**Corpus stylistics**


- a finite-sized body of machine-readable texts, sampled in order to be maximally representative of a particular language or the language variety under consideration” (McEnery and Wilson 2001: 32).

- REPRESENTATIVENESS => corpus compilation = principled way

---

**Corpus stylistics**

- corpus stylistics
  - frequency + concordances of a search word
  - patterns of repetition & patterns of co-occurrence among lexical items

- linguistic data: empirical (Biber et al. 1998)

---

**Corpus stylistics**

- corpus stylistics
  - linguistic data: empirical (Biber et al. 1998)

---

**Corpus stylistics**

- corpus stylistics
  - linguistic data: empirical (Biber et al. 1998)

---

**Corpus stylistics**

- corpus stylistics
  - linguistic data: empirical (Biber et al. 1998)

---

**Corpus stylistics**

- corpus stylistics
  - linguistic data: empirical (Biber et al. 1998)
วัจนิยาม

วัจนิยามของ corpus styliks (corpus stylistics)

Form – Meaning Relations
- repeated patterns of co-occurrences
  - predominant language uses
  - NORMATIVE USAGE OF LANGUAGE
    - รูปแบบ - ความหมายที่เป็นมาตรฐานในการใช้ภาษา

Form – Meaning Relations
- ภาษาของตัวบทที่มีความหมายซ้ำๆ (meanings) และผลกระทบ (effects) ต่อผู้อ่าน/ผู้ฟังได้อย่างไร
  - LINGUISTIC DEVIATION IN TEXT

Linguistic Norm

Linguistic Deviation

วัจนิยาม:

Linguistic Norm

Linguistic Deviation

Linguistic Norm

“[...] a text is a selection from the potential of the language [...]. Comparative corpus methods [...] allow us to study how far texts consist of recurrent phrasal patterns which are widespread in the language as a whole.” (Stubbs 2005: 21)
Corpus stylistics

- Corpus-based
  - ตรวจสอบ intuition
  - RQ: yes-no question
  - e.g. Hidalgo-Downing (2001), Louw (1993)

- Corpus-driven
  - คำนวณลักษณะการทำงานที่เป็นลักษณะแสวงหาของวัฒนา
  - RQ: “what” question
  - e.g. Stubbs (2005), Mahlberg (2007)

Corpus-based

- Louw (1993)’s analysis of Philip Larkin’s *First Sight*

Lambs that learn to walk in snow
When their bleating clouds the air
Meet a vast unwelcome, know
Nothing but a sunless glare.
Newly stumbling to and fro
All they find outside the fold,
Is a wretched width of cold.

As they wait beside the ewe,
Her fleeces wetly caked, there lies
Hidden round them, waiting too,
Earth’s immeasurable surprise.
They could not grasp it if they knew,
What so soon will wake and grow
Utterly unlike the snow.

Concordances of “utterly”

Corpus-driven approach

- หลักการ: การเปรียบเทียบตัวกับคลังข้อมูล

- วิธีวิจัย
  - การเลือกใช้คลังข้อมูลที่เหมาะสมเป็น norm:
    - “contextually related” (Enkvist 1976)
  - Software: AntConc, WordSmith

- Descriptive tools
  - คำสำคัญ (Key words)
  - Co-occurrence patterns
**Descriptive tool 1:**

**Definition**
- lexical items of significance to a text in frequency question, because of their “unusual in comparison with a reference corpus of some suitable kind”.
  
  (Scott and Tribble 2006: 55)

- Positive and negative keywords
- NB: definition of “style marker”
  - “linguistic features whose densities in a text are appreciably different from those found in its contextually related norm” (Enkvist 1973: 21).

**Keyword list**
- a comparison involved;
- a statistical measure of saliency of lexical items
- the starting point to discover “style markers” of a text
- affected by the choice of reference corpora.
- multiple reference corpora strongly recommended

---

**Descriptive tool 2:**

**Patterns of co-occurrences**

**“Co-selection”** (cf. Sinclair 1991)
- Word choice not an entirely independent selection but depends in part upon other words that the author has (not) chosen to use
- meanings tend to be expressed not through a single word but through a lexical item and its co-textual features.

**The relation between pattern and meaning**
- our interpretation of a text is influenced by patterns of lexical items that occur in the text.

**Categories of lexical co-occurrence patterns**

**Collocation:**
- a + nice + house (cf. Sinclair 1991: 20)

**Colligation:**
- adjective + noun, adjective + “house”

**Semantic preference:**
- “naked eye” + visibility

**Semantic prosody**
- evaluative (cf. Louw 1993), e.g. “utterly” = negative
- discourse function (cf. Sinclair 2004), e.g. “naked eye” = difficulty

**Semi-fixed phrase**
- “one of the world’s” + superlative adj. + plural noun
Methodology

- Data preparation
  - EMDI (21,412 tokens): Emma’s direct speech
- Reference corpora (NB: corpus creation criteria)
  - OTHERS: other characters’ speech
  - NOVDI: fictional direct speech during 1775-1820
  - JADI5: direct speech from the other 5 novels by Austen
  - LModE: late modern English (1780-1850)
- Corpus software
  - WordSmith Version 3 (Scott 1999)

EMDI vs. OTHERS

- Keyword extraction (cf. handout)
  - Setting:
    - Minimum frequency = 10 occurrences
    - Log-likelihood
    - P value ≤ 0.05
- Categorisation of the keywords
  - proper nouns
  - content words
  - function words

Lexical keywords (1)

- Semantic groups:
  2. Thoughts and feelings:
     - thinking: “idea”
     - feeling: “feelings”, “attachment”, “love”

Lexical keywords (2): concordance analysis

- “marry” + “marrying” (40 occurrences)
- Who + “marry”/ “marrying”?

<table>
<thead>
<tr>
<th>Character</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mr Elton</td>
<td>2</td>
<td>5%</td>
</tr>
<tr>
<td>Emma</td>
<td>4</td>
<td>10%</td>
</tr>
<tr>
<td>Harriet</td>
<td>10</td>
<td>25%</td>
</tr>
<tr>
<td>Mr Knightley</td>
<td>12</td>
<td>30%</td>
</tr>
<tr>
<td>Mr. Martin</td>
<td>2</td>
<td>5%</td>
</tr>
<tr>
<td>Mr. Weston</td>
<td>3</td>
<td>7.5%</td>
</tr>
<tr>
<td>generic</td>
<td>7</td>
<td>17.5%</td>
</tr>
<tr>
<td>Total</td>
<td>40</td>
<td>100%</td>
</tr>
</tbody>
</table>
Lexical keywords (3): Mr. Knightley and “marry(ing)”

- Mr. Knightley must not marry!
- I cannot at all consent to Mr. Knightley’s marrying; and I am sure it is not likely.
- He has no occasion to marry, either to fill up his time or his heart.
- But Mr. Knightley does not want to marry.
- I have not the smallest wish for your marrying Jane Fairfax or Jane anybody.

Lexical keywords (4): Emma and “marry(ing)”

- If I thought I would be like Miss Bates, [...] I would marry tomorrow.
- If I were to marry, I must expect to repent it.
- My being charming, Harriet, is not quite enough to induce me to marry;
- I have none of the usual inducements of women to marry.
- And I am not only, not going to be married, at present, but have very little intention of ever marrying at all.”

Lexical keywords (4): love, attachment, feelings

“feelings” in EMDI

have been giving way to any feelings ---- Nothing could be one peculiar circumstances—feelings rather of a totally entered into with doubtful feelings, with half a heart. I all in suspense I kept my feelings to myself, but now that ion in Mrs. Weston for my feelings, were you to guess her to y to fine sounds than to my feelings. How did Miss Campbell e quickest, and I think our feelings will lose no time there. I must settle with your own feelings." I lay it down as a you to give way to your own feelings?—I never had the the party those pleasantest feelings of our nature, eager ssible for me to have such feelings. I assure you mine are ve rade.-- You will betray your feelings improperly, if you are too ecure him many others. Your feelings are singular. They seem to wisest in you to check your feelings while you can: at any rate Colligational patterns of “feelings” in EMDI, OTHERS and LModE

<table>
<thead>
<tr>
<th></th>
<th>EMDI (14)</th>
<th>OTHERS (12)</th>
<th>LModE (454)</th>
<th>EMDI vs. OTHERS LL value</th>
<th>EMDI vs. OTHERS P value</th>
<th>EMDI vs. LModE LL value</th>
<th>EMDI vs. LModE P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Possessive determiners</td>
<td>9 (64.29 %)</td>
<td>7 (58.33%)</td>
<td>176 (38.77 %)</td>
<td>0.04</td>
<td>0.841481</td>
<td>1.89</td>
<td>0.172625</td>
</tr>
<tr>
<td>Other types of determiners</td>
<td>3 (21.43 %)</td>
<td>2 (16.67%)</td>
<td>157 (34.58 %)</td>
<td>0.08</td>
<td>0.777297</td>
<td>-0.79*</td>
<td>0.3741</td>
</tr>
<tr>
<td>No determiners</td>
<td>2 (14.29 %)</td>
<td>3 (25%)</td>
<td>121 (26.65 %)</td>
<td>-0.39*</td>
<td>0.532299</td>
<td>-0.94*</td>
<td>0.332278</td>
</tr>
</tbody>
</table>
Possessive determiners that colligate with "feelings" in EMDI, OTHERS and LModE

<table>
<thead>
<tr>
<th>feelings + pos. pron.</th>
<th>my</th>
<th>our</th>
<th>your</th>
<th>his</th>
<th>her</th>
<th>its</th>
<th>they</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMDI (9 cases)</td>
<td>3 (33.33%)</td>
<td>1 (11.11%)</td>
<td>5 (55.56%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>OTHERS (7 cases)</td>
<td>3 (42.86%)</td>
<td>1 (14.29%)</td>
<td>0 (0%)</td>
<td>3 (42.86%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>LModE (176 cases)</td>
<td>65 (36.93%)</td>
<td>13 (7.39%)</td>
<td>20 (11.36%)</td>
<td>40 (22.73%)</td>
<td>13 (7.39%)</td>
<td>1 (0.57%)</td>
<td>24 (13.64%)</td>
</tr>
<tr>
<td>E-O LL value</td>
<td>-0.09</td>
<td>-0.03</td>
<td>5.75</td>
<td>-4.96</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>E-O P value</td>
<td>0.764177</td>
<td>0.86249</td>
<td>0.016489</td>
<td>0.02594</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>E-L LL value</td>
<td>-0.03</td>
<td>0.14</td>
<td>7.21</td>
<td>-3.99</td>
<td>-1.30</td>
<td>-0.1</td>
<td>-2.39</td>
</tr>
<tr>
<td>E-L P value</td>
<td>0.86249</td>
<td>0.86249</td>
<td>0.007455</td>
<td>0.045231</td>
<td>0.254213</td>
<td>0.75183</td>
<td>0.120562</td>
</tr>
</tbody>
</table>

Emma’s “your feelings”:

- Unusual collocations + normal grammar (McIntosh 1966)
  - You will betray your feelings improperly, if you are too conscious and too quick;
  - This is a point which you must settle with your own feelings.
  - It will be wisest in you to check your feelings while you can.
  - “Can you seriously ask me, Harriet, whether I imagined him attached to another woman at the very time that I was-- tacitly, if not openly-- encouraging you to give way to your own feelings?

Function keywords (1): “your”, “yourself” and “you” in EMDI

- Density of second-person possessive determiner and pronouns
  - “your”, “yourself”, “you”
    - semantic connection
    - p value: ≤ 0.001
- Biber (1988)
  - characteristic of spoken discourse
  - involvement
Function keywords (2) “yourself” and word distribution

- Sample concordances of “yourself” in EMDI
  - Harriet, Harriet, do not deceive yourself;
  - Read it in comfort to yourself
  - Exert yourself, Harriet, for my sake.
  - I want you to save yourself from greater pain.
  - But take care of yourself, I would not have you too sanguine.
  - Take a little time, consider, do not commit yourself.
  - Command yourself enough to say no more, and I will endeavour to forget it.
  - These are difficulties which you must settle for yourself.

Distribution of “yourself” in EMDI

Collocates of “your” in EMDI and OTHERS

<table>
<thead>
<tr>
<th></th>
<th>Concrete collocates</th>
<th>Abstract collocates</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMDI</td>
<td>42 (31.11%)</td>
<td>93 (68.89%)</td>
</tr>
<tr>
<td>OTHERS</td>
<td>118 (55.92%)</td>
<td>93 (44.08%)</td>
</tr>
<tr>
<td>LL value</td>
<td>- 11.57*</td>
<td>9.20</td>
</tr>
<tr>
<td>P value</td>
<td>0.000831</td>
<td>0.00242</td>
</tr>
</tbody>
</table>

Function words (3): collocational analysis of “your”

- Abstract collocates in EmDi
  - 40.86 % = words and phrases related to cognition and emotions
    - e.g. “your views”, “your affection”, “your expectations”, “your raising your thoughts”, etc.
Function keywords (4): from collocations to speech acts

- Directive:
  “Perhaps, it will be wisest in you to check your feelings while you can.”
  “Let his behaviour be the guide of your sensations.”
  “You must extend your suspicions and comprehend Mr. Dixon in them.”
  “Choose your own degree of crossness.”
  “It will do you good to speak of your distress,...”

Function keywords (4): from collocations to speech acts

- Representative:
  “That, is what you have in your head, I know.”
  “The service he rendered you was enough to warm your heart.”
  “Your countenance testifies to me that your thoughts on this subject are very much like mine.”
  “Your feelings are singular.”
  “The extent of your admiration may take you by surprise.”

EMDI vs. NOVDI

- Keyword extraction
  - Setting:
    - Minimum frequency = 10 occurrences
    - Log-likelihood
    - P value ≤ 0.00001 (NB.: different from EMDI vs. OTHERS)

Keywords in EMDI in comparison with the two corpora

A. your marry yourself likeness idea in be you which either to of with feelings longer might or match attachment his whenever love own appear because side perfectly doubt manner

B. very be it not marry you do must Mrs idea is likeness good quite always much think sort so anybody such sure likely doubt there being nothing everybody perfectly any
Categorisation of keywords extracted from EMDI vs. NOVDI

- **Core keywords:**
  - Occurring in both lists
  - P value ≤ 0.00001 across the two different lists
  - “papa”, “Mr.”, “marrying”

- **Highly significant keywords**
  - Occurring in both lists
  - Differing p values
  - “marry”, “likeness”, “idea”, “doubt”, “perfectly”, “you”

- **Keywords that only occur in EMDI-NOVDI**
  - related to those in EMDI-OTHERS, e.g. “likely”, “must”, “think”, etc.
  - dominant in EMDI-NOVDI only: hyperbole triggers, e.g. “quite”, “perfectly”, “very”, “always”, “everybody”, etc.

---

Hyperbole-related words in EMDI and OTHERS

- **Degree adverbs**
  - Degree intensifiers: “very”, “so”
  - Degree maximisers: “quite”, “perfectly”

- **Non-intensifiers:** “nobody”, “always”

- **Description of the style of Emma’s speech => significant contributions to the study of Jane Austen’s style of characterisation**

---

Observations from quantitative findings

- the critics’ remarks about hyperbole and Jane Austen’s characterisation
  - Validate the critics’ intuitive observations
  - offering more precise descriptions
    - less sophisticated characters: degree max.
    - more sophisticated characters: degree inten.

- Adding linguistic evidence
  - the critics’s observation: less freq. = marked words, e.g. “dreadfully”, “vastly”, etc.
  - frequent words also have a role, e.g. “so”, “quite”
  - hyperbolic non-intensifiers

- Emma’s hyperbolic speech style
  - Emma’s speech style closer to less sophisticated characters but relative freq. not as high as those less admired ones => her mixed character
Hyperbole in EMDI: the lexicogrammatical creation of extreme hyperbole in discourse

- Emma vs. Miss Bates
- Collocational analysis: a case of “quite”
  Sample concordance of “quite” in EMDI
  for you, I suppose she is quite strengthened yours. You were quite possible! You mistake me, you quite ou; but I assure you you are quite he instrument is indifferent. Quite between Miss Fairfax and me is quite kept young Martin. There I was quite of yourself to-night. You do quite

in the dark. The in the right; it was most mistake me. It is not th mistaken. Mr. Elton and otherwise indeed, if I u out of the question. I right. That was well right. We will make your

Hyperbole in EMDI: the lexicogrammatical creation of extreme hyperbole in discourse

- “quite” in EMDI and Miss Bates’ speech have different semantic preference patterns => distinction between JA’s characterisation of the heroine and comic characters
  - EMDI: 53.13% = “perception”, “judgement” or “understanding”, e.g. “right”, “understand” => extreme about the validity of her own and other characters’ understandings/ truth
  - Miss Bates: 65.85% = “emotion/feelings”, “physical states” and “good/bad evaluation” (trivial)

Hyperbole in EMDI: the lexicogrammatical creation of extreme hyperbole in discourse

- Sample concordance of “quite” in Miss Bates’ speech
  I t thinking the asparagus quite boiled enough, sent it all out concerned, I am sure you hit decided against accepting. delightful; so many friends
delightful to have you all depend upon it; nothing can displeased at their being all
distressed that I had owned the frightened if you saw it. I frightened with, “Bless me!

Hyperbole in EMDI: the lexicogrammatical creation of extreme hyperbole in discourse

- Lexical chain among hyperbole-related words => cumulative hyperbole
  you. There is so pointed, and so take disgust towards a girl so is so perfectly satisfied, so a; there is nothing they like so But I shall always think it a very fond of the nephew: he is so very of fortune.--Her allowance is very ving a very particular friend always truly beloved and important; so always sions of some famous ox. Oh! I always d not think such beauty, and such w that I am so entirely without any particular a meaning in idolized and so cried u delighted even -- Pray much. It is such enjoym abominable sort of proce fond of the nephew: he is so very great a favourite. Now, liberal; nothing has ever at hand, to do everything first and always right deserve the best because temper, the highest claim claim to it. You have
Mrs. Weston's manners were always particularly good. Their propriety, simplicity, and elegance would make them the safest model for any young woman.

and if I thought I should ever be like Miss Bates! so silly-- so satisfied-- so smiling-- so prosing-- so undistinguishing and unfastidious-- and so apt to tell everything relative to everybody about me, I would marry tomorrow.

He is very plain, undoubtedly -- remarkably plain: -- but that is nothing compared with his entire want of gentility. I had no right to expect much, and I did not expect much; but I had no idea that he could be so very clownish, so totally without air. I had imagined him, I confess, a degree or two nearer gentility.

It would be dreadful to be standing so close. Nothing can be farther from pleasure than to be dancing.

There will be the leg to be salted, you know, which is so very nice, and the loin to be dressed directly in any manner they like.

Within such apparently trivial conversational scenes lie salient lexical patterns in Emma’s idiolect, which bear implications on the heroine’s extremity.

Methodology:
- JA6 vs. 19CNov
- JA6 vs. BNC

Findings:
- Degree words: irony + insincerity
- Modal auxiliaries: hypothetical + necessity
- BE: good/ bad evaluation
- Female personal pronouns: female characters
- Thoughts and beliefs
Degree words

- Emma could not feel a doubt of having given Harriet's fancy a proper direction and raised the gratitude of her young vanity to a very good purpose, for she found her decidedly more sensible than before of Mr. Elton's being a remarkably handsome man, with most agreeable manners; [...] she was soon pretty confident of creating as much liking on Harriet's side, as there could be any occasion for. She was quite convinced of Mr. Elton's being in the fairest way of falling in love, if not in love already. She had no scruple with regard to him He talked of Harriet, and praised her so warmly, that she could not suppose anything wanting which a little time would not add.

- Mr. Collins on his return highly gratified Mrs. Bennet by admiring Mrs. Phillips's manners and politeness. He protested that, except Lady Catherine and her daughter, he had never seen a more elegant woman; for she had not only received him with the utmost civility, but even pointedly included him in her invitation for the next evening, although utterly unknown to her before. Something he supposed, might be attributed to his connection with them, but yet he had never met with so much attention in the whole course of his life.

Conclusion

- Corpus stylistics is not simply a computer-assisted analysis of literary texts
  - Theoretically: distinctive descriptive tools
  - Methodologically: objectivity in item selection + combined quantitative – qualitative analysis
  - Findings: check the critic’s intuition + new insights into a particular textual and/or authorial style, via additional linguistic evidence
  - Yet, it still needs the researcher’s understanding of the texts under investigation to interpret the patterns.