Quantification and Approximation in Business Language. A Contrastive Approach (Dutch, English, French)

Présentation du projet FSR

Diane Goossens*

Promoteurs:
Philippe Hiligsmann
Sylvie De Cock*

*Centre for English Corpus Linguistics
Plan de la présentation

- Présentation du projet FSR
- Les notions de quantification et d’approximation
- Les données
- Méthodologie
- Implications
OUR PROJECT

- **Aim**: explore and contrast the contents of the ‘quantification linguistic toolbox’ in business language on the basis of computerized corpora

- Linguistic devices used to express quantity and approximate quantity → uncover the various compartments of the toolbox

- Onomasiological approach

- Empirical investigation
OUR PROJECT

- **Aim**: explore and **contrast** the contents of the ‘quantification linguistic toolbox’ in business language on the basis of **computerized corpora**

- **Contrastive approach**:
  - English
  - Dutch & French
QUANTIFICATION AND APPROXIMATION
**Quantification and Approximation**

- **Quantification**: expression of quantity and amounts

  - Variety of devices that can be used to express quantity/amounts (e.g. 153, *both*, *large*, *substantially*, *a series of*, *the majority*, *to increase*, *many*, *a vast amount of*)

  - Quantities and amounts can be expressed in an *imprecise* manner → quantities can be *approximated*

- ‘Quantification’ and ‘approximation’ = two closely connected notions

  - Precise or imprecise quantity?

- Explore the relationship between quantification and approximation on a theoretical level
QUANTIFICATION AND APPROXIMATION

- Previous linguistic studies
- Very little empirical research for Dutch and French
  - E.g. papers presented at the conference ‘La quantification et ses domaines’ (2006)
Quantification and Approximation

A number of corpus-based descriptive studies for English

- Channell (1994): *Vague Language*
- Drave (2002) *Vaguely speaking: a corpus approach to vague language in intercultural conversations*
- Jucker et al. (2003) Interactive aspects of *vagueness* in conversation
- Koester (2009) ‘About Twelve Thousand or So’: *Vagueness in North American and UK Offices*
QUANTIFICATION AND APPROXIMATION

- Kennedy (1987) Quantification and the use of English: A case study of one aspect of the learner’s task
- Smith (2009) Non-numerical quantifiers

Corpus-based: pre-established lists (e.g. dictionaries, grammars, Channell’s list 1994)
THE DATA
**The Data**

- Business language: heterogeneous
- A variety of genres:
  - *Talking/writing about business*
    - Business newspapers, books, radio & tv
    - Academic business texts
  - *Doing business*
    - Written: business reports, emails, press releases...
    - Spoken: job interviews, meetings, telephone calls
THE DATA

- English – a number of genres
  - Picture of the linguistic devices used to express quantity and to approximate quantity in Business English
  - Investigate the impact of genre (e.g. genre-specific devices?, differences in frequency?)

- Written and spoken genres
- Language used to talk/write about business and language used to do business
THE DATA

- English

- Written genres:
  - Business news reporting (newspapers, journals, magazines)
    - BENews corpus (Business English News corpus, Centre for English Corpus Linguistics, UCL)
  - Academic Business English
    - LOCRA_Business corpus (LOuvain Corpus of Research Articles, Centre for English Corpus Linguistics, UCL)
  - Other genres?
    - Formal and informal business correspondence (business letters and faxes; e-mails):
      - British National Corpus and Business English Corpus (Nelson 2000)
THE DATA

- English

- Spoken genres:
  - Meeting and negotiating (dialogic)
    - British National Corpus, Business English Corpus (Nelson 2000)
  - Speeches (monologic)
    - British National Corpus, Business English Corpus (Nelson 2000)
THE DATA

- Dutch and French: limited number of genres
  - Two written genres
    - Business news reporting
    - Academic Business texts
  - Comparable corpora (not translation corpora)
    - Original texts in Dutch and French
    - Sampled using similar design criteria as BENews and LOCRA_BUSINESS
Method
METHOD

- Explore the contents of the ‘quantification linguistic toolbox’
- Not on the basis of pre-established lists
- Corpus-driven approach (Tognini-Bonelli 2001)
  - inductive approach
  - (semi-) automatic corpus linguistic techniques as a discovery procedure; e.g. frequency lists of single words or of sequences of words; keywords analysis
  - starting point to uncover the variety of devices that can be used
  - focus on specific devices and the patterns in which they occur
**METHOD**

- Language users’ perception of some of the approximating devices uncovered
- Experimental elicitation involving both native and non-native speakers
  - Combinations of approximators
    - *more than 50 to 80 per cent*
  - Semantically ‘ambiguous’ approximators
    - *Nearly 500 people*
    - *About 500 people*
METHOD

- One possible starting point: **numbers**

- Exploratory study: ‘Around Numbers: Collocational patterns involving numbers in Business English – Combinations of numbers and approximators’

- 2 corpora (500,000 words):
  - Academic Business English (Business subcorpus of LOCRA)
  - Business News Reporting (BENews)
AROUND NUMBERS

• Starting point: numbers (e.g. 15, hundred, three quarters)

• Identify any device used around numbers to approximate a quantity

• Part-of-Speech-tagged corpora (CLAWS 7):
  • 10 tags for numbers (e.g. MC or MD) automatically retrieved using Concord in WordSmith Tools (Scott 2004)
A ROUND NUMBERS

price_NH1 rtses_VZ a_AT1 key_JU priority_NN1 for_IF 2007_MC _
OWER_RPs as-CSA Microsoft_NP1 CEO_NN1 in_IL 2000_MC _
capable_JU handhelds_NN2 reaching_VVY 50m_NN1 in_IL 2009_MC _
fall_VVD 2.8_MC per_NN121 cent_NN22 to_IL 141_MC _
down_RP 4.9_MC per_NN121 cent_NN22 at_IL 88.4_MC _
2_MC per_NN121 cent_NN22 higher_JU at_IL 22.9_MC _
outweighs_VVZ the_AT A_s OF [drawbacks_NN2 &s<]<_] _
The_AT_Financial_JU Times_NR122 Limited_JU 2006_MC _

placed_VVn_a_AT1 50p_NN1 acumulator_NN1 in_IL eight_MC horses_NN2 and_CC won_VVY 1million_F0 after_CS

especially_RR in_IL Delhi_NR1 _ where_RQ only_RR two_MC dots_NN2 are_VBB on_IL ever_NN1 in_IL some_DD

kisses_NN2 Heard_VVD this_ID1 one_PNN1 ?_Two_MC women_NN2 are_VBB walking_VVZ through_IL the_AT

The_AT figures_NN2 are_VBB based_VVZ on_IL 10_MC analyst_NR122 forecasts_NN2 collated_VVD by_IL Standard_JU

currently_NR122 operates_VVZ 33_MC aircraft_NR122 on_IL 43_MC short_Haul_NR122 flights_NR122 with_JV fewier_DAR_daily_JU

If_PPH1 would_VM deliver_VVD in_IL initial_NR1 only_RR 9_MC A380s_F0 rather_NR1 122 then_122 the_AT 25_MC

and_CC stamina_NR1 from_IL which_DD0 only_RR two_MC outcomes_NN2 are_VBB possible_JU Either_RR Mr_NR122

" For_IF a_AT1 small_JU employer_NN1 with_IL only_RR five_MC to_IL ban_MC staff_NN1 _ it_PPH1 is_VBB very_RG

mergers_NN2 and_CC acquisitions_NR122 new _ with_IL only_RR 7_MC per_NN121 cent_NN22 saying_VVY that_CST they_PPH32

months_NR122 of_IO last_VD_year_NR122 _122 with_IL only_RR 9_MC per_NN121 cent_NN22 reporting_VVZ a_AT rise_NR122 ninew_NN122

free_VVZ it_PPH1 from_IL their_APRFZV_122 but per_NR122 cent_NN22 interest_NR122 charge_NR122 _

" For_IF a_AT1 small_JU employer_NN1 with_IL only_RR five_MC to_IL ban_MC staff_NN1 _ it_PPH1 is_VBB very_RG

and_CC has_VVZ been_VVZ working_VVZ only_RR two_MC days_NN2 and_AT 1 year_NR122 since_IL September_NR122

tones_NR122 a_AT1 lead_NR122 of_IO 13_MC points_NR122 on_IL 45_MC per_NN121 cent_NN22 say_VVZ that_CST they_PPH32

instead_RR _ Mr_NR122 Ellison_NN3_F0 is_VBB keen_NR122 on_IL 22 MC to_CONN2 of_IO Sun_NN2 software_NR122 portfolio_NR122

Mr_NR122 Abarne_NN2_PNP122 recomb ine_NR122 to_IL two_MC difficult_JU years_NR122 into_IL second_RR commodity_NR122

The_AT banks_NR122 resources_NR122 on_IL 10_MC of_IO the_AT riskier_JU areas_NR122 of_IO the_AT

If_PPH1 would_VM announce_VVZ_122 results_NR122 on_IL 7_MC March_NR122 _ HSBC_NR122 has_VVZ not_OK ruled_VVZ

eight_MC in_IL the_AT series_NR122 NN1 released_VVZ on_IL 18_MC November_NR122 _ but_CC the_AT game_NR122 met_VVD

which_DDQ have_VVZ left_VVZ it_PPH1 relant_JU _ NN1 met_VVD

sales_NR122 via_VVZ 23_MC per_NN121 cent_NN22 on_IL 2007_MC _ Ford_NR122 s old_VVZ yesterday_RT that_CST sales_NR122

up_RP about_24_MC per_NN121 cent_NN22 on_IL 2005_MC despite_IL Vodafone_NR122 14_Sin_F0 Los_NR122 for_IF

by_IL some_DD 15_MC per_NN121 cent_NN22 on_IL 2008_MC which_DDQ itself_PPH1 saw_VVZ a_AT1 reduction_NR122

down_RP 18_MC per_NN121 cent_NN22 on_IL 2007_MC http_NR122 _ www.timesonline.co.uk

so up_14_MC per_NN121 cent_NN22 on_IL 2008_MC according_NR122 I21 to_I22 the_AT recent_JU figures_NR122

two-year_NR122_ju one_PNN1 instead_RR _ focusing_VVZ on_IL two_MC things_NR122 _ internal_NR122 reorganisation_NR122 and_CC

operating_NR122 system_NR122 _ - _ installed_VVZ on_IL new_MC out_I21 of_I22 ten_MC personal_JU computers_NR122

such_I21 as_I22 service_NR122 charge_NR122 _ on_IL 70_MC empty_JU stores_NR122 _ Market_VVZV sources_NR122

of 2003

concordance

distributes
calculate

datasets
patterns

discern
fingerprints

source

notes
Around Numbers - Conclusions

- Corpus-driven approach: diversity
  - Number of types (> existing lists)
    - 47 types including words, parts of words and punctuation
      - Nearly, at most, approaching, multi-billion euro,
      - two, three, four years
  - Grammatical diversity: 11 word classes (adverbs, prepositions, MORE/LESS + than, verbs etc.)
  - Semantic diversity: 5 semantic categories (not just « actual quantity is larger or smaller than number »)
AROUND NUMBERS - CONCLUSIONS

- More approximators in **BENews** than LOCRA_B
  (proportion of numbers approximated and tokens): 14.19% vs 7.88%
  - Less imprecision/approximation in academic Business English

- Genre-specific approximators
  - Plural number nouns (e.g. millions of): more in **BENews** + combined (e.g. hundreds of billions of pounds)
    - 234 in **BENews** > 14 in LOCRA_B
  - Amounts of money
OTHER APPROACHES / METHODS – WORK IN PROGRESS

- Semantic tagging: tags for “quantities”
- KeyWords Analysis (WordSmith Tools): compare corpora to a reference corpus (fiction) – extract list of genre-specific words
- WordLists (WordSmith Tools): frequency lists of single words or of sequences of words
LOOKING AHEAD: POSSIBLE APPLICATIONS
LOOKING AHEAD: POSSIBLE APPLICATIONS

- Improved treatment of quantification and approximation
  - Lexicography
    - Inclusion and description in dictionaries and more specifically specialized business dictionaries
    - Examples reflecting preferred patterns of usage
  - Pedagogical applications
    - Business textbooks
REFERENCES