

Sketch Engine

Introduction

DAAD 2021 training course

IÜD, Universität Heidelberg

Monolingual search

- Choose BNC; look for “cause”
 - Concordance, change to 50
 - Advanced search > restrict to verbal meanings
 - Word sketch (collocations)
 - Which field indicate the prosody? (go to examples)
 - Check synonyms (thesaurus) for “cause” (Verb)
 - Do synonyms share semantic prosodies? Test:
 - cause ‘affect’, ‘produce’
 - set in; establish;
 - refugee, asylum seeker, migrant
 - soft power; influence
 - politically correct; political correctness
 - Generate word sketch differences
 - which fields are indicative of the prosody?

Construction-based search

- Look forward to VVG
 - “He looks forward to working with you”
- What’s X doing Y:
 - “what's this fly doing in my soup?”
 - Cf. “What are you doing eating cold pizza?”
- X let alone Y
 - “Max won’t eat shrimp, let alone squid.”

Searching a parallel corpus

- Choose Europarl-En aligned with your favorite target language
 - Generate aligned concordances
 - Test native speakers of different languages at the SL
 - Do translation equivalents change? (translationese)
 - Which translation equivalents are used? German corpus: check translation equivalents
- [lemma="mit"] [lemma="alle"] [lemma="Macht" | lemma="Kraft" | lemma="Gewalt" | lemma="Wucht"]
 - Test semantic prosodies for translation equivalents in a Target Language corpus
 - Other resources to convey meanings of semantic prosodies?

Creating your own corpus in Sketch Engine

- From own (downloaded) texts
- From web search
- From TMX

Semantic prosodies

- Translation of semantic prosodies
 - Reading:
 - Hauser, D. J., & Schwarz, N. (2016). Semantic prosody affects valence inferences about ambiguous concepts. *Journal of Experimental Psychology: General*, 145, 882-896 URL: https://www.researchgate.net/publication/299859911_Semantic_Prosody_and_Judgment

Example corpus queries

- Negative vs. positive semantic prosodies
 - ‘regime’ vs. ‘government’; ‘attain’ vs. ‘get’; ‘arise’ vs. ‘evoke’, cause vs. produce; symptomatic, indicative, characteristic
 - Loom large
- Constructions
 - Look* forward to VVG
- Terminology:
 - block exemption; block exemption regulations

Exploring concordances

Do dictionaries tell everything?

<http://www.ldoceonline.com/dictionary/>

<http://dictionary.reference.com/>

Key-word in context (KWIC view): how the word is used

Semantic prosodies

- *Ensure vs. warrant*
- *Regime vs. leadership*
- *Visionary vs. speculative*

Usage preferences:

- *Unforeseen vs. unexpected?;*
- Unknown and absent words (*nail down, gamut, a litany of complaints*)

Exploring collocations

Collocations can be generated as a 'summary' of a concordance

- E.g., which things can be “*strong*”, “*powerful*”, etc.

Testing semantic prosodies

- *Regime, leadership, ensure, warrant, visionary, speculative*

Exploring words' meaning: Nouns, Verbs, Adjectives which go with a word... (e.g., in a window of +/- 5 words)

- *Vengeance* (left sort cc); typical constructions
 - Back with a vengeance...

Semantic prosodies across languages

- Task: test observations for translation with corpus
 - Do translation equivalents share the same semantic prosodies? How often? How this could be addressed?
 - To what extent dictionaries convey semantic prosodies?
 - <https://www.dictionary.com/browse/cause?s=t>
 - <https://dict.leo.org/german-english/cause>
 - <https://www.linguee.com/english-german/search?source=auto&query=cause>

Example task

- Identify a word or sets of words
 - with different semantic prosodies
- Collect evidence of semantic prosodic meaning
 - Create ‘prosodically aware’ dictionary entries
 - What would be the structure to convey meanings?
 - Share with the group

Collocations

Typical, recurrent combinations of words

(J.R.Firth)

day & night

ring & bell

milk & cow

kick & bucket

brush & teeth

Native speakers have strong & widely shared intuitions about such collocations

Collocational knowledge is essential for non-native speakers in order to sound natural
“idiomatic English”

Specialized monolingual corpus

- Log in a new window
 - leave the BNC window open for comparison
- In the new window choose Covid-19 corpus
 - Generate concordance, thesaurus, word sketches, sketch differences
 - Any changes compared to the balance corpus (BNC)?

Investigating linguistic problems with corpus

Words / texts with political bias

- Read the discussion in BBC magazine:
 - <https://www.bbc.com/news/magazine-34061097>
“The battle over the words used to describe migrants”
- Investigate usage differences in a corpus, e.g.
 - [https://app.sketchengine.eu/#sketchdiff?
corpname=preloaded%2Feuoparl7_en&tab=basic
&lemma=refugee&lemma2=migrant&showresults
=1](https://app.sketchengine.eu/#sketchdiff?corpname=preloaded%2Feuoparl7_en&tab=basic&lemma=refugee&lemma2=migrant&showresults=1)

How usage changes with time

- Migrant, refugee, asylum seeker, immigrant, alien
 - Modifiers ‘undocumented’, ‘failed (asylum seeker)’, ‘illegal’, ‘economic’, etc.
- English corpora produced at different times:
 - BNC, Europarl, English Web 2008, English Web 2015
 - Any evidence of the change of attitude?

How these terms are translated?

- Use a parallel corpus (Europarl) and search for each term
 - Which translation equivalents have been used?
 - Do they have similar usage patterns in the target language?

Find other politically charged lexical items or phrases

- In any language:
 - En: ally vs. supporter vs. crony
 - Affront, attack, criticism
- Investigate their usage patterns & translation

Example task

- Find culturally or linguistically marked concepts which may require overt translation, e.g.:
 - Right Honourable Friend
 - Scottish nationalism
 - Loom large; symptomatic; cause; happen...
- Explore parallel and monolingual corpora
 - What are semantic prosodies of these words
 - Which translation strategies have been used

Dictionary vs. corpus

Combining methods: Dictionary research: OED

dubious

adjective

- **1** hesitating or doubting: *I was rather **dubious** about the whole idea*
- **2** not to be relied upon; suspect: *extremely dubious assumptions*
- morally suspect: *timeshare has been brought into disrepute by dubious sales methods*
- of questionable value: *he holds the dubious distinction of being relegated with every club he has played for*

doubtful

adjective

- **1** feeling uncertain about something: *he looked doubtful, but gave a nod I was **doubtful of** my judgement*
- **2** not known with certainty: *the fire was of doubtful origin*
- improbable: *[with clause] it is doubtful whether these schemes have any lasting effect*
- not established as genuine or acceptable: *of doubtful legality*

Dictionary search: Cobuild

dubious	doubtful
<p>1 adj If you describe something as dubious, you mean that you do not consider it to be completely honest, safe, or reliable. (=questionable) This claim seems to us to be rather dubious...</p> <p>2 adj If you are dubious about something, you are not completely sure about it and have not yet made up your mind about it. v-link ADJ, oft ADJ about n (=doubtful) My parents were dubious about it at first but we soon convinced them.</p> <p>3 adj If you say that someone has the dubious honour or the dubious pleasure of doing something, you are indicating that what they are doing is not an honour or pleasure at all, but is, in fact, unpleasant or bad. ADJ n Nagy has the dubious honour of being the first athlete to be banned in this way...</p>	<p>1 adj If it is doubtful that something will happen, it seems unlikely to happen or you are uncertain whether it will happen. usu v-link ADJ, oft it v-link ADJ that/wh For a time it seemed doubtful that he would move at all..., It is doubtful whether Tweed, even with his fluent French, passed for one of the locals...</p> <p>2 adj If you are doubtful about something, you feel unsure or uncertain about it. usu v-link ADJ, oft ADJ about n (=dubious) I was still very doubtful about the chances for success..., Why did he sound so doubtful?</p> <p>3 adj If you say that something is of doubtful quality or value, you mean that it is of low quality or value. usu ADJ n (disapproval) (=dubious) ...selling something that is overpriced or of doubtful quality...</p> <p>4 adj If a sports player is doubtful for a match or event, he or she seems unlikely to play, usually because of injury. (JOURNALISM) oft ADJ for n</p>

Contrastive examples

(Apresyan, 1979)

- He was *doubtful* about the outcome of this project / of the prospects of the rebellion; He was more than ever *doubtful* whether the battery would last him home.
- I am *dubious* about his stories of early success; I am *dubious* of his honesty
- (1) 'no knowledge about whether a statement is true'
 - (2) 'tend to think that it is false, based on *information* they have' → *doubtful*
 - (2) 'tend to think that it is false, based on *suspensions* or *concerns*' → *dubious*
- I am *dubious* / *doubtful* / about what I ought to do:
- 'No difference: cannot reasonably be suspicious of yourself'

Contrastive examples: predicted difference ?

- John was *doubtful* / *dubious* about his ability to cope with situation
- his \neq John's; his = someone else's ability (non-co-referential reading)
- *doubtful* = based on information;
- *dubious* = based on suspicions or concerns
- his = John's ability (co-referential reading)
- *dubious* = *doubtful*; (based on information)

Collocation-based research

dubious		doubtful	
dubious distinction	dubious privilege	doubtful whether	doubtful proposition
dubious honour	dubious moral	doubtful :	doubtful after
dubious about	dubious benefit	doubtful that	doubtful as_to
dubious pleasure	dubious proposition	doubtful if	doubtful because
dubious practice	dubious morality	doubtful starter	doubtful case
dubious provenance	dubious character	doubtful .	doubtful origin
dubious penalty	dubious decision	doubtful for	doubtful privilege
dubious claim	dubious nature	doubtful about	doubtful value
dubious quality	dubious assumption	doubtful debt	doubtful quality
dubious reputation	dubious accolade	doubtful validity	doubtful relevance
dubious legality	dubious reliability	doubtful stayer	doubtful loan
dubious taste	dubious tactic	doubtful authenticity	doubtful efficacy
dubious circumstance	dubious <u>honor</u>	doubtful legality	doubtful expression
dubious accounting	dubious source	doubtful runner	doubtful ,
dubious value	dubious achievement	doubtful with	doubtful assumption

Ideas for further investigation

- *Remember, recollect, recall*
- *Gather vs. collect*
- *Victim vs. prey*
- *Grateful vs. thankful*
- *Stay vs. remain*
- *Hard vs. difficult*
- *Honest vs. upright vs. (righteous)*
- *Lose yourself in a great story. vs. Tell him to get lost!*
- Lexicographic portraits of words (Apresyan, 1986/1990)
- coherent information about meaning & grammar
- prosodic, communicative and pragmatic properties

Parallel concordances

Aligned segments → concordance search

Terminological search (more limited domains)

artificial intelligence,

Member states,

Vengeance, realism... (in games)

Parallel concordances: investigating translation equivalents

- DE: Mit aller Kraft / Gewalt / Macht / Wucht
- collocation with verbs:
 - versuchen (to try), sich wehren gegen (to resist), vermeiden (to avoid)
- + knowledge of the agent needed: making explicit whose effort has been expended
- Translations:
 - ... *The nurse had **to force** her physically to get up*
 - ... *She tried **with all** her strength to keep a clear head*
 - ... *They were trying **as hard as** they could to escape...*
 - ... *She beats the chair against the door **with all** her might*

Parallel concordances

- Lexical priming: Associating words with domains & contexts via use (*an open and shut case; shut up* → friends vs. committee meeting)
- Es: *hora* vs. *a la hora*
- En: *ride a bus, etc....*
- En: *as a consequence* vs *consequence*.
- Semantic prosodies: translation of the effect not the means
- *symptomatic of ...*
- *loom large*