

Practical Post-Editing Lexicography with Lexonomy and Sketch Engine

Miloš Jakubíček, Vojtěch Kovář, Michal Měchura, Pavel Rychlý



Brno, Czech Republic

EURALEX 2018, Ljubljana

sketchengine.co.uk

1 Post-editing lexicography

Three steps of utilizing corpora in lexicography
(Rundell's "corpus revolutions")

You've got nothing

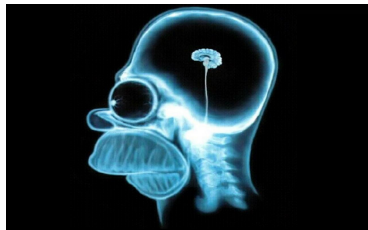


© CC-BY-SA, user OUCHcharley, Flickr

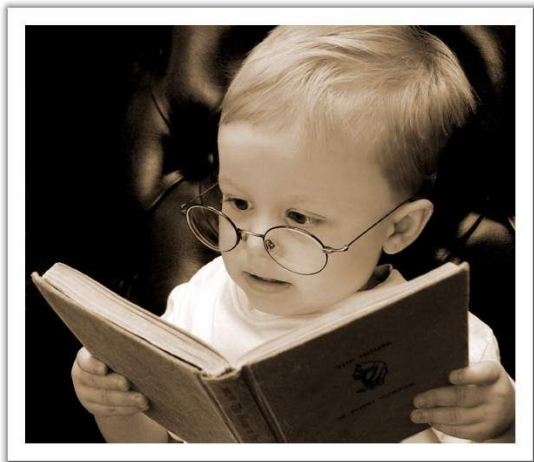
Step 1: Getting a corpus



vs.



You have a corpus



© CC-BY-SA, User Omarfaruquepro, Wikimedia

You have a corpus



© "Here's how many books you can expect to read before you die", mentalfloss.com

Step 2: Extracting data from a corpus



You have a corpus AND data extracted from it

goal

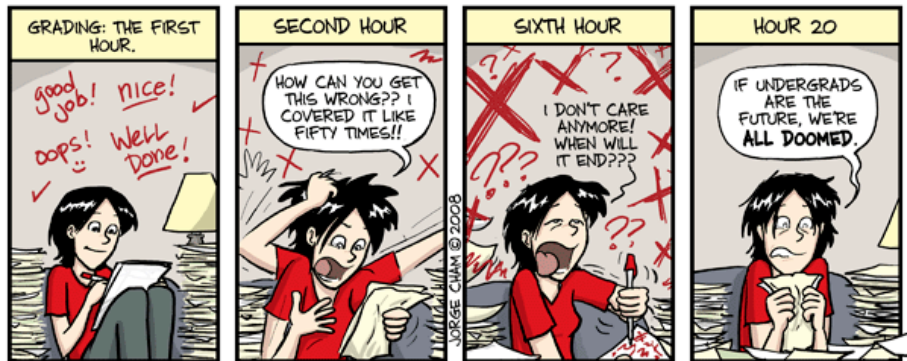
| object of | 58924 | 3.0 | subject of | 25451 | 2.3 |
|------------|----------------------|-------|------------|----------------------|------|
| score | 8390 | 11.18 | score | 903 | 8.45 |
| achieve | 9422 | 9.76 | | | |
| concede | 14 | 9.37 | concede | 204 | 7.5 |
| accomplish | 585 | 7.9 | gape | 76 | 6.5 |
| reach | 1924 | 7.57 | kick | 76 | 5.27 |
| net | 337 | 7.4 | orientate | 34 | 5.03 |
| pursue | 648 | 7.35 | rule | 61 | 5.02 |
| grab | 406 | 7.33 | come | 1316 | 4.96 |
| attain | 400 | 7.32 | cap | 20 | 4.32 |
| pull | 501 | 6.69 | beat | 53 | 4.18 |

choice is the key reform to **achieve** this **goal**, is that s
are going to do the tasks to **achieve** these **goals**. For exa
ous recommendations on how to **achieve** this **goal**. The loc
n Union, and help ensure this work **achieves** its **goals**. To help
ent departments, in a fun environment to **achieve** a **goal** for charit
strong opposition of the old spiritual forces could the **goal** be **achieve**
actors may intend to use IO tools to **achieve** specific **goals**. Recent
winningater environmentally friendly ways of **achieving** target **goals**. In the ci
primaryas of developing countries. **Achieving** these avowed **goals** will rema
seconding â that local solutions are key to **achieving** global **goals** â should
stratollective resources to identify and **achieve** system-wide **goals**. One ad
commie for plugging particular gaps or **achieving** some local **goals**, for getti
ment of an interoperable Federal PKI. To **achieve** the **goal** of an inte
realist realise that he has carried out a task and **achieved** a **goal**. To conf
t. an outcome or a clear confirmation that the learning **goal** was **ach**

Step 3: Automatic dictionary drafting



You have a corpus AND a dictionary draft which you post-edit



WWW.PHDCOMICS.COM

sketchengine.co.uk

Post-editing lessons from MT

- introduced too early with technology not mature enough
 - pushed hard by translation agencies (cost-cutting)
 - processes and methodology initially not adjusted
- ⇒ translators got furious about post-editing and it took long time until it became an established practice

therefore does not mean only automatic dictionary drafting but also

- development of tools that support efficient post-editing
- development of methodology that accounts for post-editing
- teaching and training

Sketch Engine

- corpora for 100 languages
- processing tools, analytic tools
- web platform
- SaaS + open source NoSketch Engine

Lexonomy

- simple & lightweight
- web-based
- SaaS + open source
- designed for easy post-editing
- interconnected with Sketch Engine (push & pull)

The case of Sketch Engine and Lexonomy

- bidirectional data transfer
- push model
 - creating automatic dictionary drafts from Sketch Engine corpora into Lexonomy
 - “One-Click Dictionary”
 - presented at eLex 2017
- pull model
 - accessing Sketch Engine corpora from existing dictionaries in Lexonomy

- headwords: wordlist
- word senses: clustering through word sketches
- collocations: word sketches
- examples: GDEX
- definitions: GDEF
- labels: word sketch highlights
- translations: bidicts from parallel corpora and/or MT



British National Corpus (BNC)



GO

Looking for your previously created dictionaries? Go to [Lexonomy](#) to find them.

ADVANCED

- ▼ Headwords generation
- ▼ Word sense induction
- ▼ Collocation extraction
- ▼ Definition finding
- ▼ Examples finding
- ▼ Thesaurus

[Back to the original interface](#)

- headwords
- collocations
- word senses
- definitions
- examples
- thesaurus



New

ID 316



Save



Cancel

309

<entry>



::<headword>account</headword>

::<partOfSpeech>n</partOfSpeech>

+ ::<sense> a record or narrative description o...

Merge <sense> with next Alt + Shift + Down

+ Child elements

+ Sibling elements

Find examples <example>

Find collocations <collocation>

Find thesaurus items <relatedWord>

Find definitions <definition>

Edit

Configure

Download

Upload

 New

ID 316

 Save

 C

309


 <entry>



 Find examples <example>

 Find collocations <collocation>

 Find thesaurus items <relatedWord>

 Find definitions items <definition>

 Show word sketch...

 Show concordance...

 Show thesaurus...

- improvements of word sense clustering
- currently being tested on a Danish to English and Korean bilingual project with K Dictionaries
- improvements of definition/explanation finding
- maturing the prototype, completing its settings
- problems, bugs, requests → support@sketchengine.co.uk

Conclusions

- post-editing lexicography = third corpus revolution?
- Sketch Engine & Lexonomy ready for experiments
- lot of development ahead (framed under ELEXIS)
- methodology, teaching, training extremely important