

**LINGUISTIC LINKED OPEN DATA AS A  
SOURCE FOR TERMINOLOGY-  
QUANTITY VERSUS QUALITY**

SHANSHAN WANG, LAURI CARLSON



UNIVERSITY OF HELSINKI

# Introduction

This paper studies the accessibility and quality of linguistic linked open data (LLOD) for multilingual terminology and localization.

The paper mainly discusses the quality of LLOD data for terminological use, which we are in the process of studying with our ontology-based terminology tool TermFactory, designed to search, merge, verify and improve the quality of linked term data.

We study a large derivative multilingual lexical data sources: BabelNet 3.0 to analyse and evaluate its usefulness for professional special language terminology.



UNIVERSITY OF HELSINKI

# Query Terms

3 domains, 10 terms both in general and specific terms

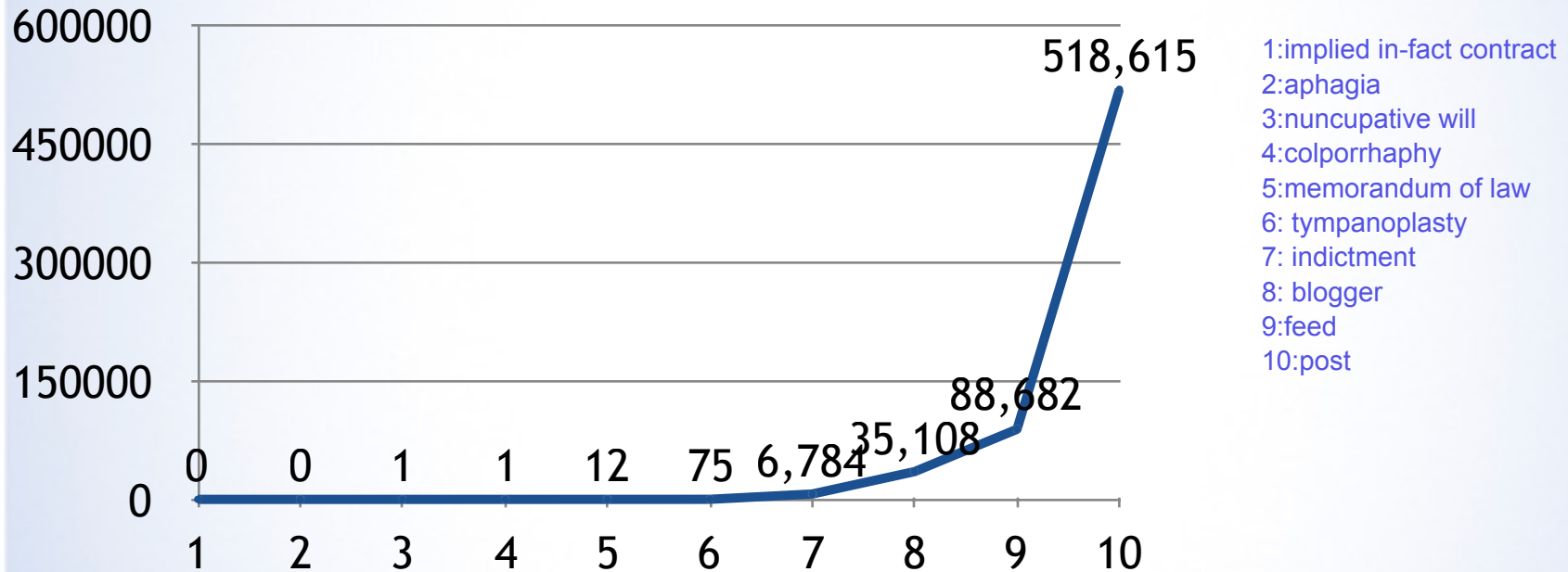
- Medical domain: colporrhaphy , tympanoplasty, aphagia
- Law domain: indictment, implied-in-fact contract, nuncupative will, memorandum of law
- Social network domain: post, blogger, feed

The aim is to test the coverage and reliability of BabelNet data and compare with other sources such as IATE, Eurotermbank, Wiktionary and Termwiki.



UNIVERSITY OF HELSINKI

# Glowbe (Corpus Of Global Web-Based English) Frequencies



UNIVERSITY OF HELSINKI

# Google Book Frequencies

## 2006

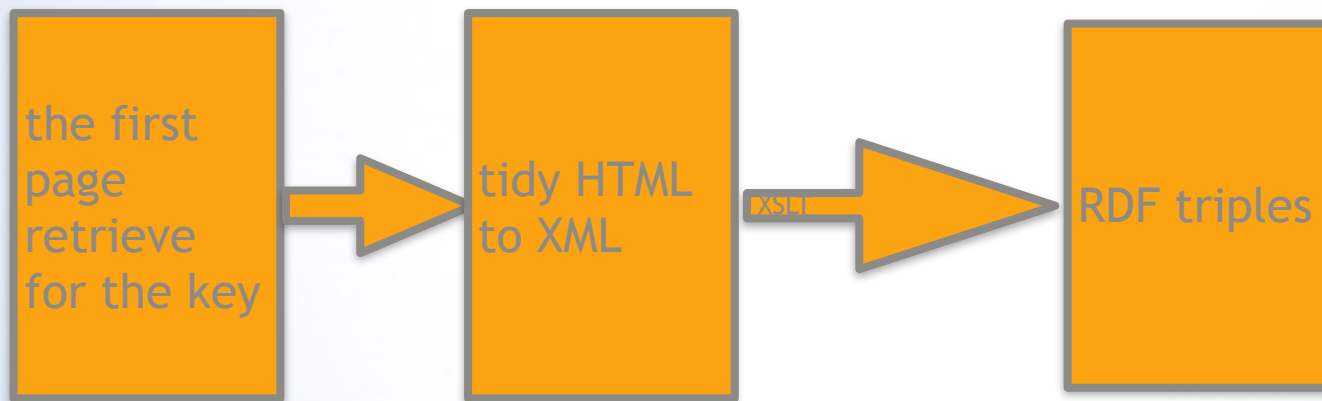
memorandum of law (All)	0.0000031808%
indictment (All)	0.0003728161%
nuncupative will (All)	0.0000007216%
aphagia (All)	0.0000004180%
colporrhaphy (All)	0.0000050880%
tympanoplasty (All)	0.0000066229%
blogger (All)	0.0000540348%
feed (All)	0.0026545580%
<b>post (All)</b>	<b>0.0108030071%</b>



# Methodology

We trawled the sources for the sample keys (specifying the string and the language code) using our TermFactory toolkit

TermFactory toolkit is established by Prof. Lauri Carlson, University of Helsinki



# Query results

Termbase	Triples	language code	Property
BabelNet	15203	12	18
Euroterm bank	472	30	4
IATE	3256	25	11
Termwiki	280	1	5
Wiktionary	1775	69	25



## How Many Entries (Concepts, Synsets, Meanings) Per Termbank And Key?

	Babel Net	ETB	IATE	Term wiki	Wik tionary
memorandum of law	1	0	1	1	1
nuncupative will	1	0	1	1	0
aphagia	1	0	1	1	1
colporrhaphy	1	2	2	0	0
blogger	2	6	1	3	1
indictment	2	11	6	9	3
tympanoplasty	2	1	1	1	0
feed	14	32	10	16	8
post	22	40	7	25	17
implied-in- fact contract	0	0	0	0	0

