Around Numbers: Collocational Patterns Involving Numbers in Business English

Combinations of numbers and approximators

Joint 10th ABC Europe convention + 2nd GABC conference
27-29 May 2010
Antwerp, Belgium

Diane Goossens*
Sylvie De Cock*
Philippe Hiligsmann

*Centre for English Corpus Linguistics
AROUND NUMBERS

Research project: ‘quantification linguistic toolbox’ in business language
  ◦ = linguistic devices used to express quantity and approximate quantity

Exploratory study
  ◦ Focus on numbers
  ◦ Collocational patterns involving approximating devices and numbers
  ◦ Computerized corpora of Business English:
    • Academic Business English
    • Business News Reporting
AROUND NUMBERS

- Link between quantification, approximation and vague language
- Channell 1994 ‘Vague Language’
- Corpus-based studies
  - Jucker et al. (2003); Ruzaitè (2007); Koester (2007)
- Make use of pre-established lists of approximators
  - Limited:
    - *about, around*
    - *approximately, roughly, round, like and or so*
METHOD

4
Method

- Corpus-driven method (Tognini-Bonelli 2001): using corpus linguistics as a discovery procedure
- Starting point: numbers (e.g. 15, hundred, three quarters)
- Identify any device used around numbers to approximate a quantity
- POS-tagged corpora (CLAWS 7)
  - 10 tags for numbers (e.g. MC or MD) automatically retrieved using Concord in WordSmith Tools (Scott 2004)
METHOD
METHOD

Methodology and assumptions include:

- An initial database of 7,000 MC companies
- A dataset of 100,000 MC employees
- A sample size of 100,000 MC transactions
- A statistical analysis of 500,000 MC contracts

Concordance of keywords:

- Company
- MC
- Transaction
- Contract
- Employee
DATA
Data

- Business English (heterogeneous)
- Includes a variety of genres:
  - *Talking/writing about business*
    - Business newspapers, books, radio & tv
    - Academic Business English
  - *Doing business*
    - Written: business reports, emails, press releases...
    - Spoken: job interviews, meetings, telephone calls
- In this study: only language used to talk about business (Nelson 2000)
DATA

- Two 500,000 word corpora (Centre for English Corpus Linguistics, Université catholique de Louvain)
  - VESPA_Business: Academic Business English
  - BENews Corpus: Business English News
    - Financial Times, The Economist, The Independent (business section), Times (Economics)
IDENTIFYING APPROXIMATORS

- **Approximator**: any device used around numbers to approximate a quantity
- ‘Any device used around numbers’:  
  - (parts of) words, punctuation  
  - Added before, after and/or to a number  
    - *More or less 20, 50 or something, elevenish,*
- ‘To approximate a quantity’  
  - = to denote imprecision of quantity
- Were discarded: numbers that did not denote a quantity (year, page, scale, ...)

11
APPROXIMATORS IN VESPA_B AND BENews: RESEARCH QUESTIONS

- Corpus-driven approach vs existing lists?
- Impact of genre?: Less imprecision/approximation in Academic Business English than in Business News?
  - ‘It is often assumed that the goal of academic writing is clarity and lack of ambiguity’ (Myers 1996)
CORPUS-DRIVEN APPROACH AND EXISTING LISTS OF APPROXIMATORS
Approximators Uncovered by the Corpus-Driven Approach: Diversity of Approximators

- 47 types

Words

- Single words:
  - *nearly* $1\ billion$ (VESPA_B)
  - *around 1,800 staff* (BENews)
  - *revenues approaching €12bn* (BENews)

- Combinations of words:
  - *at most 2 per cent of economic output* (BENews)
  - *between 3 and 5 years ago* (VESPA_B)
  - *a figure in the region of 8 per cent* (BENews)
Approximators Uncovered by the Corpus-Driven Approach: Diversity of Approximators

- Parts of words
  - Inflectional suffix:
    - *thousands* of households (VESPA_B)
    - *hundreds* of billions of pounds (BENews)
  - Prefix
    - *the multi-* billion euro scramble (BENews)
- ≠ words: punctuation
  - *a total of EUR 20 billion-25 billion* (BENews)
  - *that's going to take two, three, four years* (BENews)
Approximators Uncovered by the Corpus-Directed Approach: Diversity of Approximators

- Combinations of approximators:
  - There were about 5 to 10 top global firms considering entering the Indian market (BENews)
  - The company never lends more than 50 to 80 per cent of the value of the item (BENews)
  - The available data indicate that over the last two decades this factor has netted the US at most $300bn-$400bn (BENews)
  - The goodness-of-fit index (GFI), adjusted goodness-of-fit index (AGFI), normed fit index (NFI), non-normed fit index (NNFI), and comparative fit index (CFI) should be near or higher than .90 (VESPA_B)
Grammatical Diversity

1. Adverbs (14 types)
2. Prepositions (7 types)
3. Comparative adjective + than (6 types)
4. MORE/LESS + than (5 types)
5. Verbs (4 types)
6. Coordinating conjunction + adverb (3 types)
7. Determiners (2 types)
8. Coordinating conjunction (1 type)
9. Prefix (1 type)
10. Suffix (1 type)
11. Prepositional phrase (1 type)
SEMANTIC DIVERSITY

- Lists of approximators in other studies are mostly limited to *about, around, like, approximately, roughly, round, or so*:

  - Actual quantity is larger or smaller than number (expressing approximate amounts)

    - 9/47 types in VESPA_B and BENews
      - *About, around, approximately, roughly, TO APPROXIMATE, some, circa, in the region of, or SO*
**Semantic Diversity**

- Four additional semantic categories were identified in this study.

- Actual quantity $\geq n$ (Expressing minimum amounts): $\frac{15}{47}$ types in VESPA_B and BENews
  - at least, over, above, beyond, more than, greater than, higher than, longer than, larger than, no less than, to exceed, plural number nouns, several, multi-, or more
**Semantic Diversity**

- Actual quantity ≤ n (Expressing maximum amounts): 17/47 types in VESPA_B and BENews
  - *under, below, up to, nearly, near, TO NEAR, close to, almost, towards, TO APPROACH, less than, fewer than, not more than, smaller than, lower than, at most*
  - *Or less*
**Semantic Diversity**

- Actual quantity is between $n$ and $n$ (interval):
  - $\frac{5}{47}$ types in VESPA_B and BENews
  - *Between $n$ and $n$; $n$ to $n$; $n-n$; from $n$ to $n$; $n$, $n$, $n$.***
**Semantic Diversity**

- Actual quantity is either **n** or **n**: 1/47
  
  type in VESPA_B and BENews

  - *N* or *n* (often *one or two*)
# Comparing the 47 Types in Our Corpora with Existing Lists of Approximators

<table>
<thead>
<tr>
<th></th>
<th>Number of approximators in existing lists</th>
<th>Number of approximators also found in our corpora</th>
<th>Additional types of approximators identified in our corpora (not in existing lists)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collins Cobuild English Grammar</td>
<td>29</td>
<td>21</td>
<td>26</td>
</tr>
<tr>
<td>Cambridge Grammar of English</td>
<td>13</td>
<td>8</td>
<td>39</td>
</tr>
<tr>
<td>Channell (1994)</td>
<td>15</td>
<td>14</td>
<td>33</td>
</tr>
<tr>
<td>Ruzaitè (2007)</td>
<td>5</td>
<td>4</td>
<td>43</td>
</tr>
</tbody>
</table>
COMPARISON WITH OTHER LISTS OF APPROXIMATORS

- Not found in our corpora:
  - Frequency? Level of formality? Genre-specific?
    - *round about*
    - *(a) minimum/maximum (of)*
    - *or thereabouts*
    - *or something*
    - *-odd*
    - *-ish*
    - *something like*
AROUND APPROXIMATORS: PREFERRED COMPANY

- Rio’s stockmarket value **hovers around** $35 billion (BENews)
- **IMRG estimates** internet sales account for **about 10 per cent** of all UK retail spending (BENews)
- The HBOS investment is undisclosed but is **thought to be about £2m.** (BENews)
- The researchers found that a sizable amount of variance (**around 32% on average**) could be accounted by how much a subordinate liked (VEPSA_B)

- Percentages: **more than** and **about**
COMPARISON OF ACADEMIC BUSINESS ENGLISH (VESPA_B) AND BUSINESS ENGLISH NEWS (BENews)
## Overall picture in VESPA_Business and BENews

<table>
<thead>
<tr>
<th></th>
<th>VESPA_B</th>
<th>BENews</th>
</tr>
</thead>
<tbody>
<tr>
<td>Types</td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td>Tokens</td>
<td>388</td>
<td>&lt; 2042</td>
</tr>
<tr>
<td>Ty/to ratio</td>
<td>9.02</td>
<td>&gt; 1.96</td>
</tr>
</tbody>
</table>
## Proportion of Numbers Denoting Approximated Quantities in VESPA_B and BENews

<table>
<thead>
<tr>
<th></th>
<th>VESPA_B</th>
<th>BENews</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of ‘numbers’</td>
<td>4921</td>
<td>&lt;</td>
</tr>
<tr>
<td>Proportion of approximations</td>
<td>388/4921</td>
<td>&lt;</td>
</tr>
<tr>
<td>(tokens)/Numbers</td>
<td></td>
<td>2042/14386</td>
</tr>
<tr>
<td>% of approximations/Numbers</td>
<td>7,88</td>
<td>&lt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14,19</td>
</tr>
</tbody>
</table>
# Approximators in VESPA_B and BENews

<table>
<thead>
<tr>
<th>VESPA_B only (7 types)</th>
<th>Shared (28)</th>
<th>BENews only (12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>greater than</td>
<td>more than</td>
<td>10% or so</td>
</tr>
<tr>
<td>lower than</td>
<td>about</td>
<td>under</td>
</tr>
<tr>
<td>not more than</td>
<td>approximately</td>
<td>several</td>
</tr>
<tr>
<td>TO APPROXIMATE</td>
<td>above</td>
<td>multi-</td>
</tr>
<tr>
<td>From one to two</td>
<td>140m-145m</td>
<td>towards</td>
</tr>
<tr>
<td>no less than</td>
<td>up to</td>
<td>beyond</td>
</tr>
<tr>
<td>smaller than</td>
<td>around</td>
<td>TO NEAR</td>
</tr>
<tr>
<td></td>
<td>over</td>
<td>at most</td>
</tr>
<tr>
<td></td>
<td>at least</td>
<td>circa</td>
</tr>
<tr>
<td></td>
<td>almost</td>
<td>larger than</td>
</tr>
<tr>
<td></td>
<td>nearly</td>
<td>two, three, four</td>
</tr>
<tr>
<td>plural number nouns</td>
<td>between 75% and 80%</td>
<td>in the region of</td>
</tr>
<tr>
<td></td>
<td>below</td>
<td></td>
</tr>
<tr>
<td></td>
<td>less than</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

47 types found in VESPA_B + BENews
# Top Ten Approximators in VESPA_B and BENews

<table>
<thead>
<tr>
<th>Top ten approximators in VESPA_B</th>
<th>Number of tokens in VESPA_B (total: 388 tokens)</th>
<th>Top ten approximators in BENews</th>
<th>Number of tokens in BENews (total: 2042 tokens)</th>
</tr>
</thead>
<tbody>
<tr>
<td>at least</td>
<td>53</td>
<td>more than</td>
<td>447</td>
</tr>
<tr>
<td>more than</td>
<td>44</td>
<td>about</td>
<td>325</td>
</tr>
<tr>
<td>over</td>
<td>41</td>
<td>plural number nouns (dozens, hundreds, thousands, millions, billions)</td>
<td>234</td>
</tr>
<tr>
<td>about</td>
<td>37</td>
<td>up to</td>
<td>173</td>
</tr>
<tr>
<td>n or more</td>
<td>29</td>
<td>around</td>
<td>135</td>
</tr>
<tr>
<td>approximately</td>
<td>27</td>
<td>almost</td>
<td>89</td>
</tr>
<tr>
<td>greater than</td>
<td>16</td>
<td>over</td>
<td>89</td>
</tr>
<tr>
<td>less than</td>
<td>16</td>
<td>nearly</td>
<td>85</td>
</tr>
<tr>
<td>almost</td>
<td>15</td>
<td>at least</td>
<td>63</td>
</tr>
<tr>
<td>between n and n</td>
<td>14</td>
<td>some</td>
<td>58</td>
</tr>
</tbody>
</table>
Approximators in VESPA_B and BENews

- Genre differences => different approximators?
  - *Approximately*: ‘Generally confined to rather specific official, semi-official, and scientific registers’ (Channell 1994)
    - 27 in VESPA_B > 7 in BENews
  - *Two, three, four* (BENews only): reported speech!
    - Business News Reporting: hybrid genre => also speech!
  - Plural number nouns: more in BENews + combined
    - 234 in BENews > 14 in VESPA_B
  - Amounts of money
CONCLUSIONS
CONCLUSIONS

- Corpus-driven approach: diversity
  - Number of types (> existing lists)
  - Grammar
  - Semantics
- More approximators in **BENews** than **VESPA_B** (proportion of numbers approximated and tokens)
- Genre-specific approximators
REFERENCES