

Compiling and analysing journalistic and academic texts

Credibility, Honesty, Ethics, and Politeness in Academic and Journalistic Writing
(CHEP 2018)

Summer school, Split July 2018

Jessica Dheskali & Ivana Mitić

Table of content

1. Introduction
2. Corpus compilation
 - 2.1. General aspects
 - 2.2. Compiling academic texts
 - 2.3. Compiling journalistic texts
3. Analysing journalistic texts: a case study
4. Examples
5. Analysing academic texts
6. Bibliography

1. Introduction



2. Corpus compilation

2.1. General aspects

What is a corpus?

corpus (pl: **corpora**):

- a collection of (1) *machine-readable* (2) *authentic* texts (incl. transcripts of spoken data or signs) which is (3) *sampled* to be (4) *representative* of a particular language or language variety
- is different from a random collection of texts or an archive

(cf. McEnery & Wilson 2001)

2. Corpus compilation

2.1. General aspects

What is corpus linguistics?

- the study of language in use through corpora
- it serves to answer fundamental research questions such as:
 1. What particular patterns are associated with certain lexical or grammatical features?
 2. How do these patterns differ within varieties, genres and registers?

(Bennet 2010: 2)

2. Corpus compilation

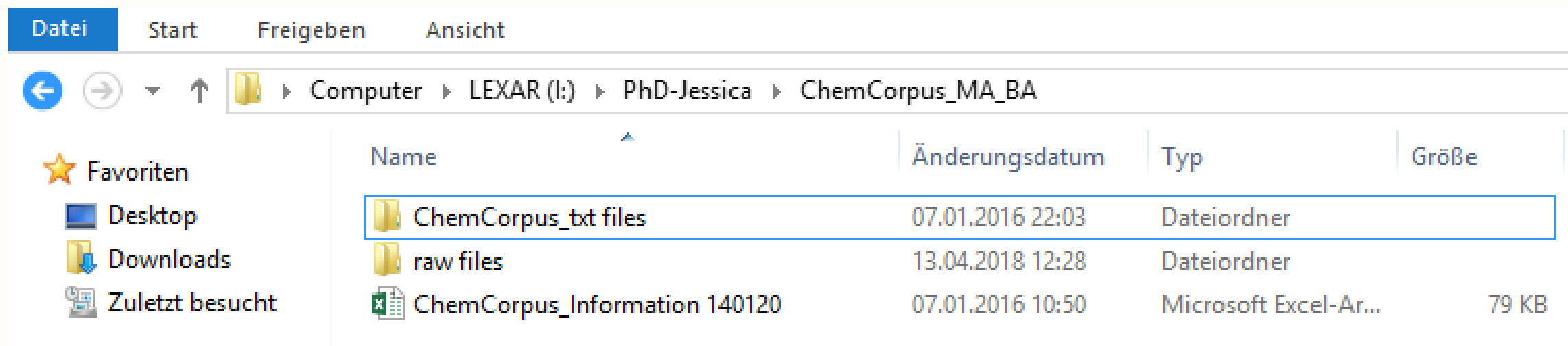
2.1. General aspects

Corpus design & compilation, managing files

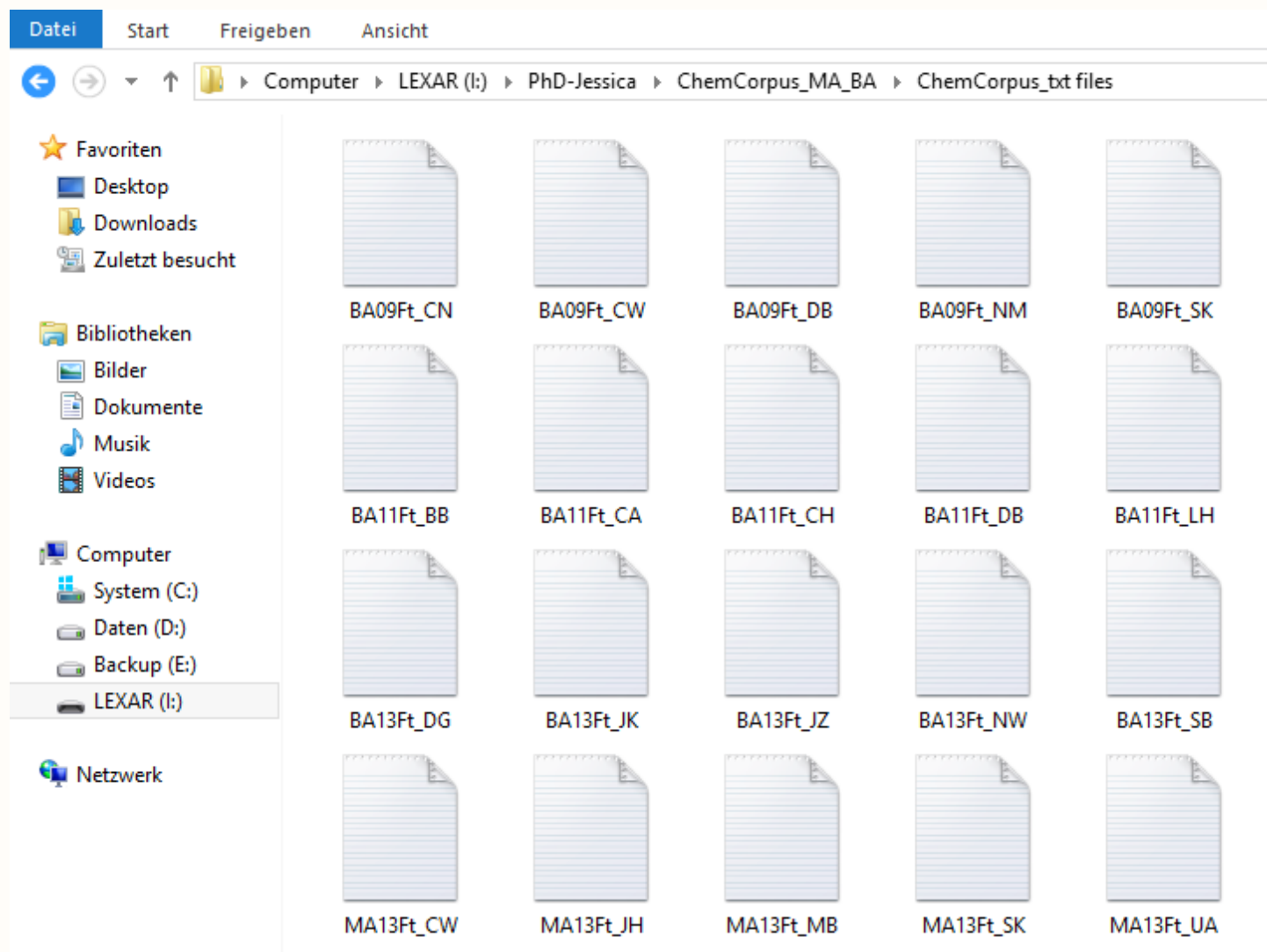
- (1) Decide on a website/platform, journal, university, magazine,... !
- (2) Download as many texts as possible or necessary and save them in a folder named “raw texts”!
- (3) Convert all files into text files and clean them!
- (4) Rename and annotate your files!
- (5) Save all the extra mark-up information about the texts in an excel sheet!

2. Corpus compilation

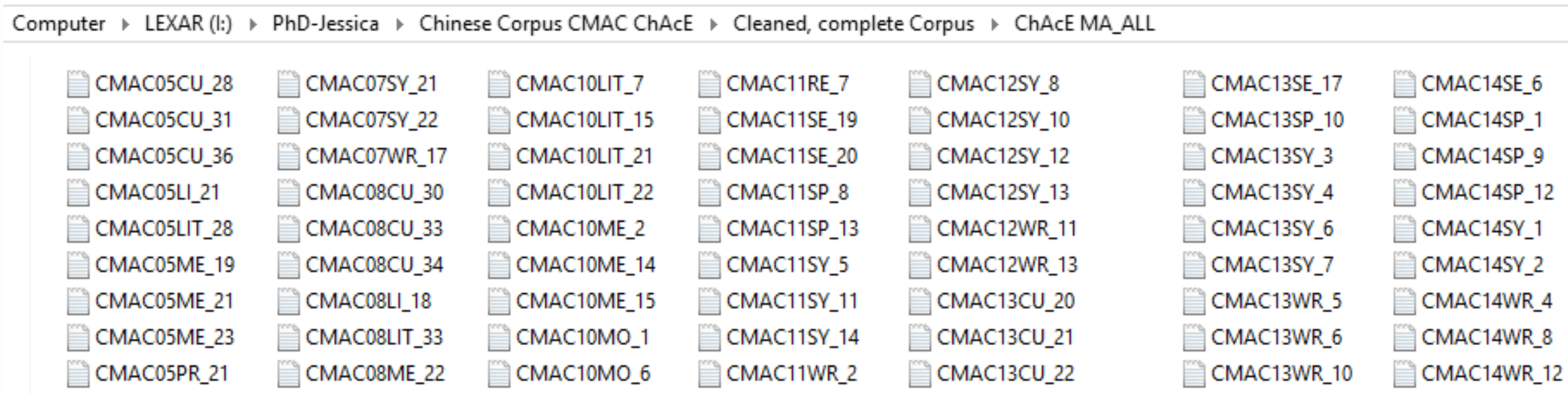
2.1. General aspects



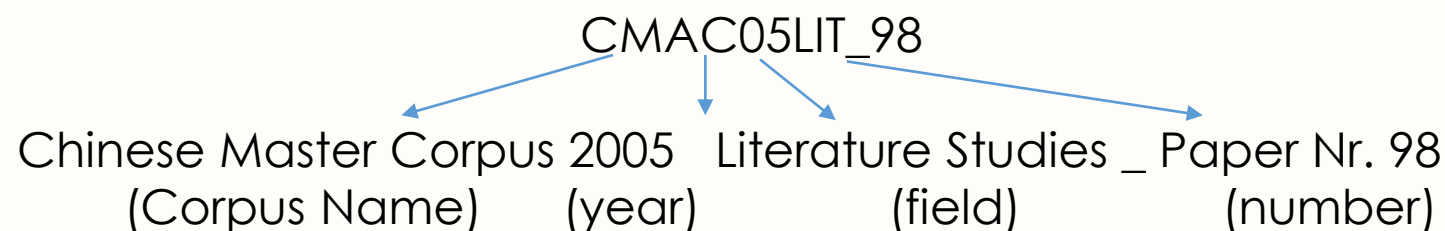
Picture 1: The compilation of the ChemCorpus



Picture 2: Text files included in the ChemCorpus



Picture 3: Text files included in the ChAcE Corpus



additional: male/female (gender), BA/MA/PhD (level), thesis/term paper/report (text type), home country...

2. Corpus compilation

2.1. General aspects

What is representativeness?

- “A corpus is thought to be representative of the language variety it is supposed to represent if the findings based on its contents can be generalized to the said language variety.” (Leech 1991: 27)
- refers to the extent to which a sample includes the full range of variability in a population. (cf. Biber 1993)
- is a defining feature of a corpus (language is infinite, but a corpus has to be finite in size; proportionally include a wide range of text types to ensure maximum balance and representativeness)

2. Corpus compilation

2.1. General aspects

Representativeness is closely related to your research questions:

- for a corpus representative of general English, a corpus of newspapers will not be enough
- for a corpus representative of newspapers, a corpus representative of *The Times* will not be enough

2.2. Compiling academic texts

- collecting (parts of) theses/term paper/project reports from your/a university by asking students and teachers
- downloading theses etc. from online data bases
- downloading research articles from open access online journals
- downloading abstracts online
- downloading e-books
- ...

2.3. Compiling journalistic texts

1. corpus of the whole texts (one or more – connection with the topics),
 2. corpus of headlines,
 3. corpus of headlines and frames,
 4. corpus of frames
- Analyzing: connection between the type of compilation and the main analysis

- Compiling a journalistic corpus
- Must be sizeable so that it can be defined quantitatively – the exact number of excerpts and source(s) must be indicated
- Analyzing
- Connection between the corpus, topics, type of research questions, and analysis

3. Analysing journalistic texts

- Selecting a specific approach of analysis is crucial – relying on relevant and credible sources is essential.
- **Textual and linguistic analysis** (connection between structures and rhetorical devices, for example see Bonyadi and Samuel 2013, Botella et al. 2015)
- **Relevance theory** (Dor 2003 investigated communicative function of headlines and claimed that „headlines are designed to optimize the relevance of their stories for their readers: Headlines provide the readers with the optimal ratio between contextual effect and processing effort, and direct readers to construct the optimal context for interpretation.“)
- **Critical discourse analysis** (language is a social and practical construct which is characterised by a symbiotic relationship with society, see Dubrovskaya et al. 2015, Mohammedwesam 2017; Mohammedwesam 2017 examined the case of the Gaza war of 2008–2009 in two British newspapers (The Guardian, The Times London) and two American newspapers (The New York Tim and The Washington Post) and claimed „that analysis of discourse practices was crucial to illuminate the representation patterns of the social groups. This study reveals that the supposed Palestinian danger and threat to Israel was prevalent across the four selected newspapers“).

- Identification of context in which we examine the problem
- Explicit discourse analysis requires partial analysis of some of the relevant dimensions of the context, such the:
 - *setting* (time, place),
 - *participants* (their identities, roles and relations),
 - the ongoing communicative or social *Action*,
 - activity type or social practice,
 - the *Goals* of the action,
 - as well as the (mutual) *knowledge* (common ground),
 - and even the *attitudes* or *ideologies* of the participants. (Van Dijk, Journal submission guide)
- Discourse processing
- The aim is not to summarize, paraphrase or repeat pieces of discourse BUT *analyse* them.

3. Analysing journalistic texts: a case study

Bonyadi and Samuel 2013

- The topic: **Headlines in Newspaper Editorials: A Contrastive Study**
- Aim: exploring the kind of textual and rhetorical strategies the two newspapers used for propagating their preferred ideologies.
- Corpus: Headlines (40) from the editorials of the English newspaper, The New York Times (NYT), and those of Persian newspaper, Tehran Times (TT).
- Results:
 - 1) Different structures of headlines: in terms of verbal/ nonverbal distinction, in *TT* the headlines mostly in a form of full sentences, while those of *NYT* the headlines in short phrases (see Bonyadi and Samule 2013: 8).
 - 2) Similarities in using presuppositions: useness of existential presupposition and lexical presupposition for persuasion purposes.
 - 3) Different rhetorical devices: in *TT* used literary devices such as allusion, neologism, and irony which made the headlines look persuasive, while the writers in *NYT*, preferred another set of devices such as pun, testimonial, and quoting out-of-context (see Bonyadi and Samuel 2013).

Bonayadi and Samuel 2013 – how to analyse headlines

- Conclusions: functions of headlines is not only to inform about topic of the editorial. Editor writers use rhetorical devices to expresses ideology of the papers which can influence the readers' understanding of the editorial text.
- What is important for analysis of headlines?
- Focus on the textual analysis of headlines by trying to identify presuppositions and rhetorical devices.
- Pay attention on how headlines are used to manipulate the readers' understanding of the news events.

A possible new subject for investigation

- The subject of this analysis is the politically and socially important topic of “gifting Niš airport” as it has been presented in Serbian online media.
- Important to know: Some background information related to the explored subject seems important. The Administration of the City of Niš has decided to transfer the ownership of Niš airport *Constantine the Great* to the Republic of Serbia.

What could be the source for the new corpus?

Aspects to consider:

- Which criteria can be used for compiling the corpus?
- State media and/or local media
- Which kind of material? Online, or not?
- Who will be the sender of the news? Who will be the receiver?

An example of a corpus

- The corpus is based on an analysis of headlines conducted from March 30th to April 30th , and June 23th
- The headlines used in the analysis were excerpted from online newspapers (local media newspapers, such as *Južne vesti*, *Niške novine*, and state newspapers, such as *Večernje novosti*, *Politika*, *Danas*) and from news published on local television channels, such as *Zona plus* and from the public media service *Radio Television of Serbia* (RTS).
- The criteria based on which the headlines were excerpted from daily newspapers are the importance of the topic and the possibility of changes on a daily basis.

- The primary focus in this research can be the degree of credibility whereby the headlines have been analysed from two perspectives:
 - 1) who is behind them (the Republic of Serbia as the state, the City of Niš, the president and the minister who present the power of the state) and 2) who is the transporter (local media and state media).
- „Media credibility can be defined as perceptions of a news channel’s believability, as distinct from individual sources, media organizations, or the content of the news itself“(Busy 2003: 248).
- Source credibility focuses on characteristics of message senders or individual speakers, such as trustworthiness and expertise“ (Busy 2003: 248).

4. Examples

State media or local media?

[1] *Mihajlović: Niš nema novca za aerodrom* / 'Mihajlović: Niš doesn't have money for the airport.'

[2] *Mihajlović: Nastavićemo da ulažemo u niški aerodrom* / 'Mihajlović: We will continue to invest in Niš airport.'

From state media or from local media?

[3] *Ministarka Mihajlović: Niški aerodrom je bio i ostaće državni /*
'Minister Mihajlović: Niš airport has been and will be state ownership.'

[4] *MIHAJLOVIĆ: DRŽAVA ĆE RAZVIJATI NIŠKI AERODROM! /*
'MIHAJLOVIĆ: THE STATE WILL DEVELOP NIŠ AIRPORT!'

Possibilities for analyses

- Type of speech?
- Identification of the the sender of information? Who is the sender, receiver?
- Type of contexts?
- Can you recognized the effect of credibility?
- Type of discursive strategies and effect of credibility?
- Function of headlines and effect of credibility?
- Headlines and structures and effect of credibility?

Examples

- Do you have some examples from your country? – TOPIC (example)
- Which kind of newspapers and media? – SOURCE (example)
- Effect of credibility, honesty, ethics – ways of exploring?
- How to create a corpus?

5. Analysing academic texts

S 1: “It is true that many problems and issues regarding the Balkan have been shrugged off and overlooked, not just by European leaders but also by the leaders of the Balkan countries; as they were not viewed as something of great importance, or as something that urgently needs to be fixed. Either because the leaders of the countries were afraid and reluctant to take any kind of action, or just because there were seemingly more relevant problems to be dealt with.” (SS17DOFT_eb)

S 2: “To get more clarification and support for this issue, in the following part, I would like to discuss the danger of cultural identity in China.” (CMAC05CU_28)

S 3: In Plagâ’s (1999) analysis, phonological repair plays a major role, and phonologically conditioned gaps in this suffixation are not clearly explained. He also claims that -ify, another verbal suffix, and -ize can be regarded as phonologically conditioned allomorphs (Plag, 1999, p. 195).” (CMAC08PH_20)

5. Analysing academic texts

S 1: “**It is true that** many problems and issues regarding the Balkan have been shrugged off and overlooked, not just by European leaders but also by the leaders of the Balkan countries; as they were not viewed as something of **great** importance, or as something that **urgently** needs to be fixed. Either because the leaders of the countries were afraid and reluctant to take any kind of action, or just because there were **seemingly more relevant** problems to be dealt with.” (SS17DOFT_eb)

S 2: “To get more clarification and support for this issue, in the following part, **I would like to** discuss the **danger** of cultural identity in China.” (CMAC05CU_28)

S 3: In Plagâ’s (1999) analysis, phonological repair plays a **major** role, and phonologically conditioned gaps in this suffixation are not **clearly** explained. He also **claims** that -ify, another verbal suffix, and -ize can be regarded as phonologically conditioned allomorphs (Plag, 1999, p. 195).” (CMAC08PH_20)

5. Analysing academic texts

Lexical analysis:

SEMANTIC PROSODY

- is the discourse function of the word: it describes the speaker's communicative purpose
- writers can use it to convey negative or positive evaluation without stating their views explicitly (cf. Lindquist 2009: 57)
- “When the usage of a word gives an impression of an attitudinal or pragmatic meaning, this is called a *semantic prosody*.” (Sinclair 1999)
- others (e.g. Channell 1999) call it ‘evaluative lexis’

The following words have been claimed to have a negative (‘unpleasant’, or ‘unfavorable’) semantic prosody. You could investigate whether this is true and identify near-synonyms (a term whose meaning is similar, but not identical):

happen, commit, set in, cause, peddle, dealings

(Sinclair 1999)

Do you know other examples of words which have hidden meanings?

5. Analysing academic texts

- authorial identity
 - engagement markers
 - attitude markers
 - sentence structure
 - hedges & booster
 - article usage
 - Connectors
 - modality
 - ...
- cultural differences
 - gender studies
 - native vs. non-native texts
 - differences across disciplines
 - differences across academic levels (BA, MA, PhD)
 - differences across genres (research article, thesis , handbook article, review, conference presentation ...)
 - ...

Defining research questions:

Basic criteria for a research question:

- It must be possible to answer it (within a reasonable time span)
- It must be specific

In choosing your topic and defining your research question, it may be useful to ask yourself the following questions:

- (1) Will it be possible to answer my question(s) with the help of my corpus material?
- (2) Will it be possible to work through the project in the limited time that you have available?
- (3) How could you classify your examples?
- (4) Do you know what else has been written on the topic (library, google scholar, previous theses,...)?

Examples:

1. **What factors determine** the placement of nouns and noun phrases as direct objects of phrasal verbs?
2. **What are the major differences** between the usage of ... in Chinese theses?
3. **In how far do** the usages of the two modals *may* and *will* **differ** in Chinese and German student writings?
4. Do German L2 learners of English set themselves **more often** as the agent in their writings when expressing (deontic) volition?
5. **How** have the connotations of *nerd* and *geek* **changed over time**?

Task: How would you analyze these examples?

[1] *Protest “Ne damo niški aerodrom” danas u centru Niša (Južne vesti)* / 'The protest “We will not give away Niš airport” today in the centre of Niš.'

[2] *NIŠLIJE „GLASALE” ZA AERODROM (Zona plus)* / 'THE CITIZENS OF NIŠ “VOTED” FOR THE AIRPORT.'

[3] *Нишки аеродром постаје државни, грађани протестују (RTS)* / 'Niš airport is turning into state ownership, the citizens are protesting.'

[4] *Niš: Protest zbog "Konstantina", leteli avioni od papira... (B 92)* / 'Niš: A protest because of “Constantine”, paper planes were flying.'

[5] *PROTEST ZBOG "KONSTANTINA VELIKOG" Nišlije ne daju da aerodrom bude u vlasništvu države (Blic)* / 'A protest because of “CONSTANTINE THE GREAT” The citizens of Niš refuse to give the airport to the state.'

6. Bibliography

- Bennett, G.R. (2010). *Using corpora in the language learning classroom: corpus linguistics for teachers*. Ann Arbor, MI: University of Michigan Press.
- Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, 8(4), 243–257.
- Botella, A, Stuart, K, Gadea, L. (2015). A journalistic corpus: a methodology for the analysis of the financial crisis in Spain. *Procedia - Social and Behavioral Sciences* 198: 42–51.
- Bucy, E. (2003). Media credibility reconsidered: Synergy effects between on-air and online news. *Journalism & Mass Communication Quarterly* 80(2): 247-264.
- Dubrovskaya T, Dankova N, Gulyaykina S. (2015). Judicial power in Russian print media: Strategies of representation. *Discourse & Communication* 9(3): 293-312.
- Dor, D. (2003). On newspaper headlines as relevance optimizers. *Journal of Pragmatics* 35: 695-721.
- Elorza, I. (2009/2010). Positioning readers in newspaper discourse: A contrastive case study. *BISAL* 4: 1–22.
- Leech, G. (1991). The state of art in corpus linguistics. *English Corpus Linguistics: Linguistic Studies in Honour of Jan Svartvik*, London: Longman, 8-29.

6. Bibliography

Lindquist, H. (2009). *Corpus linguistics and the description of English*. Edinburgh: Edinburgh University Press.

McEnery & Wilson 2001

McCarthy, M. (2004). *Touchstone: From corpus to course book*. Cambridge: Cambridge University Press.

Mohammedwesam, A. (2017). Critical discourse analysis of war reporting in the international press: the case of the Gaza war of 2008-2009. *Palgrave Communications* 3: 1–11.

Semetko, A, Valkenburg, P. (2000). Framing European politics: A content analysis of press and television news. *Journal of Communication* 50(2): 93-109.

Pollak, S, Coesemans, R. Daelemans, W, and Lavra, N. (2011). Detecting contrast patterns in newspaper articles by combining discourse analysis and text mining. *Pragmatics* 21(4): 647-683.

Sinclair, J. 1999. *Concordance tasks*. Retrieved from <http://www.twc.it/happen.html>. Accessed on May 20, 2018. The Tuscan Word Centre.

Zhang, W. (2015). Discourse of resistance: Articulations of national cultural identity in media discourse on the 2008 Wenchuan earthquake in China. *Discourse & Communication* 9(3): 355-370.