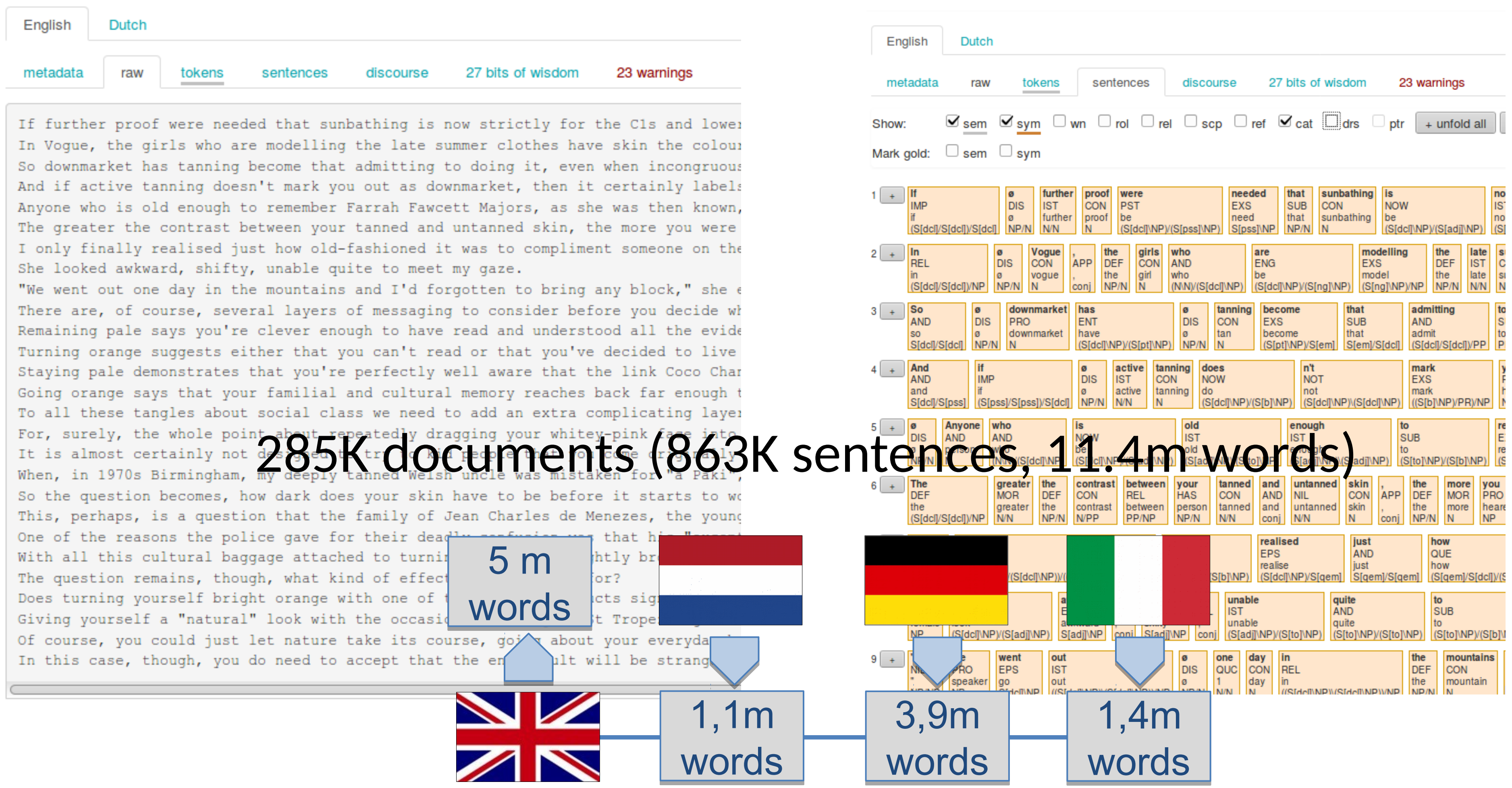
http://gmb.let.rug.nl

PMBots

From Raw Parallel Documents to the Documents with several annotation layers



English | Dutch

metadata | raw | tokens | sentences | discourse | 27 bits of wisdom | 23 warnings

```
If further proof were needed that sunbathing is now strictly for the C1s and lowe
In Vogue, the girls who are modelling the late summer clothes have skin the colou
So downmarket has tanning become that admitting to doing it, even when incongruous
And if active tanning doesn't mark you out as downmarket, then it certainly labels
Anyone who is old enough to remember Farrah Fawcett Majors, as she was then known,
The greater the contrast between your tanned and untanned skin, the more you were
I only finally realised just how old-fashioned it was to compliment someone on the
She looked awkward, shifty, unable quite to meet my gaze.
"We went out one day in the mountains and I'd forgotten to bring any block," she e
There are, of course, several layers of messaging to consider before you decide wh
Remaining pale says you're clever enough to have read and understood all the evide
Turning orange suggests either that you can't read or that you've decided to live
Staying pale demonstrates that you're perfectly well aware that the link Coco Char
Going orange says that your familial and cultural memory reaches back far enough t
To all these tangles about social class we need to add an extra complicating laye
For, surely, the whole point about repeatedly dragging your whitey-pink face into
It is almost certainly not designed to try to kid people that you come originally
When, in 1970s Birmingham, my deeply tanned Welsh uncle was mistaken for "a Paki",
So the question becomes, how dark does your skin have to be before it starts to wo
This, perhaps, is a question that the family of Jean Charles de Menezes, the young
One of the reasons the police gave for their deadly confusion was that his "expect
With all this cultural baggage attached to turni                 htly br
The question remains, though, what kind of effect                  for?
Does turning yourself bright orange with one of t                 cts sig
Giving yourself a "natural" look with the occasi                 t Trope
Of course, you could just let nature take its course, going about your everyda
In this case, though, you do need to accept that the en      ult will be strang
```

English | Dutch

metadata | raw | tokens | sentences | discourse | 27 bits of wisdom | 23 warnings

Show: ☑ sem ☑ sym □ wn □ rol □ rel □ scp □ ref ☑ cat □ drs □ ptr [+ unfold all]

Mark gold: □ sem □ sym

1 [+] If IMP if (S[dcl]/S[dcl])/S[dcl] | ø DIS ø NP/N | further IST further N/N | proof CON proof N | were PST be (S[dcl]\NP)/(S[pss]\NP) | needed EXS need S[pss]\NP | that SUB that NP/N | sunbathing CON sunbathing N | is NOW be (S[dcl]\NP)/(S[adj]\NP) | no IS no (S

2 [+] In REL in (S[dcl]/S[dcl])/NP | ø DIS ø NP/N | Vogue CON vogue N | , APP , conj | the DEF the NP/N | girls CON girl N | who AND who (N\N)/(S[dcl]\NP) | are ENG be (S[dcl]\NP)/(S[ng]\NP) | modelling EXS model (S[ng]\NP)/NP | the DEF the NP/N | late IST late N/N | s C s N

3 [+] So AND so S[dcl]/S[dcl] | ø DIS ø NP/N | downmarket PRO downmarket N | has ENT have (S[dcl]\NP)/(S[pt]\NP) | ø DIS ø NP/N | tanning CON tan N | become EXS become (S[pt]\NP)/S[em] | that SUB that S[em]/S[dcl] | admitting AND admit (S[dcl]/S[dcl])/PP | to S to P

4 [+] And AND and S[dcl]/S[pss] | if IMP if (S[pss]/S[pss])/S[dcl] | ø DIS ø NP/N | active IST active N/N | tanning CON tanning N | does NOW do (S[dcl]\NP)/(S[b]\NP) | n't NOT not (S[dcl]\NP)\(S[dcl]\NP) | mark EXS mark ((S[b]\NP)/PR)/NP | y P h N

5 [+] ø DIS ø NP/N | Anyone AND person N | who AND who (N\N)/(S[dcl]\NP) | is NOW be (S[dcl]\NP)/(S[adj]\NP) | old IST old (S[adj]\NP)/(S[to]\NP) | enough IST enough (S[adj]\NP)\(S[adj]\NP) | to SUB to (S[to]\NP)/(S[b]\NP) | re E re (S

6 [+] The DEF the (S[dcl]/S[dcl])/NP | greater MOR greater N/N | the DEF the NP/N | contrast CON contrast N/PP | between REL between PP/NP | your HAS person NP/N | tanned CON tanned N/N | and AND and conj | untanned NIL untanned N/N | skin CON skin N | , APP , conj | the DEF the NP/N | more MOR more N | you PRO heare NP

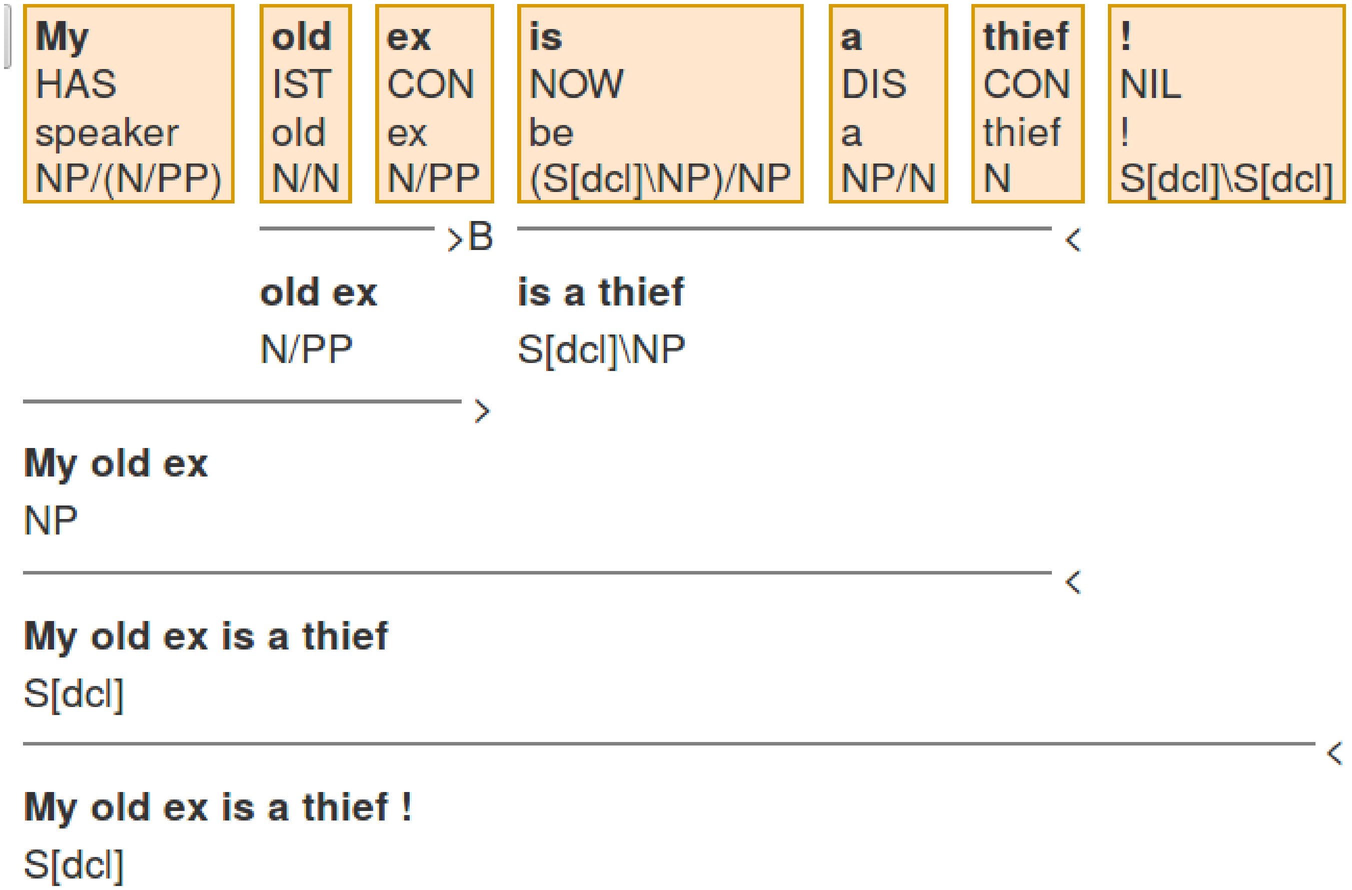
(S[dcl]\NP))/(| (S[b]\NP) | realised EPS realise (S[dcl]\NP)/S[qem] | just AND just S[qem]/S[qem] | how QUE how (S[qem]/S[dcl])/(S

NP | (S[dcl]\NP)/(S[adj]\NP) | a E a S[adj]\NP | , conj | S[adj]\NP | , conj | unable IST unable (S[adj]\NP)/(S[to]\NP) | quite AND quite (S[to]\NP)/(S[to]\NP) | to SUB to (S[to]\NP)/(S[b]\NP)

9 [+] " NIL " NP/NP | we PRO speaker NP | went EPS go S[dcl]\NP | out IST out ((S[dcl]\NP)\(S[dcl]\NP))/NP | ø DIS ø NP/N | one QUC 1 N/N | day CON day N | in REL in ((S[dcl]\NP)\(S[dcl]\NP))/NP | the DEF the NP/N | mountains CON mountain N

285K documents (863K sentences, 11.4m words)

Annotated English Document



My	**old**	**ex**	**is**	**a**	**thief**	**!**
HAS	IST	CON	NOW	DIS	CON	NIL
speaker	old	ex	be	a	thief	!
NP/(N/PP)	N/N	N/PP	(S[dcl]\NP)/NP	NP/N	N	S[dcl]\S[dcl]

>B

old ex
N/PP

<

is a thief
S[dcl]\NP

>

My old ex
NP

<

My old ex is a thief
S[dcl]

<

My old ex is a thief !
S[dcl]

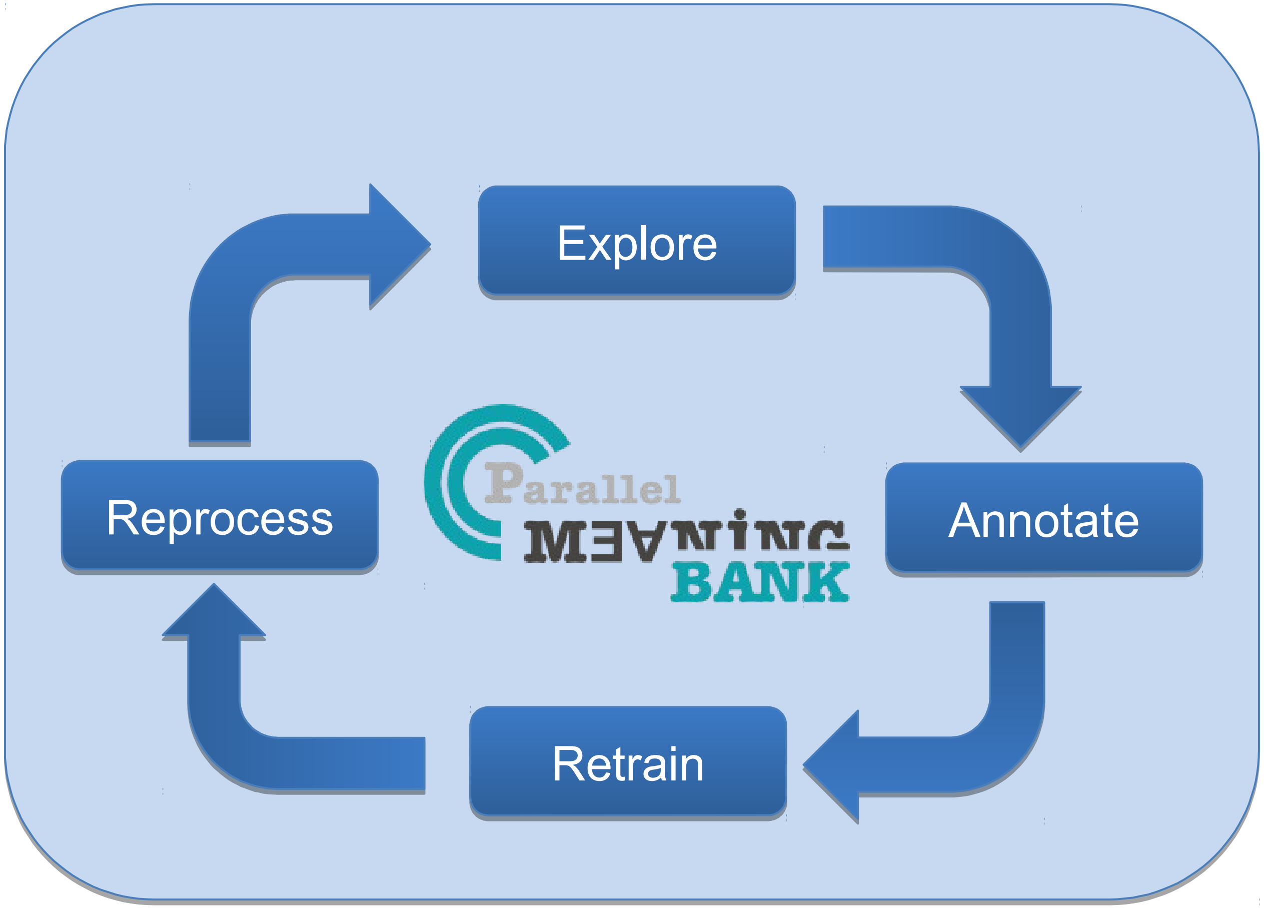
Usage of the PMB

- Train and evaluate several systems for the languages
  - Tokenizer
  - Semantic tagger
  - Coreference resolution system
  - CCG parser
  - Semantic parser
- Studying compositional semantics in a data-driven fashion
- Cross-lingual projection of annotations:

  From an annotated English document

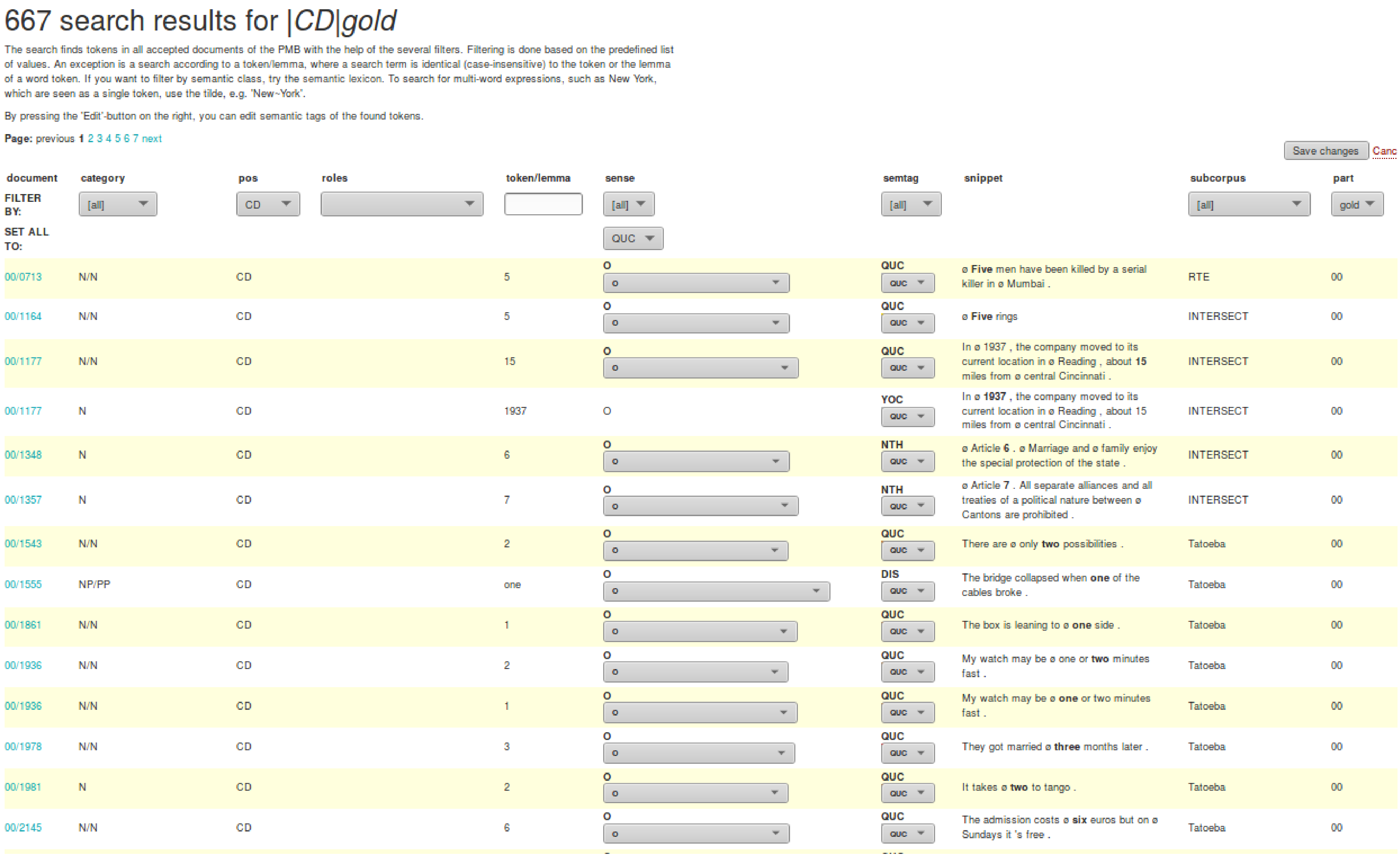
  to the (DE/NL/IT) document

Annotation with Bootstrapping



Adding BoWs
(Bits of Wisdoms)

Adding Bulk of BoWs



667 search results for |*CD*|*gold*

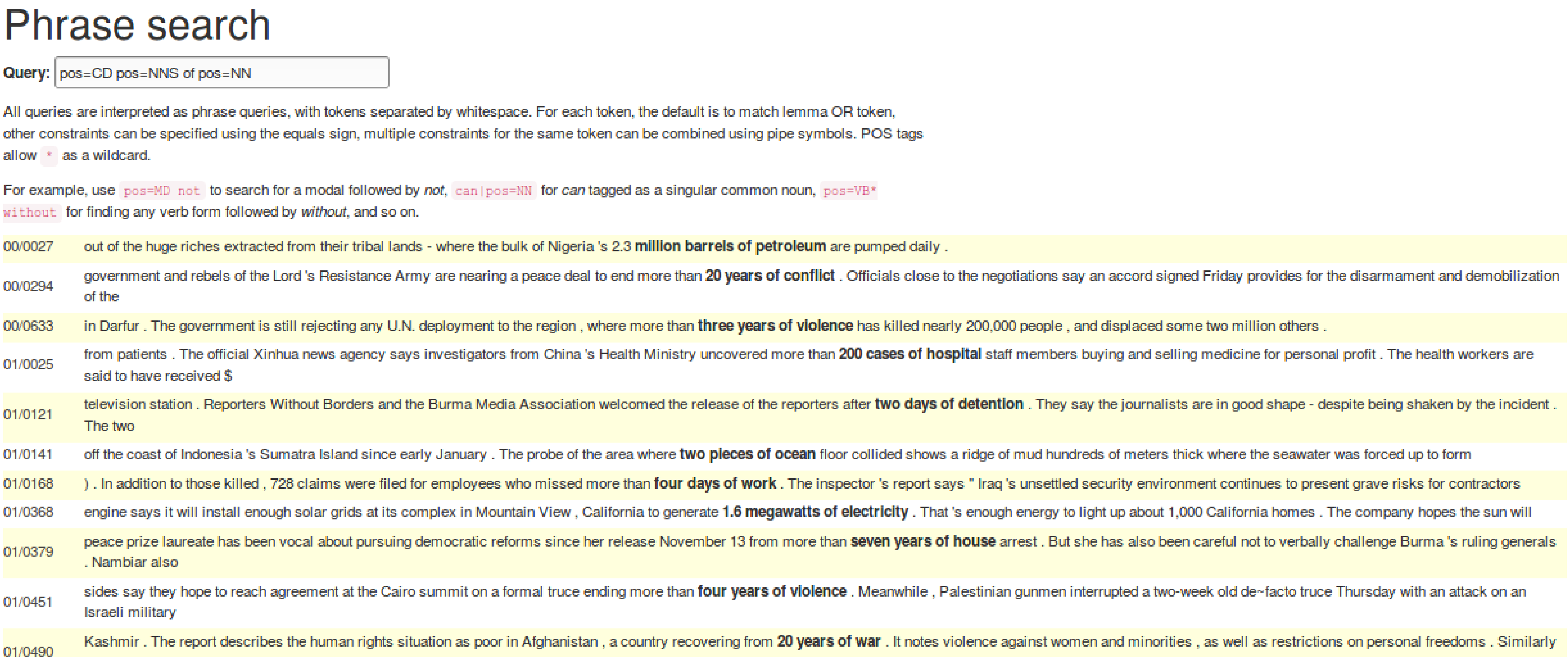
The search finds tokens in all accepted documents of the PMB with the help of the several filters. Filtering is done based on the predefined list of values. An exception is a search according to a token/lemma, where a search term is identical (case-insensitive) to the token or the lemma of a word token. If you want to filter by semantic class, try the semantic lexicon. To search for multi-word expressions, such as New York, which are seen as a single token, use the tilde, e.g. 'New~York'.

By pressing the 'Edit'-button on the right, you can edit semantic tags of the found tokens.

Page: previous **1** 2 3 4 5 6 7 next

document	category	pos	roles	token/lemma	sense	semtag	snippet	subcorpus	part
FILTER BY:	[all]	CD			[all]	[all]		[all]	gold
SET ALL TO:					QUC				
00/0713	N/N	CD		5	**O** o	**QUC** QUC	ø **Five** men have been killed by a serial killer in ø Mumbai .	RTE	00
00/1164	N/N	CD		5	**O** o	**QUC** QUC	ø **Five** rings	INTERSECT	00
00/1177	N/N	CD		15	**O** o	**QUC** QUC	In ø 1937 , the company moved to its current location in ø Reading , about **15** miles from ø central Cincinnati .	INTERSECT	00
00/1177	N	CD		1937	O	**YOC** QUC	In ø **1937** , the company moved to its current location in ø Reading , about 15 miles from ø central Cincinnati .	INTERSECT	00
00/1348	N	CD		6	**O** o	**NTH** QUC	ø Article **6** . ø Marriage and ø family enjoy the special protection of the state .	INTERSECT	00
00/1357	N	CD		7	**O** o	**NTH** QUC	ø Article **7** . All separate alliances and all treaties of a political nature between ø Cantons are prohibited .	INTERSECT	00
00/1543	N/N	CD		2	**O** o	**QUC** QUC	There are ø only **two** possibilities .	Tatoeba	00
00/1555	NP/PP	CD		one	**O** o	**DIS** QUC	The bridge collapsed when **one** of the cables broke .	Tatoeba	00
00/1861	N/N	CD		1	**O** o	**QUC** QUC	The box is leaning to ø **one** side .	Tatoeba	00
00/1936	N/N	CD		2	**O** o	**QUC** QUC	My watch may be ø one or **two** minutes fast .	Tatoeba	00
00/1936	N/N	CD		1	**O** o	**QUC** QUC	My watch may be ø **one** or two minutes fast .	Tatoeba	00
00/1978	N/N	CD		3	**O** o	**QUC** QUC	They got married ø **three** months later .	Tatoeba	00
00/1981	N	CD		2	**O** o	**QUC** QUC	It takes ø **two** to tango .	Tatoeba	00
00/2145	N/N	CD		6	**O** o	**QUC** QUC	The admission costs ø **six** euros but on ø Sundays it 's free .	Tatoeba	00

Phrase Search



Phrase search

Query: pos=CD pos=NNS of pos=NN

All queries are interpreted as phrase queries, with tokens separated by whitespace. For each token, the default is to match lemma OR token, other constraints can be specified using the equals sign, multiple constraints for the same token can be combined using pipe symbols. POS tags allow `*` as a wildcard.

For example, use `pos=MD not` to search for a modal followed by *not*, `can|pos=NN` for *can* tagged as a singular common noun, `pos=VB* without` for finding any verb form followed by *without*, and so on.

00/0027	out of the huge riches extracted from their tribal lands - where the bulk of Nigeria 's 2.3 **million barrels of petroleum** are pumped daily .
00/0294	government and rebels of the Lord 's Resistance Army are nearing a peace deal to end more than **20 years of conflict** . Officials close to the negotiations say an accord signed Friday provides for the disarmament and demobilization of the
00/0633	in Darfur . The government is still rejecting any U.N. deployment to the region , where more than **three years of violence** has killed nearly 200,000 people , and displaced some two million others .
01/0025	from patients . The official Xinhua news agency says investigators from China 's Health Ministry uncovered more than **200 cases of hospital** staff members buying and selling medicine for personal profit . The health workers are said to have received $
01/0121	television station . Reporters Without Borders and the Burma Media Association welcomed the release of the reporters after **two days of detention** . They say the journalists are in good shape - despite being shaken by the incident . The two
01/0141	off the coast of Indonesia 's Sumatra Island since early January . The probe of the area where **two pieces of ocean** floor collided shows a ridge of mud hundreds of meters thick where the seawater was forced up to form
01/0168	) . In addition to those killed , 728 claims were filed for employees who missed more than **four days of work** . The inspector 's report says " Iraq 's unsettled security environment continues to present grave risks for contractors
01/0368	engine says it will install enough solar grids at its complex in Mountain View , California to generate **1.6 megawatts of electricity** . That 's enough energy to light up about 1,000 California homes . The company hopes the sun will
01/0379	peace prize laureate has been vocal about pursuing democratic reforms since her release November 13 from more than **seven years of house** arrest . But she has also been careful not to verbally challenge Burma 's ruling generals . Nambiar also
01/0451	sides say they hope to reach agreement at the Cairo summit on a formal truce ending more than **four years of violence** . Meanwhile , Palestinian gunmen interrupted a two-week old de~facto truce Thursday with an attack on an Israeli military
01/0490	Kashmir . The report describes the human rights situation as poor in Afghanistan , a country recovering from **20 years of war** . It notes violence against women and minorities , as well as restrictions on personal freedoms . Similarly

Despite this adding BOWs manually is associated with a lot of human annotation efforts

Idea of PMBots

- PMBots (prudent monitoring bots) are bots that monitor the annotated documents and with high reliability identify (plausible) annotation errors.
- PMBots use context sensitive rules:

  [sem=QUC; cat=N/N]   [ ]$_X$ [tok=of]   [cat=N]

  ===> [sem=UOM,CON]$_X$

- PMBots can be prudent (verify) or confident (rewrite)