

Meaning Banking and Beyond

Valerio Basile
Wimmics, Inria

November 18, 2015



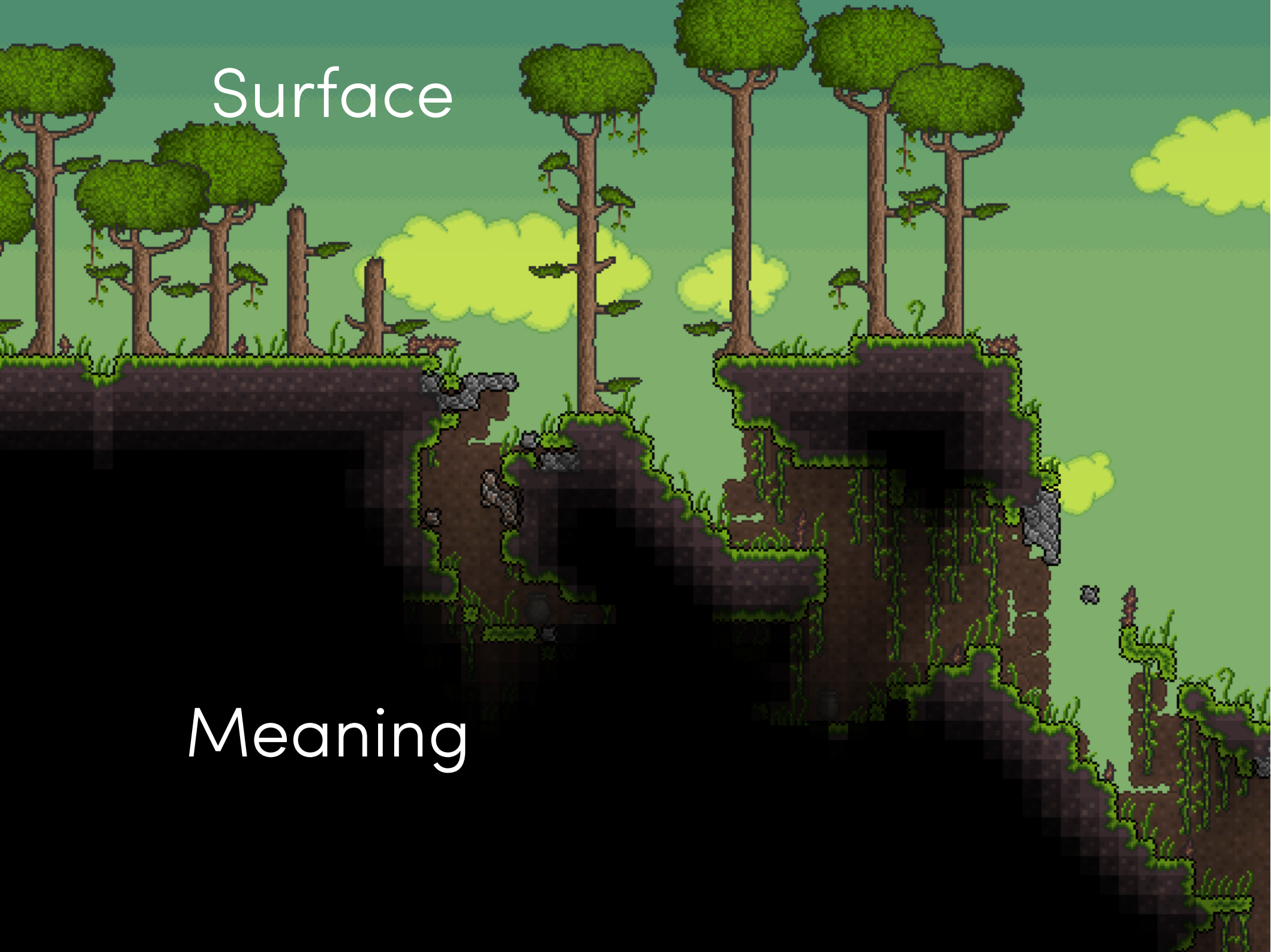
*Semantics is a well-kept secret in texts,
accessible only to humans.*

– Anonymous

I BEG TO DIFFER

Surface

Meaning



Step by step analysis

Dividing text into words

Step by step analysis

Dividing text into words

Labeling words

Emotions play an important role in decision making

p. noun verb det adjective noun prep noun noun

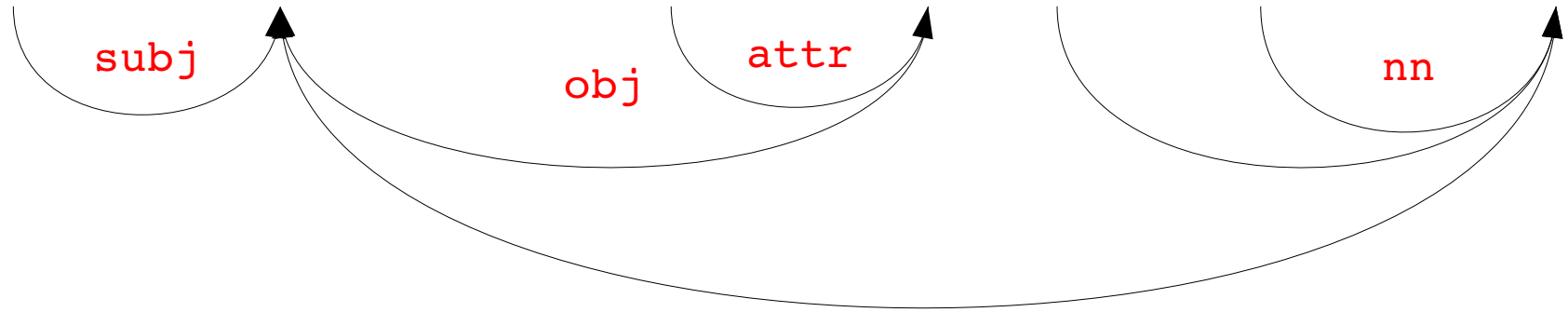
Step by step analysis

Dividing text into words

Labeling words

Finding links between words

Emotions play an important role in decision making



Step by step analysis

Dividing text into words

Labeling words

Finding links between words

Extracting the meaning

Emotions play an important role in decision making

play(emotion, role)

important(role)

|emotion| >1

...

Step by step analysis

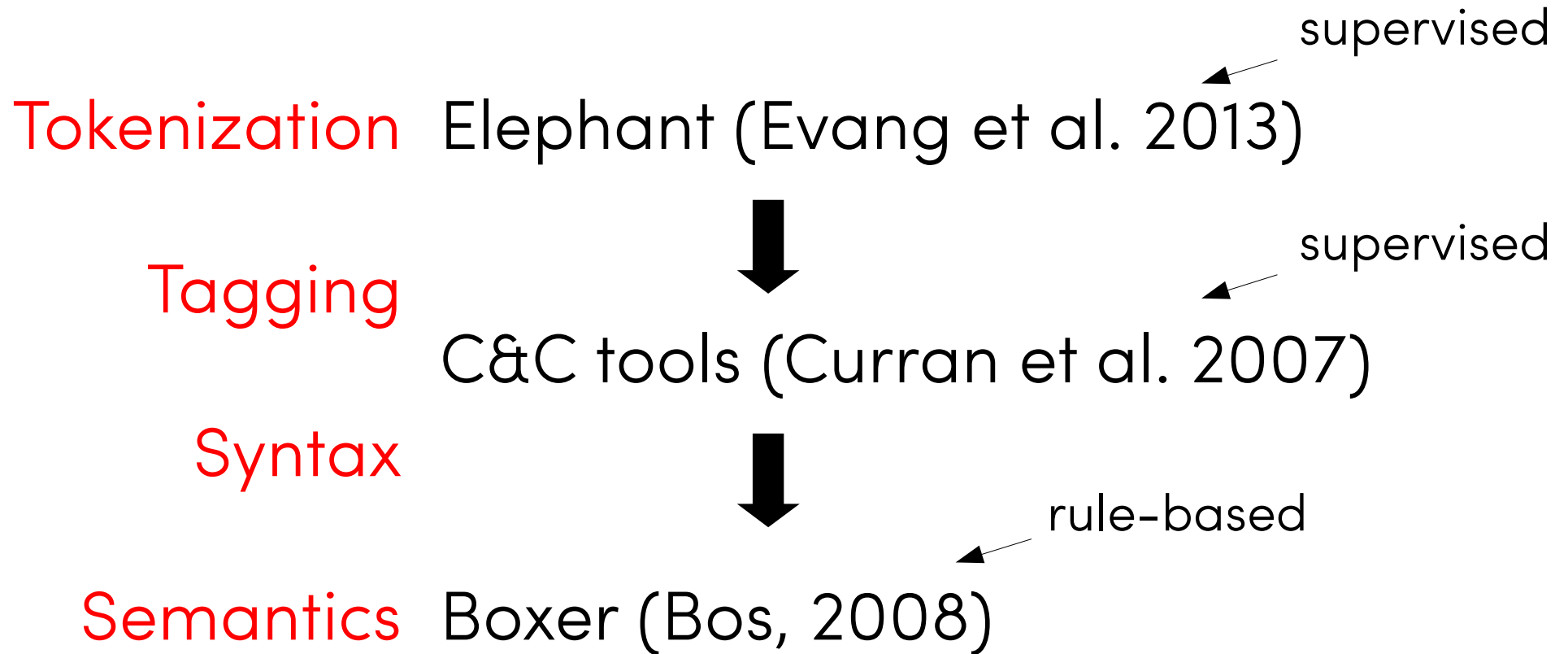
Dividing text into words **Tokenization**

Labeling words **Tagging**

Finding links between words **Syntax**

Extracting the meaning **Semantics**

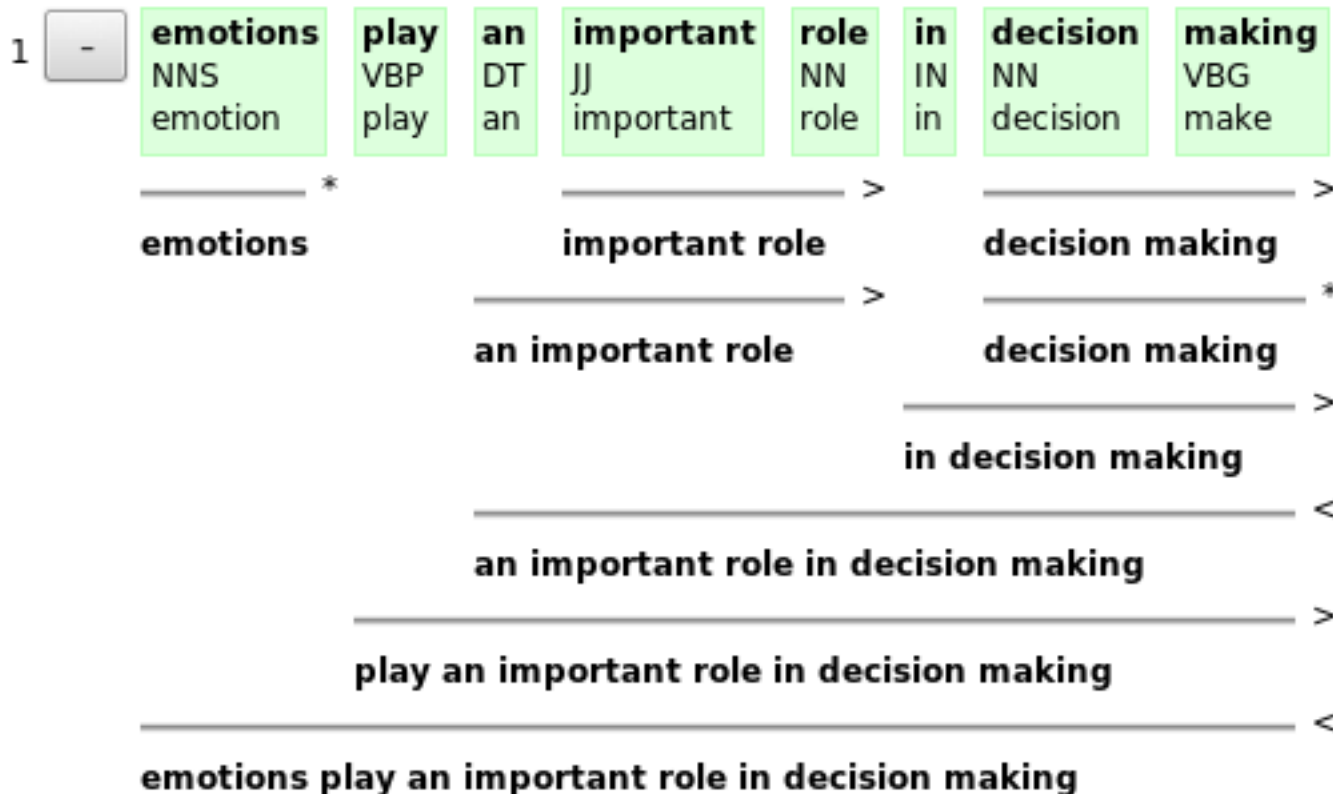
Software pipeline



Emotions play an important role in decision making

Derivation

☒ POS ☒ lemmas ☐ namex ☐ syntax ☐ semantics



DRS

x1	x2	s1	x3	x4	e1
emotion(x1)					
Experiencer(s1, x2)					
important(s1)					
role(x2)					
decision(x4)					
of(x3, x4)					
make(x3)					
in(x2, x3)					
play(e1)					
Agent(e1, x1)					
Co-Agent(e1, x2)					

<http://gmb.let.rug.nl/webdemo>

Supervised learning

Show **enough** examples to the machine
and it will **learn** and **generalize**.

As opposed to unsupervised learning,
e.g. clustering

Supervised learning

enough examples

=

millions of words + annotation

that is, an annotated text **corpus**

Examples:

Penn *treebank* (syntax)

Semcor (word senses)

TWITA (sentiment), ...

What about a semantics?

What about a semantics?

Meaning Bank
(a treebank for semantics)



How to build a meaning bank

Manually:

Collection of texts
+
Expert annotation

How to build a meaning bank

Manually:

Collection of texts
+
Expert annotation

DO NOT TRY THIS AT HOME

How to build a meaning bank

By **bootstrap**:

Collection of texts

+

Analysis software

+

Manual correction

Collection of texts

English
Public domain
Whole documents
Short
Open domain

Collection of texts

73,352 documents

10,103 (accepted)

6.3 sentence per document

~90% from VoA (newswire)

Jokes, legal text, fables, ...

Analysis software



Elephant
tokenization



C&C tools
taggins & parsing



Boxer
Semantic analysis

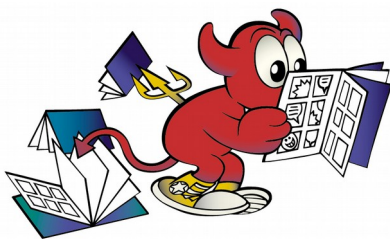


DRS

x1	x2	s1	x3	x4	e1
emotion(x1)					
Experiencer(s1, x2)					
important(s1)					
role(x2)					
decision(x4)					
of(x3, x4)					
make(x3)					
in(x2, x3)					
play(e1)					
Agent(e1, x1)					
Co-Agent(e1, x2)					



GNU Make



Daemon process

Manual correction

Silver standard

Experts and the crowd

The GMB Explorer

Document **1761 of 10103**, ID: /

[|< first](#) [<< previous](#) [next >>](#) [last >|](#) [random](#)

Status: **accepted** [history](#)

Change to: Comment:

[report issue](#)

size: 6 sentences, 112 tokens
last processed: 29 September 2015, 08:23:10
C&C tools/Boxer revision: 2561

Filter by part:
Filter by status:
Filter by subcorpus:
Warnings:
Effective BOWs:

[metadata](#) [raw](#) [tokens](#) [sentences](#) [discourse](#) [19 bits of wisdom](#) [0 warnings](#)

Show: ☐ POS ☐ lemmas ☒ namex ☐ animacy ☒ senses ☐ roles ☐ relations ☐ scope ☐ reference ☐ syntax ☒ semantics

1 **Iranian**
gpe-nam
λv1. λv2. (; (v1 @ v2))
named(x1, iran, gpe)
of(v2, x1)

officials
O
official.n.01
λv1.

say
O
state.v.01
λv1. λv2. λv3. (v2 @ λv4. (; (v3 @ e1)))
say(e1)
Cause(e1, v4)
Topic(e1, p1)
p1: (v1 @ λv5.
e1 ⊆ t2
t2 = t1

Pakistan
geo-nam
pakistan.n.01
λv1.

has
O
λv1. λv2. λv3. ((v1 @ v2) @ λv4. (; (v3 @ v4)))
now(t1)
x1 = t1
e1 ⊇ x1
v4 ⊇ e1

freed
O
free.v.01
λv1. λv2. λv3. (v2 @ λv4. (

2 **Iran**
geo-nam
iran.n.01
λv1.

's
O
λv1. λv2. λv3. (v1 @ λv4. ((; (v2 @ x1)) * (v3 @ x1)))
of(x1, v4)

state-run
O
λv1. λv2. (; (v1 @ v2))
Experiencer(s1, v2)
state-run(s1)

news
O
news.n.01
λv1. λv2. (; (v1 @ v2))
news(x1)
from(v2, x1)

agency
O
agency.n.01
λv1.

'
O
λv1. λv2. λv3. (v2 @ λv4. ((v1 @ λv5. (

Iran
λv1. (* (v1 @ x1))
named(x1, iran, geo)

news agency
λv1.
news(x1)
from(v1, x1)
agency(v1)

Iran 's
λv1. λv2. (* ((; (v1 @ x2)) * (v2 @ x2)))
named(x1, iran, geo) of(x2, x1)

state-run news agency
λv1.

, IRNA
λv1. λv2. (v1 @ λv3. (named(x1, irna) rel(v3, x1)

<http://gmb.let.rug.nl/explorer>

The GMB Explorer

33 users

173,173 annotations
(including automatically generated)


<http://gmb.let.rug.nl/explorer>

Gamification

?

wordrobe

play what you mean

 Pointers Questions left until drawer is completed: 7

Cuban *President Fidel Castro* says he is not well enough to attend celebrations in *Havana* to belatedly mark **his** 80th birthday.

☒ 1) President Fidel Castro
☐ 2) Havana

Place your bet: low high



PLAY TWINS

Noun or verb? Good game for beginners. Double the trouble, twice the fun!

1  BSenthilKumar | 489 points



PLAY SENSES

Identify the correct sense of a word. Quite a challenging game. Does it make sense?

1  vindaci_TT | 396 points



PLAY NAMES

Find out what type of proper name is used. A game for starters. What's in a name?

1  christine | 372 points



PLAY POINTERS

Select the referent of a pronoun. Not for absolute beginners! Point of no return?

1  Maikel_CD | 797 points



PLAY BRIDGES

Find out if a referent is new, or already introduced in the text. Not a bridge too far!

1  aprilwent | 189 points



PLAY BURGERS

Choose the topping of your noun burger. Which preposition fits best?

1  aprilwent | 433 points



PLAY ANIMALS

Dead or alive? Classify the level of animacy of phrases.

1  aprilwent | 328 points



PLAY ROLES

Find out what thematic roles are played in texts. A game full of drama! Are you a role model?


1  vindaci_TT | 368 points

<http://www.wordrobe.org>

?

wordrobe

play what you mean



Personal information

Name: Valerio

Email: vbasile@rug.nl

Country:

Website: <http://www.let.rug.nl/basile>

About me:

Completed drawers: 380

Names: 61

Pointers: 110

Twins: 43

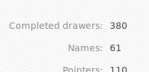
Senses: 60

Burgers: 50

Animals: 50

Roles: 3

Bridges: 3



<http://www.wordrobe.org>

Gamification

59,413 Answers

13 Games

1,732 Players

2 Datasets

<http://www.wordrobe.org>



A semantically annotated resource
that anyone can edit.

Johan Bos, Valerio Basile, Kilian Evang, Noortje Venhuizen,
Johannes Bjerva (forthcoming): *The Groningen Meaning Bank*.
In Handbook of Linguistic Annotation. Berlin: Springer.

Valerio Basile, Johan Bos, Kilian Evang, Noortje Venhuizen
Developing a large semantically annotated corpus. LREC 2012

Valerio Basile, Johan Bos, Kilian Evang, Noortje Venhuizen
A platform for collaborative semantic annotation. EACL 2012

<http://gmb.let.rug.nl>

Beyond Meaning Banking

Beyond Meaning Banking

ALOOF

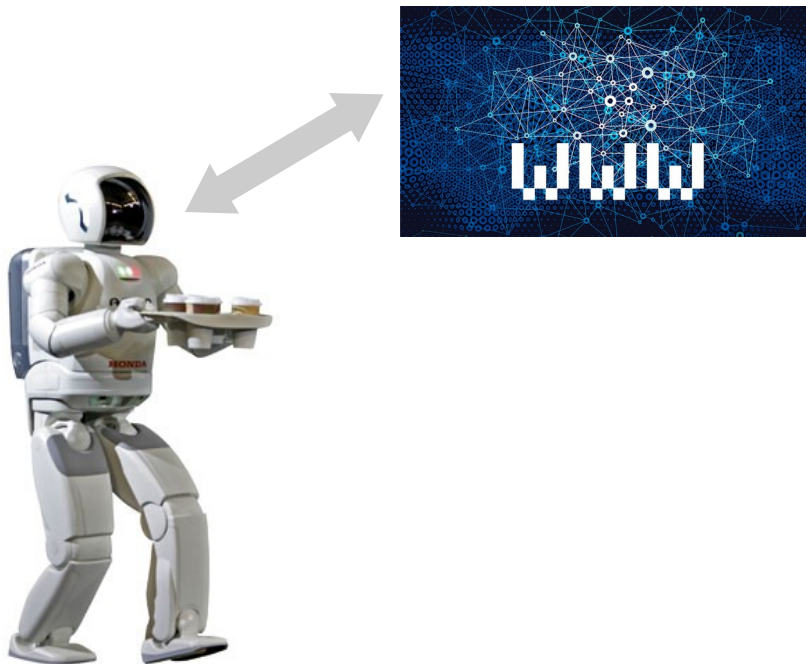
Autonomous learning of the meaning of objects



Beyond Meaning Banking

ALOOOF

Autonomous learning of the meaning of objects

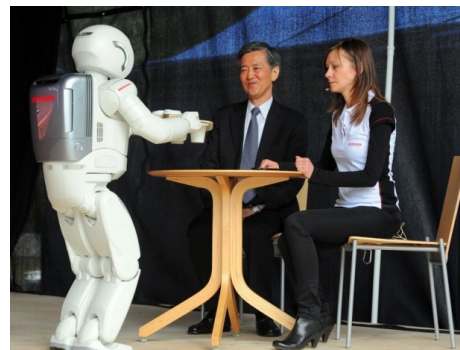


Beyond Meaning Banking

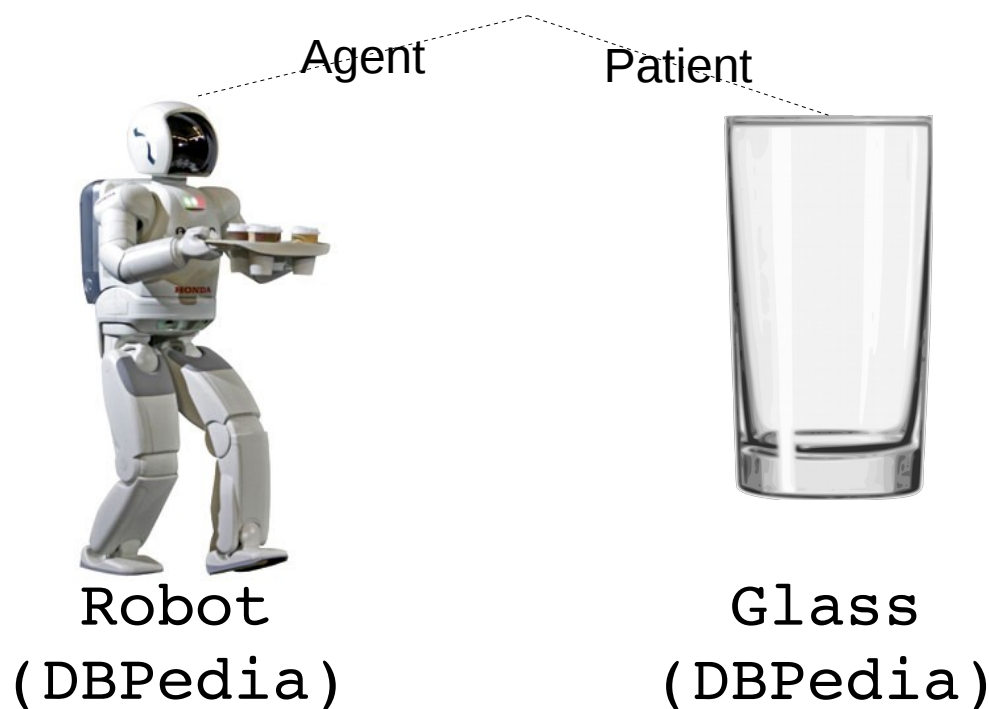
Bridging semantic analysis with entity linking
to
build a knowledge base by reading the Web

Beyond Meaning Banking

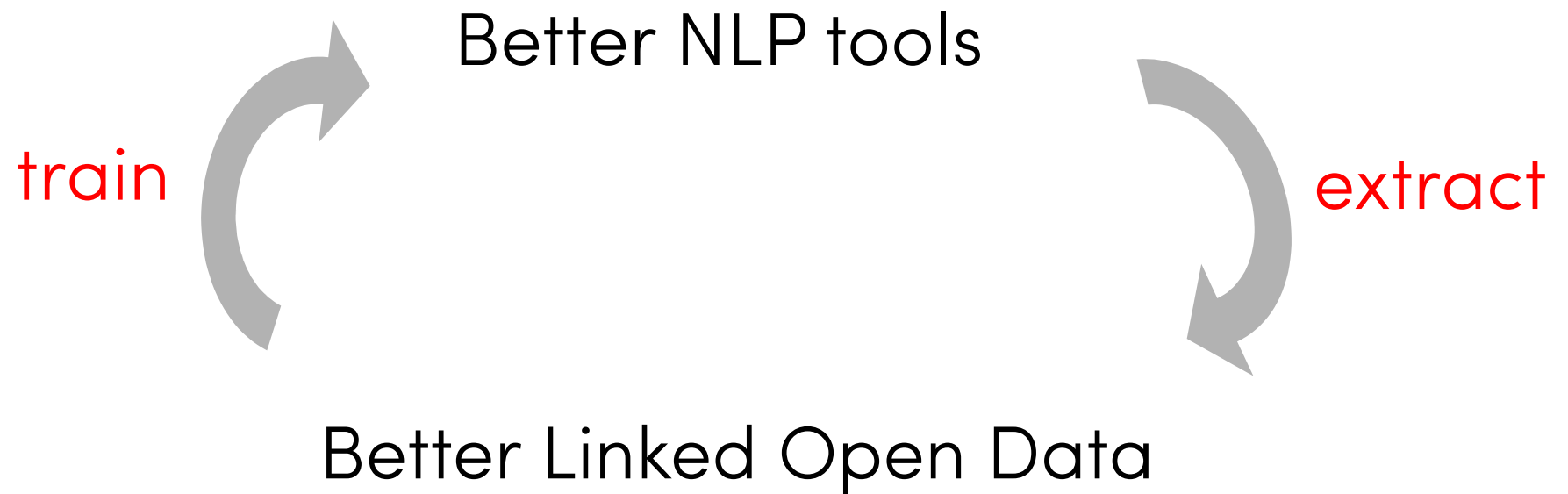
When a **glass** or **cup** is emptied,
The **robot** will
ask if **it** should
serve more



serving
(FrameNet)



Towards a Web Meaning Bank



Fin

Meaning Banking and Beyond
Valerio Basile
November 18, 2015