

Corpus-Based/ Corpus-Driven Discourse Analysis: the example of gendered discourse

Dr Sylvia Jaworska, s.jaworska@qmul.ac.uk
Summer School in Corpus Linguistics
Aston University, August 2011

Outline

- Discourse?
- Corpus Linguistics vs. (Critical) Discourse Analysis
- Case Study 1: Gendered Discourse
- Discussion
- Case Study 2: Feminism
- Discussion

Discourse?

- Linguistics (Discourse Studies, Discourse Analysis, Critical Discourse Analysis), Anthropology, Sociology, Social Psychology, History
- Definitions in Linguistics:
 - language above the sentence or above the clause (Stubbs 1983: 1)
 - language use and structures related to genres (text types) and registers
 - a variety of semiotic elements of social practice (language, non-verbal, visual) (Fairclough 1995)

Discourse?

- “ways of seeing the world, often with reference to relations of domination and power” (Sunderland 2004: 6)
- potentially constitutive: “practices that systematically form the object of which they speak” (Foucault 1972: 49); “flows of information which constructs the world through language and text, and ‘subject position’ of individuals.” (Sunderland 2004: 8)
- diachronic perspective (Discourse-Historical Approach, Wodak 2001)

Critical Discourse Analysis (1)

- “CDA is a – critical – perspective on doing scholarship: it is, so to speak, discourse analysis ‘with attitude’. It focuses on social problems, and especially on the role of discourse in the production and reproduction of power abuse and domination” (van Dijk 2001: 96)

discourse – cognition – society

discourse: communicative event (interaction, texts, gestures, images...)

cognition: beliefs, goals, evaluations, emotions, mental and memory structures

society: social and political structures (groups, movements, institutions)

- “there is no typical CDA way of collecting data” (first data collection, first analysis, finding indicators for particular concepts , expanding concepts into categories, collecting further data) (Meyer 2001: 23)
- “CDA relies strongly on linguistic categories [...] such as actors, mode, time, tense, argumentation” (Meyer 2001: 25)
- analysis of formal linguistic features such as pronoun use, modality, metaphors, agency, passivisation, nominalisation (Fairclough 1989).

Critical Discourse Analysis (2)

- Analysis of **semantic macrostructures** (topics and sub-topics)
- Analysis of **local meanings** (particularly implicit or indirect meanings, meanings of words)
- Analysis of **subtle formal structures** (linguistic markers such as: lexical styles, speech acts, schematic organisation, rhetorical figures, syntactic structures, turn taking, hesitation etc.)
- Analysis of **specific linguistic realisations**, e.g. hyperboles
- Analysis of **global and local discourse** forms (intertextuality)
- Analysis of **context** (participants, setting)

“ a full analysis of a short passage might take months and fill hundreds of pages. Complete discourse analysis of a large corpus of text or talk , as we often have in CDA research, is therefore totally out of the question. ”

(van Dijk 2001: 99)

Corpus Linguistics (1)

- large collection of data stored electronically
- researchers approach data relatively free from preconceived notions
- no prior categories – categories emerge from data (corpus-driven)
- quantitative basis for analysis (patterns identified by analysing frequencies, concordances, collocations, clusters and keywords)
- verification of results
- **“Corpus Linguistics is the closest Linguistics can get to science”** (a personal conversation with Ramesh Krishnamurthy)

Corpus Linguistics (2)

- **Frequency**

- what is frequent and typical in the given data set (saliency)
- direct researcher's attention to particular items in the corpus

language consists of thousands of words and patterns but certain lexical and grammatical choices are, in some situations, preferred over others: "Choice of words expresses an ideological position" (Stubbs 1996: 107)

- **Collocations:** the tendency of words to attract each other

- "Collocation is [...] a way of understanding meanings and associations between words which are otherwise difficult to ascertain from a small-scale analysis of a single text" (Baker 2006: 96)
- collocations are not simply lexical items, they "are also widely shared within a speech community" (Stubbs 2001:35) and are often "nodes around which ideological battles are fought" (ibid.: 188).

Corpus-Based vs. Corpus-Driven

- **Corpus-based (deductive):** analysis of patterns of use for pre-defined linguistic feature (the frequency, functions and variation of any given category or items); corpus as a source of examples
- **Corpus-driven (inductive):** rejects any pre-defined categories and starts normally with simple word forms (its strict version does not consider lemmas); categories emerge from the data

Corpus-Based Discourse Analysis

EXAMPLE 1: GENDERED DISCOURSE

- “the working of a particular set of ideas about gender in some segment or segments of society” (Eckert & McConnell-Ginet 2003: 42)
- “something to do with gender is going on” (Sunderland 2004: 21)
- Gender difference: “positioning of women and men in different ways” (Sunderland 2004: 21)
- Gender as a clear case of strong bipolarity (masculine/ feminine binary opposition)
 - ‘Gender Differences’ discourse
 - ‘Gender Equality Now Achieved’ discourse
 - ‘Poor Boys’ discourse
 - ‘Battle of Sexes’ discourse (‘Horse-Race’ discourse)
- **Gendered discourse is frequently evaluated as unfavourable to women** (Sunderland 2004)
- **Women are often relegated to a negative semantic space** (Romaine 2000: 112)

Corpus-Based Discourse Analysis

- **Exercise 1:** The Times: **'My husband does it all'**
 - the 'female success in business' discourse
 - the 'gender equality now achieved' discourse
 - the 'maternity leave as a burden for businesses' discourse
 - the 'sexism is not a problem' discourse
 - the 'role reversal' discourse

Deborah Meaden: one of the 'dragons' on the BBC 2 programme **Dragons' Den**



businesswoman

Meaden is a symphony in expensive beige

she is equally **hard-nosed**

she set up a flower stall

she works or networks

her family's holiday park **business**

he has given up his job to run her domestic life

her dismissive catchphrase

he's a fabulous cook

Corpus-Based Discourse Analysis

- **Deborah Meaden, the businesswoman - female success?**

WORD	BYU-BNC (1980s – 1993) 100,000,000	COCA (1990 – 2011) 420,000,000	WordBanks (UK) (1990s – 2005) 553,171,489
BUSINESSWOMAN	71	434	326
BUSINESSWOMEN	5	81	40
BUSINESSMAN	959	4,624	5,449
BUSINESSMEN	956	3,060	2,205

Business is a man's world

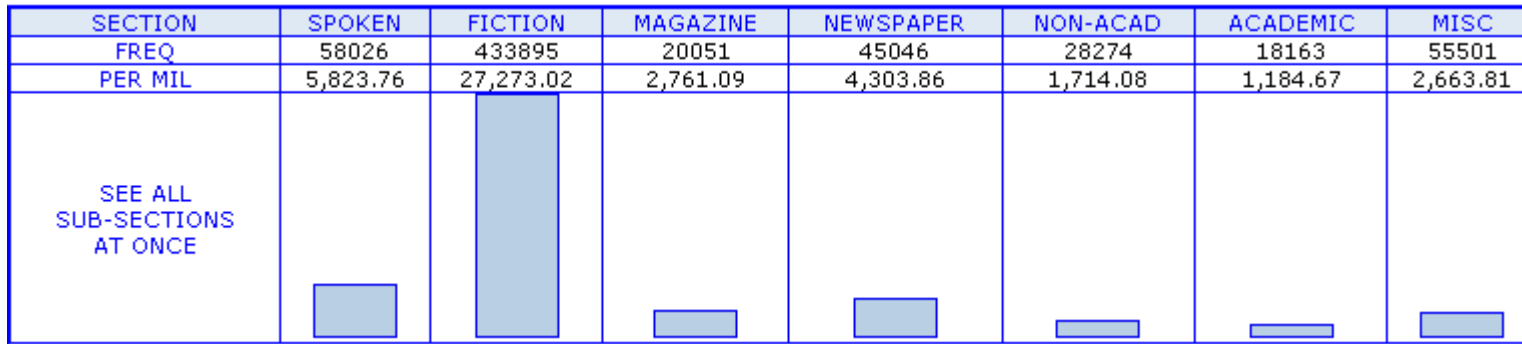
Corpus-Based Discourse Analysis

- she, her, hers, he, his, him

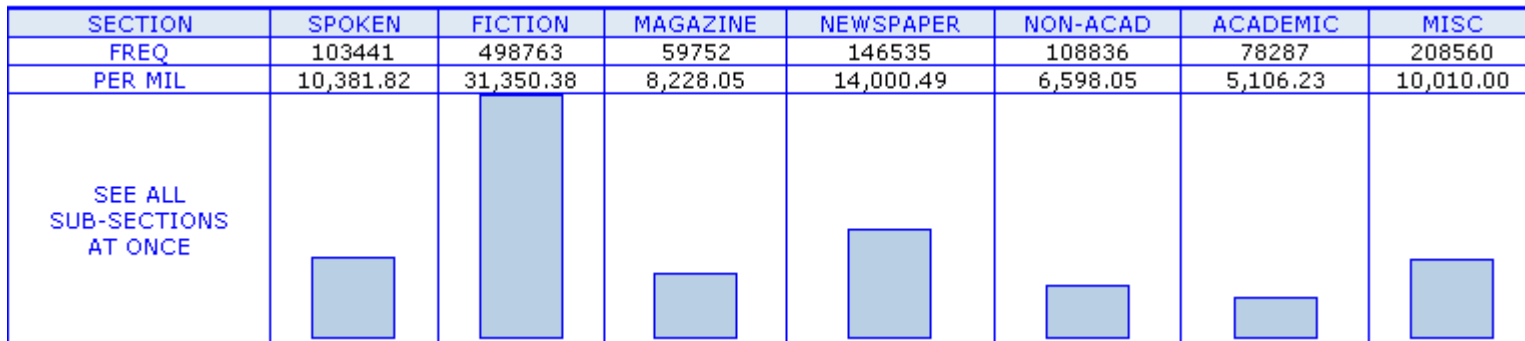
WORD	BYU-BNC (1980s – 1993) 100,000,000	COCA (1990 – 2011) 420,000,000	WordBanks (UK) (1990s – 2005) 553,171,489
SHE	352,865	1,609,166	708,363
HER	303,724	1,486,512	643,469
HERS	2,367	8,445	8,113
	= 658,956	= 3,104,123	= 1,359,945
HE	640,714	3,139,905	1,800,099
HIS	409,816	1,943,618	1,191,018
HIM	153,650	1,965,000	410,673
	=1,204,180	= 7,048,523	= 3,401,790

Corpus-Based Discourse Analysis

SHE|HER|HERS (BYU-BNC)



HE|HIS|HIM (BYU-BNC)



Corpus-Based Discourse Analysis

SHE|HER|HERS (COCA)

SECTION	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	ACADEMIC
FREQ	410541	1748359	415077	381336	148808
PER MIL	4,558.24	20,577.28	4,597.05	4,399.84	1,734.52
SEE ALL SUB-SECTIONS AT ONCE					

HE|HIS|HIM (COCA)

SECTION	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	ACADEMIC
FREQ	1057091	2341507	924324	1107767	407209
PER MIL	11,736.88	27,558.32	10,237.05	12,781.36	4,746.47
SEE ALL SUB-SECTIONS AT ONCE					

Corpus-Based Discourse Analysis

He has given up his job to run her domestic life

BYU-BNC (1980s – 1993)

[give] up * job

GIVE UP HER JOB	25	GIVE UP MY JOB	3
GIVE UP HIS JOB	10	GIVE UP A JOB	2
GAVE UP HIS JOB	8	GIVEN UP THEIR JOB	2
GAVE UP HER JOB	7	GIVING UP YOUR JOB	1
GAVE UP MY JOB	7	GIVEN UP THE JOB	1
GIVE UP THE JOB	6	GIVING UP A JOB	1
GIVEN UP HER JOB	6	GIVES UP HER JOB	1
GAVE UP THE JOB	4	GIVEN UP A JOB	1
GIVEN UP HIS JOB	4	GIVE UP ME JOB	1
GIVE UP YOUR JOB	4	GIVE UP HERE JOB	1
GIVING UP MY JOB	4	GIVE UP 'ER JOB	1
GIVING UP HIS JOB	4	GAVE UP YOUR JOB	1
GIVING UP HER JOB	3	GAVE UP THIS JOB	1

Corpus-Based Discourse Analysis

go on for ever , whether he might not give	up his job	and become Marcus 's -- what ? -- friend , secretary ,
drive because of her visual handicap ; had to give	up her job	and can not pursue her former hobbies of walking and gardening
, " she said , " if you gave	up your job	and drew unemployment money . " " I would never do that
mother 's will and decided that he should give	up his job	and go in to business for himself . Whilst the lawyers were
back at work but said he had considered giving	up his job	and moving house following the horrific incident . " I was just
was barely audible . " She had to give	up her job	and she had to buy things . She would n't consider getting
the doctor to " tell Susie " to give	up her job	and stay at home . At this point the GP had felt
insisted his only reason for urging her to give	up her job	and stay here was so she could help his grandmother get well
'd been stupid . She said she 'd give	up her job	and we 'd go back up north . Start again . "
her breathless and her doctor wants her to give	up her job	as a cook . Ruby claims her illness was caused by washing
department . In 1975 she had had to give	up her job	as a council clerkess as she had contracted asbestos-related
(Unmarried woman receiving only ICA who had given	up her job	as a local government clerk to care for her mother) CONCLUSI
, aged sixteen in 1943 ; hoped to give	up her job	as a messenger to do secretarial studies full time : Brigitte did
Helen , almost as pale as Claire , gave	up her job	as a part-time receptionist to be with her daughter . HER husban
. He has a pregnant wife yet has given	up his job	as a physio because it started to interfere with his rugby
way than I could Do you ever regret giving	up your job	as a psychiatric nurse ? No ! I 'd have ended up
. Tony began 20 years ago ; after giving	up his job	as a scientist . With skill and concentration , it can still
a high drop-out rate . " Lyn Haughey gave	up her job	as a secretary in Tring , Hertfordshire , when she moved to
and Duchess of Kent , who is just giving	up her job	as a Soho art dealer , and her husband , West End
, two children . Yes ; but Ross gave	up his job	as a uhum , he was in with the Leeds , which
surviving . " Largue has since had to give	up his job	as an apprentice turner and is now re-training as an electronic
over Mountain bike enthusiast ; Martin Richmond gave	up his job	as an environmental scientist for British Coal at Stoke Orchard
cancer two years ago . Kim has now given	up her job	as audio producer to dedicate her life to Bronson . The Death

Corpus-Based Discourse Analysis

- Collocations of 'his' and 'her'

BYU-BNC (1980s – 1993)		COCA (1990 – 2011)		WordBanks (UK) (1990s – 2005)	
HER	HIS	HER	HIS	HER	HIS
1. EYES 2. FACE 3. MOTHER 4. HUSBAND 5. HAIR 6. HANDS 7. ARMS 8. MOUTH 9. ARM 10. DAUGHTER 11. LIPS 12. SHOOK 13. FINGERS 14. SISTER 15. SHOULDERS 16. LEGS 17. BESIDE 18. NECK	WIFE EYES FATHER HANDS MOUTH BROTHER CAREER SHOOK LIPS FINGERS COLLEAGUES SHOULDER POCKET CHEST SHOULDERS THROAT GAZE KNEES	MOTHER EYES HUSBAND HAIR DAUGHTER ARMS MOUTH SISTER FINGERS LIPS SHOOK NECK SHOULDERS KNEES THROAT BREASTS GRANDMOTHER BOYFRIEND	WIFE MOUTH ARM SHOOK FINGERS SHOULDER CHEST POCKET LIPS COLLEAGUES SHIRT SHOULDER KNEES THROAT FOREHEAD GAZE CHIN GIRLFRIEND	HUSBAND MOTHER EYES HEAD FACE LIFE FATHER HOME TIME HANDS WAY FAMILY HAIR DAUGHTER BODY VOICE ARMS YEARS	WIFE LIFE HEAD FATHER HOME CAREER FACE HAND EYES TIME FAMILY TEAM MOTHER HANDS WORK YEARS GOAL DEATH

Corpus-Based Discourse Analysis

COCA (1990 – 2011),
Newspapers

HER	HIS
1. HUSBAND	1. WIFE
2. MOTHER	2. FATHER
3. SON	3. CAREER
4. FATHER	4. BROTHER
5. DAUGHTER	5. COLLEAGUES
6. SISTER	6. TEAMMATES
7. HAIR	7. GIRLFRIEND
8. MOM	8. DEBUT
9. BOYFRIEND	9. GRANDFATHER
10. GRANDMOTHER	10. HOMETOWN
11. BIRTHDAY	11. TENURE
12. DAUGHTERS	12. PREDECESSOR
13. PURSE	13. SUCCESSOR
14. AUNT	14. EX-WIFE
15. LOVER	15. CANDIDACY
16. PREGNANCY	16. BUDDIES
17. EX-HUSBANB	17. AUTOBIOGRAPHY
18. CLASSMATES	18. COUNTERPART

WordBanks (UK) (1990s – 2005)
Newspapers

HER	HIS
1. HUSBAND	1. WIFE
2. HOME	2. SIDE
3. LIFE	3. CAREER
4. MOTHER	4. HOME
5. DAUGHTER	5. LIFE
6. FAMILY	6. TEAM
7. SON	7. TIME
8. FATHER	8. GOAL
9. MUM	9. CLUB
10. TIME	10. FAMILY
11. YEARS	11. SEASON
12. LOVE	12. FATHER
13. PARENTS	13. DEBUT
14. BOYFRIEND	14. GAME
15. DEATH	15. HEAD
16. CHILDREN	16. MAN
17. SISTER	17. PLAYERS
18. CAREER	18. WORK

HER CAREER

893

HIS CAREER

5,890

Corpus-Based Discourse Analysis

- Measures of Significance (different measures favour different words)
 - **z-score**: tends to favour low frequency words
 - **Mutual Information (MI)**: tends to give high scores to low frequency words
 - **t-score**: tends to favour high frequency words
 - **log-likelihood (LL)**: favours high frequency words

Baker, P. 2006: 100 – 104

Barnbrook, G. 1996: 94 – 102

Corpus-Based Discourse Analysis

she is equally hard-nosed

WORD	BYU-BNC (1980s – 1993)	COCA (1990 – 2011)	WordBanks (UK) (1990s – 2005)
HARD-NOSED	57	378	229
collocations (MI)	BUSINESSMAN SCEPTICS BITCH BUNCH APPROACH POLITICAL	APPROACH BUSINESS TOUGH PLAY STYLE GUY AGGRESSIVE PLAYERS COP FOOTBALL PLAYER COACH	BUSINESSMAN RUTHLESS BUSINESSMEN DETECTIVE ATTITUDE APPROACH MANAGERS LAWYER COMMERCIAL PROFESSIONAL BUSINESS ECONOMIC

Corpus-Based Discourse Analysis

- Right adjectives of ‘she is’ and ‘he is’ in the COCA press sub-corpus

COCA (1990 – 2011)

SHE IS ...	HE IS ...
MARRIED	CONCERNED
CONCERNED	READY
WILLING	WILLING
PROUD	AWARE
PREGNANT	CONFIDENT
AFRAID	HAPPY
READY	PROUD
HAPPY	INTERESTED
AWARE	INNOCENT
SMART	CONVINCED
QUICK	AFRAID
EAGER	CAPABLE
INTERESTED	QUICK
CONFIDENT	SURPRISED
COMFORTABLE	GUILTY
CONVINCED	OPTIMISTIC
ANGRY	WORRIED
BEAUTIFUL	EAGER

Corpus-Based Discourse Analysis

- Further gendered areas worth exploring:
 - “words with negative overtones are still more frequently used together with ‘girl/woman’ than with ‘boy/man’” (Romaine 2000: 109)
 - many kinds of ‘wives’ and ‘mothers’: ‘working mothers’, ‘housewives’, but no ‘working fathers’, ‘single fathers’ or ‘househusbands’
 - women never grow up: the usage of the term ‘girl’ (e.g. Angela Merkel was referred to as ‘mein Mädchen’ by the former chancellor Helmut Kohl) (Sigley & Holmes 2002)
 - women: frequently described by references to their marital status and appearance
 - naming practices and titles: *Mrs*, *Mr*, *Ms* and *Miss* (Scott & Tribble 2006)

Corpus Linguistics: merits and limitations

- **Merits:**

- large data sets as opposed to a few articles
- quantitative basis (guards against over- and under-interpretation)
- salient lexico-grammatical patterns
- identification of lexico-grammatical patterns otherwise unobserved
- identification of repeated patterns (incremental effect of discourse, Baker 2006: 13) “Repeated patterns show that evaluative meanings are not merely personal or idiosyncratic, but widely shared in a discourse community” (Stubbs 2001: 215)
- reduction of researcher's bias (primacy effect, confirmation bias)

Corpus Linguistics: merits and limitations

- Limitations:
 - discourse is not confined to language only (verbal, non-verbal, gestures, prosody, images)
 - socio-cultural and historical context
 - frequent patterns do not always point to dominant discourses (the power of individual texts and people)
 - production and reception of texts

References

- Baker, P. (2006). *Using Corpora in Discourse Analysis*. London, Continuum.
- Eckert, P. & McConnell-Ginet, S. (2003). *Language and Gender*. Cambridge, Cambridge University Press.
- Fairclough, N. (1989). *Language and Power*. London, Longman.
- Fairclough, N. (1995). *Media Discourse*. London, Hodder Arnold.
- Foucault, M. (1972). *The Archaeology of Knowledge*. London, Tavistock Publications.
- Meyer, M. (2001). Between theory, method and politics: positioning of the approaches to CDA. In: Wodak, R. & Meyer, M. *Methods of Critical Discourse Analysis*. London, Sage: 14– 31.
- Romaine, S. (2000). *Language in Society*. Oxford, Oxford University Press.
- Scott, M. & Tribble, Ch. (2006). *Textual Patterns*. Amsterdam/Philadelphia, Benjamins.
- Stubbs, M. (1983). *Discourse Analysis: the Sociolinguistic Analysis of Natural Language*. Chicago, University of Chicago Press.
- Stubbs, M. (2001). *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford, Blackwell.
- Sunderland, J. (2004). *Gendered Discourses*. Basingstoke, Palgrave.
- van Dijk, T. (2001). Multidisciplinary CDA: a plea for diversity. In: Wodak, R. & Meyer, M. *Methods of Critical Discourse Analysis*. London, Sage: 95–120.
- Wodak, R. (2001). The discourse-historical approach. In: Wodak, R. & Meyer, M. *Methods of Critical Discourse Analysis*. London, Sage: 63–94.

***** **Corpus-Based or Corpus Driven Discourse Studies** *****

- Baker, P. & McEnery, T. (1996). A corpus-based approach to discourses of refugees and asylum seekers in UN and newspaper texts. In: *Journal of Language and Politics*, 4(2): 97–226.
- Baker, P. & Gabrielatos, C. (2008). Fleeing, Sneaking, Flooding: A Corpus Analysis of Discursive Constructions of Refugees and Asylum Seekers in the UK Press, 1996-2005. In: *Journal of English Linguistics*, 36 (1): 5–38.
- Grundmann, R. & Krishnamurthy, R. (2010). The Discourse of Climate Change: A Corpus based approach. In: *Critical Approaches to Discourse Analysis across Disciplines*, 4 (2): 125–146.
- Krishnamurthy, R. (1996). Ethnic, racial and tribal: The language of racism?'. In: C. R. Caldas Coulthard & M. Coulthard (eds). *Texts and Practices: Readings in Critical Discourse Analysis*. London, Routledge: 129–149.
- Mautner, G. (2007). Mining large corpora for social information: The case of elderly. In: *Language in Society*, 36: 51–72.