POS tag

DT

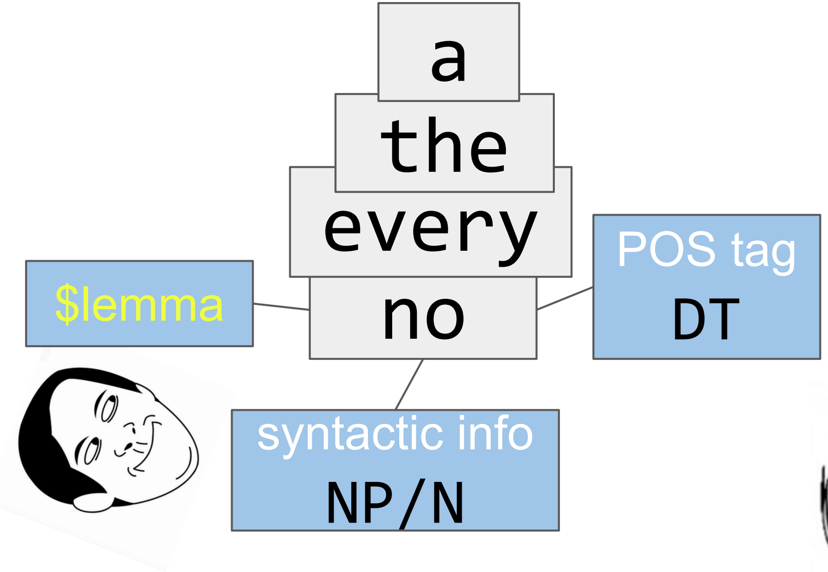
syntactic info

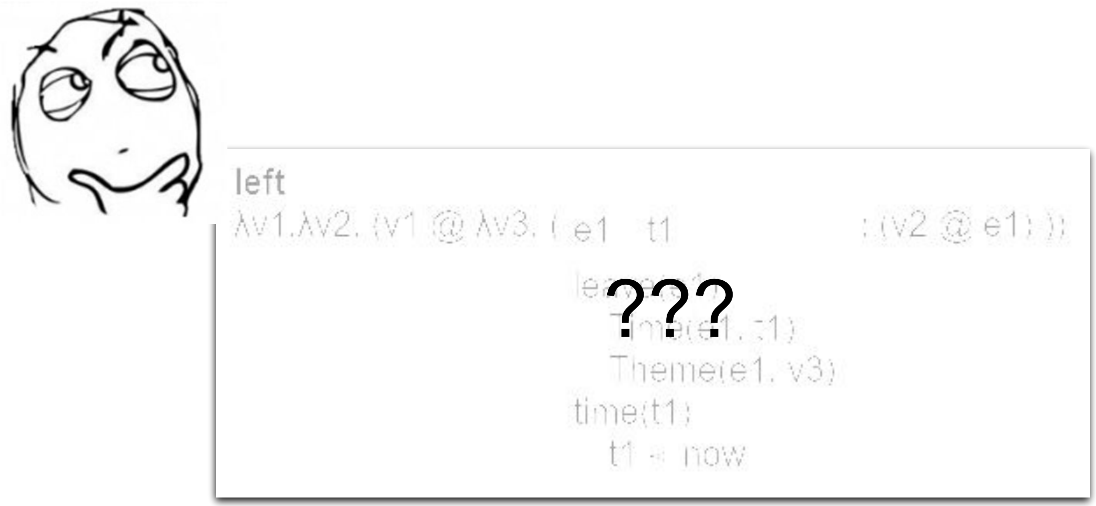
NP/N





POS tags are not enough





POS tags are not enough

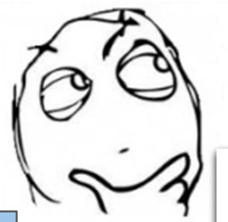
myself

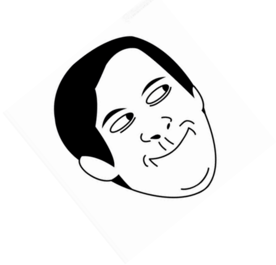
$lemma

POS tag
PRP

syntactic info
VP\VP

syntactic info
NP





left

???

Theme(e1, v3)

time(t1)

POS tags are not enough

but

or

and

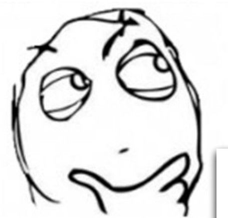
$lemma

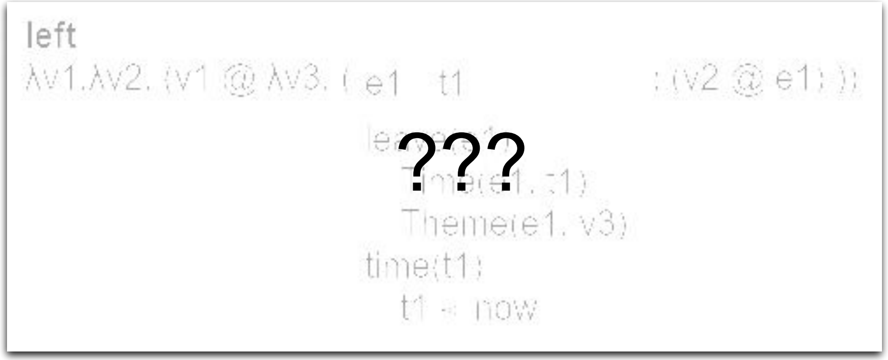
POS tag

cc

syntactic info

conj





POS tags are not enough

skillful

dead

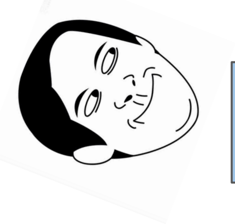
$lemma

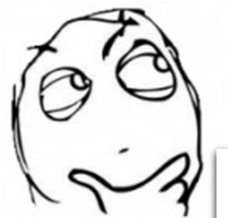
POS tag

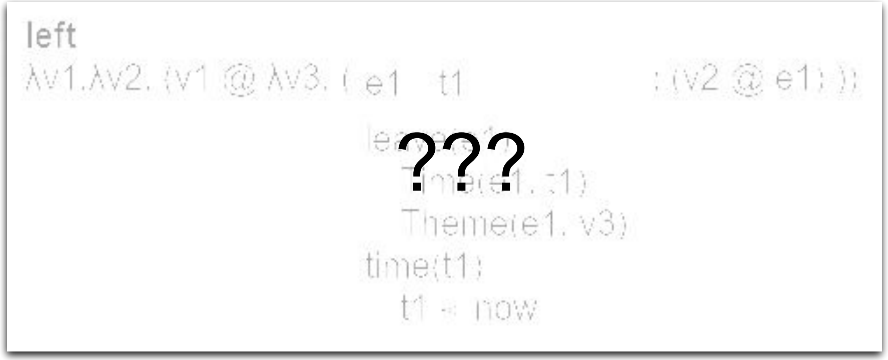
JJ

syntactic info

N/N







Something else is needed

- POS tags lack fine-grained semantic information
- Relying on lemmas → not language neutral
- Relying on CCG categories → framework/language dependent
- Sometimes even a CCG category, lemma, and a POS-tag do not suffice: and, any

Universal semantic tags

Universal semantic tags

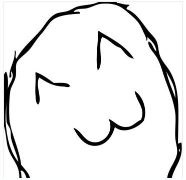
$symbol
leave

left

SEM-tag
EPS

syntactic info
S\NP

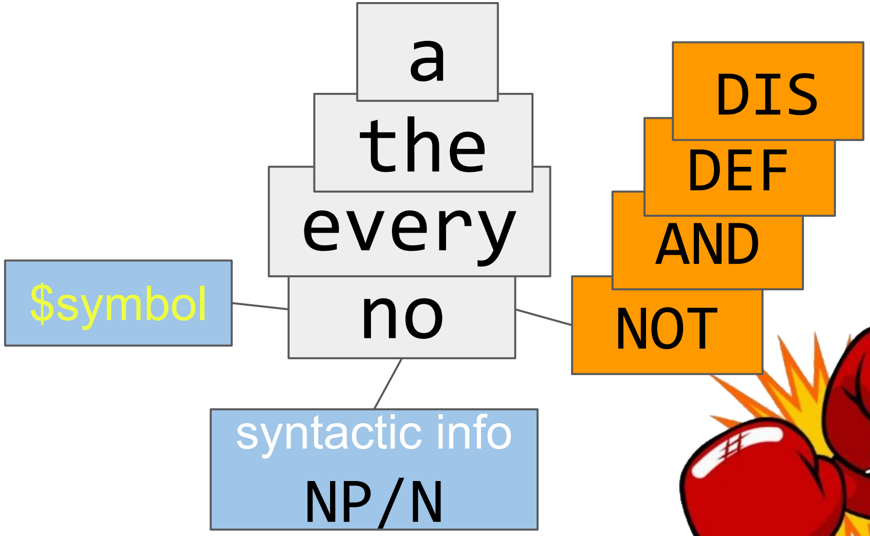
$thematic roles
theme

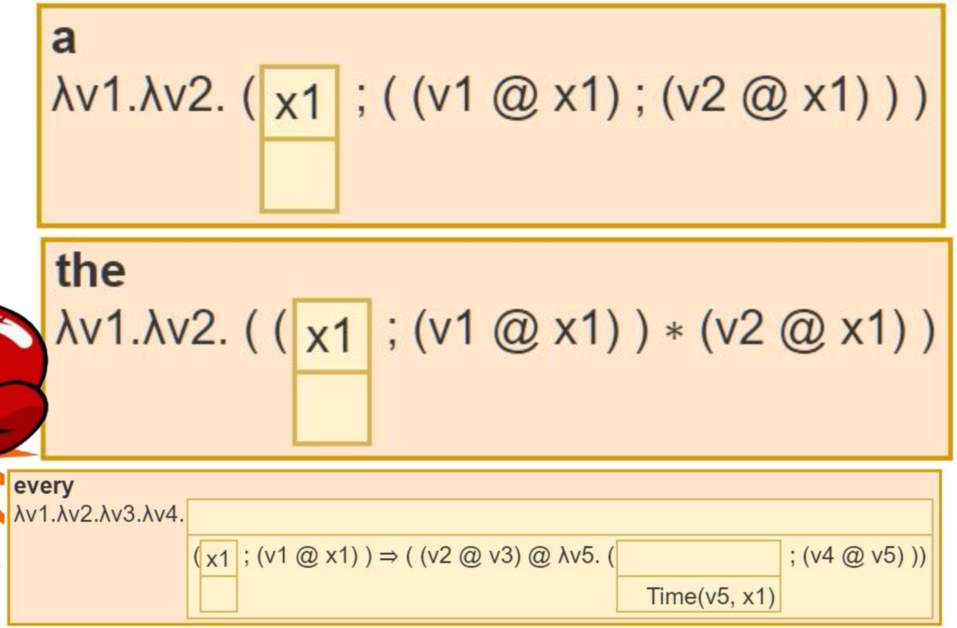


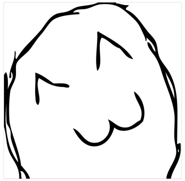
left

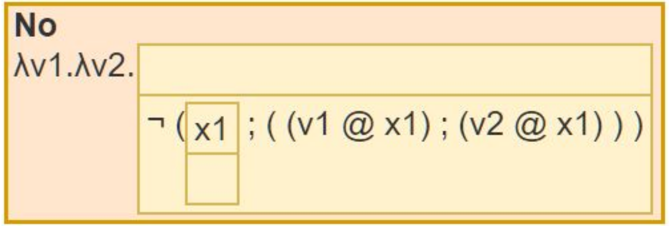
$\lambda v1.\lambda v2.\ (v1\ @\ \lambda v3.\ (\begin{array}{|l|} \hline e1\quad t1 \\ \hline \text{leave}(e1) \\ \quad \text{Time}(e1, t1) \\ \quad \text{Theme}(e1, v3) \\ \text{time}(t1) \\ \quad t1 \prec \text{now} \\ \hline \end{array}\ ;\ (v2\ @\ e1)\ ))$

Universal semantic tags

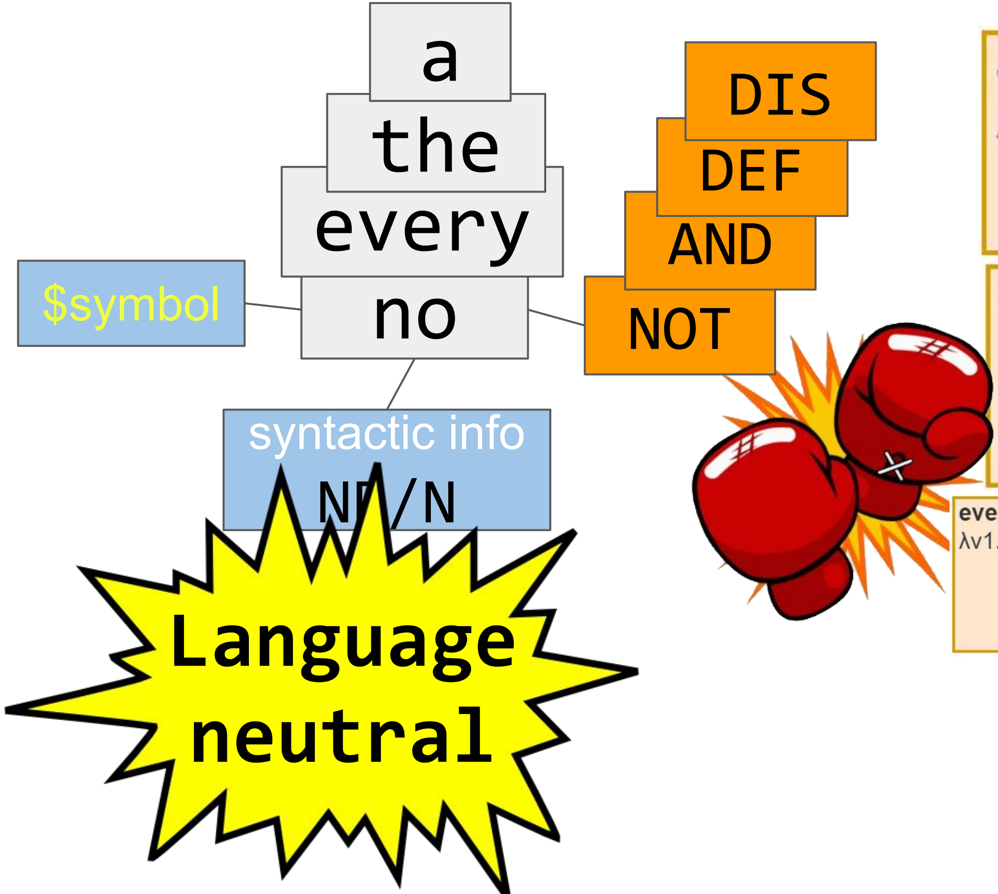








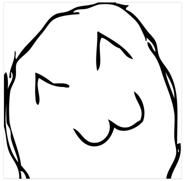
Universal semantic tags



a
λv1.λv2. (x1 ; ((v1 @ x1) ; (v2 @ x1)))

the
λv1.λv2. ((x1 ; (v1 @ x1)) * (v2 @ x1))

every
λv1.λv2.λv3.λv4. (x1 ; (v1 @ x1)) ⇒ ((v2 @ v3) @ λv5. (Time(v5, x1) ; (v4 @ v5)))



No
λv1.λv2. ¬ (x1 ; ((v1 @ x1) ; (v2 @ x1)))

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)

	QUC	Concrete quantity	two, six million, twice, 5
	QUV	Vague quantity	millions, many, enough
Attribute	COL	Colour	red, crimson, light‿blue
Attribute	IST	Intersective	open, vegetarian, quickly
Attribute	SST	Subsective	skillful surgeon, tall kid
Attribute	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

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

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

Speech Act

hi, bye	GRE	Greeting & parting
alas, ah	ITJ	Interjections & exclamations
err	HES	Hesitation
who, which, ?	QUE	Interrogative

Tense & Aspect

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

Tense & Aspect

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

Events

EXS Untensed simple — to walk, is eaten, destruction

ENS Present simple — we walk, he walks

EPS Past simple — ate, went

EXG Untensed progressive — is running

EXT Untensed perfect — has eaten

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

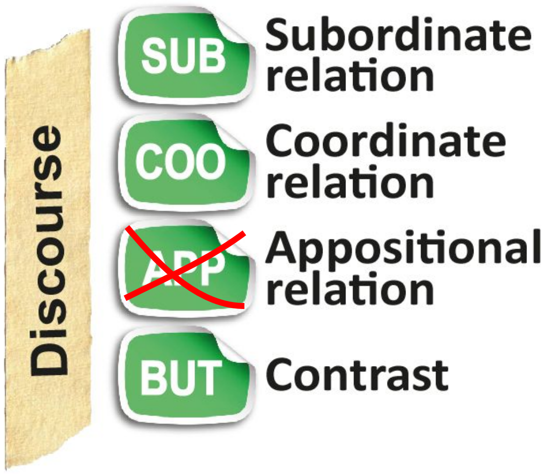
ALT	Alternatives & repetitions	another, different, again
XCL	Exclusive	only, just
NIL	Empty semantics	to, .,
DIS	Disjunction & existential quantif.	a, some, any, or
IMP	Implication	if, when, unless
AND	Conjunction & universal quantif.	every, and, who, any

Modality

not, no, neither, without	NOT	Negation
must, should, have to	NEC	Necessity
might, could, perhaps, alleged, can	POS	Possibility

Temporal Entity

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

Discourse

SUB	Subordinate relation	that, while, because
COO	Coordinate relation	so, ;, and
~~APP~~	Appositional relation	which, —
BUT	Contrast	but, yet

Deixis

here, this, above	DXP	Place deixis
just, later, tomorrow	DXT	Temporal deixis
latter, former, above	DXD	Discourse deixis

Named Entity

PER	Person	Axl Rose, Sherlock Holmes
GPE	Geo-political entity	Paris, Japan
GPO	Geo-political origin	Parisian, 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 formal semantics study modular

Tagging & Semantics

Formal compositional semantics are less favoured:

- Semantics problems
- Difficult to scale up

Make formal semantics study modular

NLP community loves tagging/labeling tasks

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

Data & Results

- Gold EN documents (34.7K)
- Silver EN documents (1.6M)

Universal Semantic Tags

version	# en	# de	# it	# nl	silver inc.	release date	download
0.1.0	5438	0	0	0	yes	01-05-2018	19 MB ZIP file

- Baseline (UniGram) ~82%
- Stanford tagger ~88.8%
- NN tagger (AUX UPOS) ~92.7% (M. Abdou)

Challenges

- Account for wide-coverage compositional semantics
- Keep UNISET independent from CCG
- Prevent the number of sem-tags from increasing

Conclusion

- Facilitates determining lexical semantics
- Contributes to cross-lingual applications
- Useful for other NLP applications
- Useful for other semantic parsers or RTE systems:

  (ccg2lambda, LangPro, UDepLambda,...)

Future work

- Cover more semantic phenomena (data-driven)
- Measure an inter-annotator agreement
- Reorganize tagset to simplify learning

Attribute

- QUC Concrete quantity
- QUV Vague quantity
- COL Colour
- IST Intersective
- SST Subsective
- PRI Privative
- DEG Degree
- INT Intensifier
- REL Relation
- SCO Score

Modality

- NOT Negation
- NEC Necessity
- POS Possibility

Tense & Aspect

- NOW Present tense
- PST Past tense
- FUT Future tense
- PRG Progressive
- PFT Perfect

Comparative

- EQU Equative
- MOR Comparative positive
- LES Comparative negative
- TOP Superlative positive
- BOT Superlative negative
- ORD Ordinal

Unnamed Entity

- CON Concept
- ROL Role
- GRP Group

Discourse

- SUB Subordinate relation
- COO Coordinate relation
- ~~APP~~ Appositional relation
- BUT Contrast

Events

- EXS Untensed simple
- ENS Present simple
- EPS Past simple
- EXG Untensed progressive
- EXT Untensed perfect

Anaphoric

- PRO Anaphoric & deictic pronoun
- DEF Definite
- HAS Possessive pro.
- REF Reflexive & reciprocal pro.
- EMP Emphasizing pro.

Logical

- ALT Alternatives & repetitions
- XCL Exclusive
- NIL Empty semantics
- DIS Disjunction & existential quantif.
- IMP Implication
- AND Conjunction & universal quantif.

Speech Act

- GRE Greeting & parting
- ITJ Interjections & exclamations
- HES Hesitation
- QUE Interrogative

Deixis

- DXP Place deixis
- DXT Temporal deixis
- DXD Discourse deixis

Named Entity

- PER Person
- GPE Geo-political entity
- GPO Geo-political origin
- LOC Geographical location
- ORG Organization
- ART Artifact
- HAP Happening
- UOM Unit of measurement
- CTC Contact info
- URL URL
- LIT Literal use of names
- NTH Other names

Temporal Entity

- DAT Full date
- DOM Day of Month
- YOC Year of century
- DOW Day of week
- MOY Month of year
- DEC Decade
- CLO Clocktime

Lexical Semantics

Parallel MEANING BANK