

Richard Eckart de Castilho

Iryna Gurevych

Ubiquitous Knowledge Processing Lab (UKP-TUDA)
Department of Computer Science
Technische Universität Darmstadt
<http://www.ukp.tu-darmstadt.de>



Sabine Bartsch

English Linguistics
Department of Linguistics and Literary Studies
Technische Universität Darmstadt
<http://www.linglit.tu-darmstadt.de>

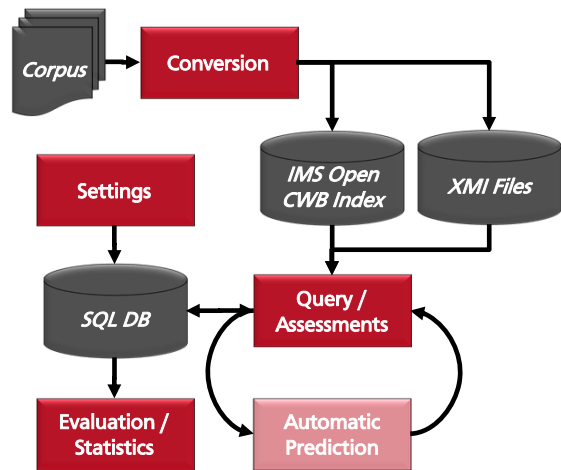
CSNIPER features

- Locate rare constructions in large corpora
- Annotate ambiguous constructions
- Define goals and reach them as a team
- Work concurrently and from different locations
- Employ integrated evaluation and statistics

Non-Canonical Construction examples

- **Canonical:** *[The media] was now [calling] [Reagan] [the frontrunner].*
- **It-cleft:** *It was [Reagan] whom [the media] was now [calling] [the frontrunner].*
- **NP-preposing:** *[Reagan] [the media] was now [calling] [the frontrunner].*

Architecture



Java-based web-application using the **Spring Framework** [1] and **Apache Wicket** [2]. Linguistic preprocessing and corpus conversion using **DKPro Core** [3] and **Apache UIMA** [4]. Linguistic analysis using **TreeTagger** [5] and **Stanford CoreNLP** [6]. Corpus stored in **IMS Open Corpus Workbench** [7] and **UIMA XMI** files. Other data stored in a **MySQL** [8] database using **Hibernate** [9].

- [1] <http://www.springsource.org/>
- [2] <http://wicket.apache.org/>
- [3] <http://code.google.com/p/dkpro-core-asl/>
- [4] <http://uima.apache.org/>
- [5] <http://www.ims.uni-stuttgart.de/projekte/complex/treeTagger/>
- [6] <http://nlp.stanford.edu/software/corenlp.shtml>
- [7] <http://cwb.sourceforge.net/>
- [8] <http://www.mysql.com>
- [9] <http://www.hibernate.org/>

Query

Corpus: Type:

Query Complete

Query history:

Query:

Assessment of the Search Results

Filter: all items not assessed items assessed items items to check

Showing 71 to 80 of 466

Doc	Left	Match	Right	Label	Comment
BOL	It was in the Cup that Northampton were	now to make the headlines .		Correct	
BOL	It was during this season that Chapman made	his first visit to Highbury , the new home of the newly-relegated Arsenal		Correct	
BOL	It was at this time that Chapman made	another of his ' discoveries ' .		Correct	
BOR	It was to Bath that the two Panisocrats walked	first , staying the night at old Mrs Southey 's house , 8 Westgate Buildings , where another visitor , Sarah Fricker , was already in residence .		Correct	
B17	It is for this reason that Interactionists stress	that labelling is crucial to understanding criminal behaviour , as the labelling process publicly identifies individuals as guilty of criminal acts .		Correct	
B1D	It was about this time that Scott found	out that Charles Barry junior had distributed a paper to MPs pressing his own claim to the Foreign Office .		Correct	
B1F	It is like a man who sows	round his garden flowers and aromatic herbs , whenever he looks out , he sees flowers and aromatic herbs and is always in paradise .		Wrong	
B1F	It is like the Buddhist priest who went	to nirvana and when he returned , the world had changed into nirvana .		Wrong	
B1J	It is in this area that religion has	a role to play .		Correct	
B1L	It was at this critical moment that good sense got	the better of the others , Big Pat was calmed down and a semblance of order was restored .		Correct	

(169956 chars) <<< ... parating the two worlds and living in both of them, not at the same time but alternately. Men in the West seem to fear this dichotomy, although it is not a separation but more a reorganisation of life so that the spiritual gains the most important part and the rest of life falls into place beneath.

Evaluation and Statistics

Corpora: BNC, DCIBB, TUEBAD25

Types: It-cleft, PP-inversion, NP-preposing

Users: William Leonard, Deforest Nichelle

Filter: Correct, Wrong, Disputed, Incomplete

Agreement: Fleiss' Kappa: 0.88

Progress: Complete: 750 / 1000

Showing items from collection [BNC] for type [It-cleft, PP-inversion], filtered by [CORRECT, WRONG, DISPUTED, INCOMPLETE].

Type	Collection	Document	Item	#Correct	#Wrong	#Incomplete
It-cleft	BNC	A00	Returning from a visit to Uganda , where he met with patients in the villages and with other agencies ... , Maurice Adams said , ' It is a beautiful country which is being devastated by a disease which can be stopped .	2	0	0
It-cleft	BNC	A02	It is the churches that provide our volunteers ; without their support we would not be able to provide a service at all .	2	0	0
It-cleft	BNC	A03	It is during this time that torture most commonly occurs .	2	0	0
It-cleft	BNC	A04	It was a Romantic notion that there were resonances in the natural world and art ; and it was Schiller ' who argued that music , ' the visual arts and poetry ' always become ... similar in their action on the mind ' .	1	0	1
It-cleft	BNC	A05	It is a work which might well shed a tender light on the novella ' Goodbye , Columbus ' .	0	2	0
It-cleft	BNC	A05	It is a self which is proving , for philosophers , hard to prove .	0	2	0
It-cleft	BNC	A05	It is a war which has begun to embarrass the interveners .	0	2	0
It-cleft	BNC	A05	It is a distinction which may in the end prove more suggestive than serviceable : the author who tells	0	2	0

Type	Collection	Query	Users	Results	Complete	Kappa	Correct	Wrong	TP	FP	UNK	Precision	Estimated Precision
It-cleft	BNC	"It" /VCC/ /NP/ /RC/	2	3734	11.9%	0.85	676	747	236	207	3291	0,0632	0,5327
It-cleft	BNC	"It" /VCC/ /PP/ /RC/	2	449	90.4%	0.90	678	191	324	82	43	0,7216	0,7980
It-cleft	BNC	"It" /VCC/ /pos="NP"/+ /RC/	2	370	25.4%	0.90	267	13	89	5	276	0,2405	0,9468

Annotation

	A	B	C	D	E	F	G	H	I
1	Collection	Document	Begin	End	Instance	Text	Type	Class	Agreement
2	BNC	A00	12998	13208	487	Returning/VVO from/PRP aIATO visit/N1 to/PRP Uganda/NPO/PUN when/AVQ ha/PPV me/I VO with/PRP patients/N2 in/PRP the/ATO villages/N2 and/CJC with/PRP other/AJO agencies/N2/PUN Maurice/NPO Adams/NPO sac/I VO/PUN /PUC/PPNP a/VBZ aIATO beautiful/AJO country/N1 which/DTO a/VBZ being/VBC devastated/VN by/PRP aIATO disease/N1 which/DTO can/VMO be/VPB stopped/VVN/PUN /PUQ	It-cleft	WRONG	1.0
3	BNC	A02	1051	1170	468	It/PPNP a/VBZ twa/ATO churches/N2 that/CT provide/VVB our/DPS volunteers/N2/PUN without/PRP the/DPS support/N1 we/PPNP would/VMO not/XXO be/VPB able/AJO to/TOO provide/VVI a/ATO service/N1 a/PPNP aIATO/PUN	It-cleft	CORRECT	1.0
4	BNC	A03	26716	25773	471	It/PPNP a/VBZ during/PRP the/DTO time/N1 that/CTJ torture/VVB most/AVQ commonly/AVQ occurs/VVZ/PUN	It-cleft	CORRECT	1.0
5	BNC	A05	177193	177278	1	It/PPNP a/VBZ aIATO work/N1 which/DTO might/VMO well/AVQ shed/VVI aIATO tender/AJO light/N1 on/PRP the/ATO novella/N1 /PUQGoodbye/VVI/PUN Columbus/NPO/PUQ/PUN	It-cleft	WRONG	1.0