

# Metaphor and Spoken Corpora

Approaches to Metaphor,  
introduction to the GeWiss  
Project: spoken academic  
discourse in contrast.



# Approaches and Importance of Metaphor

# Structure

- ▶ Overview: Approaches to corpora and metaphor
  - ▶ Is metaphor just a matter of language?
  - ▶ How can it be approached using corpora?
- ▶ Introduction to the GeWiss Project
- ▶ My PhD Project
  - ▶ Preliminary results: keywords and metaphors

# What is metaphor

- ▶ Lakoff and Johnson (1980: 4):
  - ▶ Our conceptual system is largely metaphorical
    - ▶ the way we think
    - ▶ what we experience
    - ▶ what we do every day
  - ▶ What we say reveals the metaphors.

# Approaches to corpora and metaphor

- ▶ Previous research: Deignan (2005), Musolff (2004), Semino (2008)
- ▶ Underlying research questions in all those approaches:
  - ▶ How can a corpus be used to enhance or validate existing findings or to discover new metaphors?
  - ▶ What is the specific advantage of a corpus-based analysis of metaphors?
- ▶ Identification of Metaphors in the Corpus (procedures), based on Pragglejaz (2007), Steen (2010)

# Corpus research into metaphor

- ▶ Deignan (2005: 76): corpus can be any collection of texts, processed electronically and manually.
  - ▶ How is it processed? Concordances, collocates etc. She DOES NOT state why to look for which word; she just shows pre-selected examples.
  - ▶ There seems to be no 'special' procedure (see p. 94).
  - ▶ (ibid. p. 85): discussed intuition vs. observation: basically, (based on Sinclair), human beings are **not good** at describing their own or other people's language use.
  - ▶ A corpus allows to predict unexpected uses of words (cf. ibid.).
- ▶ Can findings on the use of metaphor be generalised? See Musolff (2004: 9): notion of 'supra-individual' use.

# Qualitative distribution of metaphors

- ▶ Semino (2008: 22-26): repetition, recurrence, clustering, extension.
- ▶ Examples from my data:
  - ▶ **Repetition:** My Macbook Pro 2.0Ghz with a 7200 drive and 2GB of RAM was running perfect for 3 weeks now and then a night ago it suddenly died. I had been just browsing the web and running it off batteries.
  - ▶ **Recurrence:** i was constantly °h dropping metaphorical (0.2) wooden clogs into (0.4) the technology system (0.5) and i was kind of (0.3) i kind of (0.8) fell into technology backwards (0.4) erm (0.7) there were (0.1) a number of (0.4) reasons (0.3) for (.) my embracing of technology
    - ▶ [physical movement is recurring here]

# distribution of metaphors (continued)

- ▶ **Combination and mixing:** Dropping H264 from YouTube would be suicide for Google.
- ▶ **Signalling:** i had this virtual classroom; we\_re not going to publish (0.2) a workbook (.) the workbook is going to be (1.0) up there (0.1) on the website
  - ▶ with a virtual textbook



# Specific benefit of corpus approaches

- ▶ Baker (2006: 10): corpus approaches help to ensure “a high degree of researcher self-awareness and agency.”
  - ▶ The world is always perceived from a particular viewpoint; research is always constructed.
  - ▶ “By using a corpus, we at least are able to place a number of restrictions on our cognitive biases. It becomes less easy to be selective about a single newspaper when we are looking at hundreds - [...] overall patterns and trends should show through.” (ibid. p. 12).
- ▶ No research is free of bias, but corpora help.

# Methodological Implications

- ▶ Combined approaches of corpus-assisted discourse studies (CADS), see Duguid (2007/2010), Baker (2006) or Walsh (2011)
- ▶ **quantitative methods:** corpus analysis, based on frequency lists, the comparison of word frequencies (keyword analysis) and the search for specific expressions that consist of one or more words (concordance, n-gram) → ‘bigger picture’ / overview.
- ▶ **qualitative methods:** discourse analysis (DA), detailed analysis and discussion of selected examples from the corpora. → closer analysis.

# Metaphor Identification

- ▶ Pragglejaz (2007): MIP = Metaphor Identification Procedure easy, qualitative, straight-forward approach. A whole text is divided into its lexical items and the basic and contextual meaning are contrasted so that the researcher can decide if an expression is metaphorical.
- ▶ Steen et al (2010: 15): MIPVU: complex approach: researcher has to decide if comparisons are 'non-literal' and look up every single lexical item in a specific dictionary and manually annotate the whole corpus.
  - ▶ This procedure cannot be recommended and is too complex for practical use e.g. in my study.

# Example for MIP, Pragglejaz (2007)

First step: segmentation: Does / your computer / misbehave? / Does / it / keep / crashing / or / freezing up?

lexical unit	contextual meaning	basic meaning	metaphor?
Does	perform an action, part of question	same	no
your computer	pronoun + device for storing and processing data	same	no
misbehave	technical faults (hard or software), caused by a machine.	behave in a wrong way (mostly humans).	<b>yes</b>
Does	perform an action, part of question	same	no
it	pronoun, refers to things	same	no
keep	continue to perform an action	same	no
crashing	seize to operate, end programs abruptly, turn off suddenly	violent collision of a moving object	<b>yes</b>
or	conjunction (alternative)	same	no
freezing up	seize operations suddenly	something is cooled down into a low temperature.	<b>yes</b>

# My PhD Project

- ▶ Topic: anthropomorphic metaphors in specialised IT discourse.
- ▶ based on two corpora:
  - a) IT LSP corpus (genres: forum and blog entries) and
  - b) GeWiss corpus of spoken academic discourse (genres: oral exams and presentations).
- ▶ The academic corpus is based on a project about spoken academic discourse in contrast, GeWiss (description / website on handout).

# IT metaphors: examples, source: IT corpus

1. *their code doesn't allow h.264. Uncheck 'Allow the computer to turn off this device ...'* (linguistic metaphor); THE COMPUTER IS A HUMAN BEING (conceptual metaphor)
  2. *Safari / Internet Explorer / Konqueror / Firefox / Office; IT IS WILD NATURE, THE COMPUTER IS A WORKPLACE*  
*Open Office Suite (Allows opening and editing MS office docs)*
- ➔ target and source domains of nature, anthropomorphism and the work place (= office) dominate.

# More examples: keywords

N	Key word	Freq.	RC. Freq.	Keyness
1	#	35117	1604421	55508.80859
2	WINDOWS	3324	8554	21845.9082
3	POSTS	2177	2179	17455.64844
4	QUOTE	1843	1462	15347.28711
5	POSTED	1482	710	13223.61914
6	APPLE	1399	2502	10014.03809
7	JOIN	1679	7282	9523.893555
8	USER	1594	5848	9507.772461
9	GOOGLE	859	0	9260.083984
10	PM	1158	1355	9028.155273
11	ORIGINALLY	1388	4443	8604.695313
12	REGISTERED	1195	3406	7643.528809
13	DATE	1717	17192	7157.875977
14	OS	803	994	6194.552734
15	WEBM	570	0	6144.282715
16	MICROSOFT	878	1780	6108.692871
17	MAC	762	764	6105.754395
18	LINUX	535	0	5766.961426
19	VISTA	483	159	4489.109375
20	REPLYQUOTE	407	0	4387.088867
21	WEB	535	570	4241.42627
22	XP	405	19	4210.565918
23	I	7726	732523	4150.025391
24	PERMALINK	382	0	4117.59082
25	AM	1293	26042	3763.323242
26	SPECS	391	112	3682.193359
27	INTERNET	383	97	3646.111572
28	COM	396	281	3352.1875
29	USB	310	8	3266.805908
30	IPHONE	294	0	3168.979248
31	INSTALL	431	782	3074.419678
32	PRO	455	1081	3047.775391

33	MACBOOK	279	0	3007.287598
34	MY	2453	146775	2769.190674
35	IPAD	249	0	2683.906982
36	APPS	282	89	2631.630859
37	SOFTWARE	660	9395	2329.604736
38	THUNDERBOLT	244	64	2315.809814
39	BIT	961	26501	2270.315918
40	VIDEO	577	6615	2262.396484
41	PATENT	353	998	2262.075928
42	FLASH	374	1338	2246.162354
43	CODEC	215	11	2229.560547
44	HTTP	206	0	2220.401611
45	H	642	9910	2172.429932
46	EXPLORER	261	256	2098.934814
47	LOCATION	445	3977	1945.779419
48	OPEN	918	29600	1925.435669
49	BROWSER	187	30	1841.505127
50	RAM	295	1063	1767.978882
51	PC	356	2332	1756.629517
52	YOUTUBE	162	0	1746.12561
53	APP	233	348	1732.098999
54	DOWNLOAD	177	33	1725.452881
55	FIREFOX	163	5	1711.955322
56	REPLY	413	4223	1704.924438
57	JUN	197	140	1667.177124
58	RE	374	3248	1654.724731
59	IMAC	153	0	1649.115601
60	IOS	159	13	1621.769897
61	INSIKE	149	0	1606.000122
62	BLOG	149	0	1606.000122
63	INTEL	275	1085	1604.723877
64	LAPTOP	175	73	1586.341187
65	HARDWARE	320	2072	1585.65564
66	ROSETTA	160	24	1582.293213
67	LION	266	1014	1568.158813
68	USE	1187	62273	1567.729248

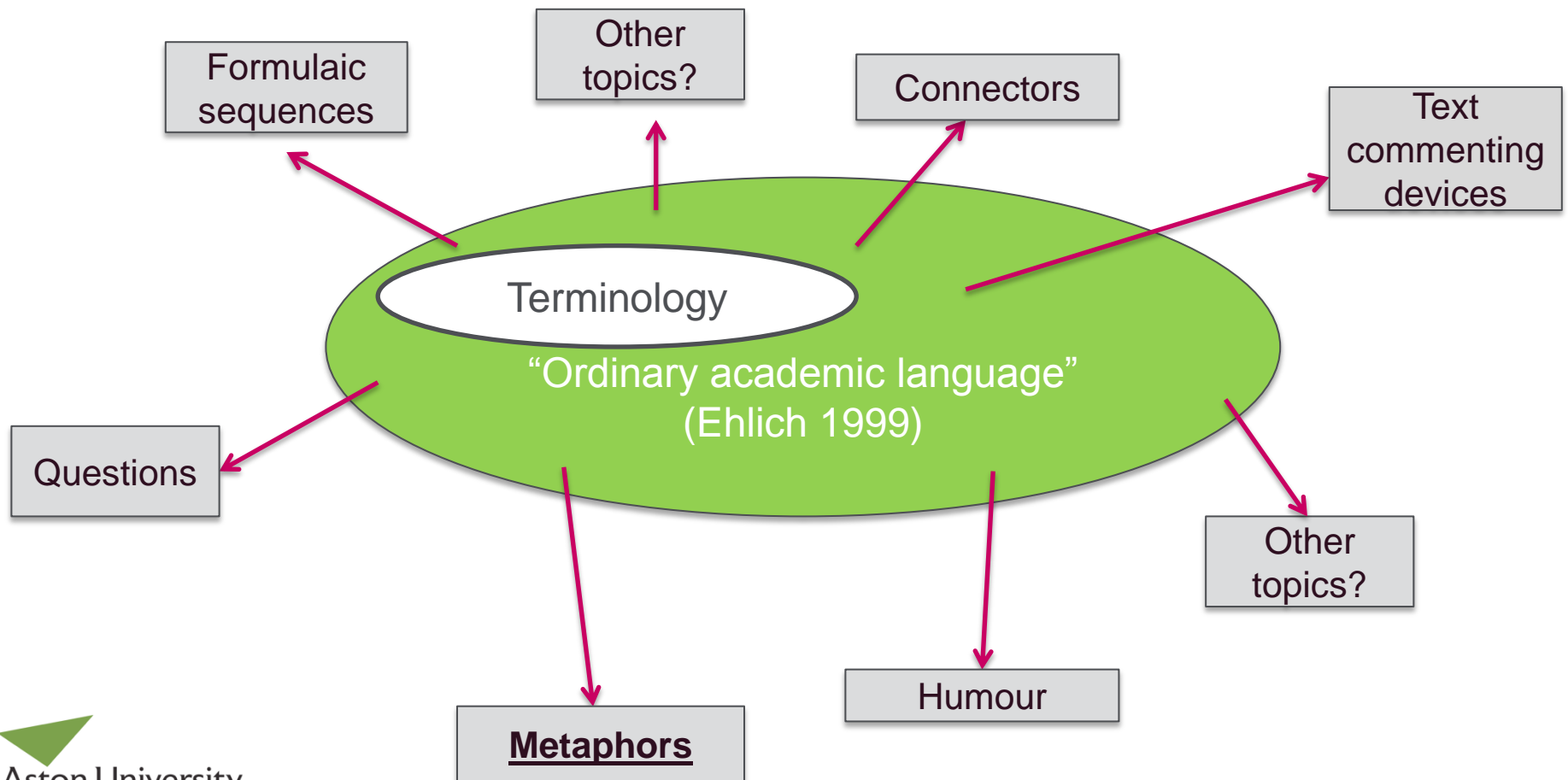
# Results based on keywords

- ▶ 'meta language'
  - Post, join, register, replyquote, permalink
- ▶ Genre-specific
  - #, e.g. the 7 in *Windows 7* or *number of posts: 1*.
  - Operating systems: *Windows, OS, Mac, Linux, Vista, XP*
  - Other: *Apple, Microsoft, Google, iPhone, iPad, flash*
  - Lower keyness (still above 1,000): *explorer, location, browser, youtube*



# Theoretical background to the study of academic discourse

## ► Academic discourse



# Academic corpus: IT metaphors: examples

- ▶ how we go about embracing technology [here: accepting is 'embracing' = physically holding / approaching something]
- ▶ Virtual as identifier of IT-related meanings:
  - ▶ i had this virtual classroom
  - ▶ with a virtual textbook
- ▶ when they hit the submit button (0.6) the (0.7) the technology will assemble (0.3) and download for each individual student (0.5) a workbook (0.3)
- ▶ an example of how technology challenges the notion that (0.6) er (0.7) the classroom is a place where (0.4) learning happens (0.6)

# (preview continued)

1. never happen (0.5) so (0.6) er (0.4) **computers** can be very (0.3) computers are very very patient
  2. of the tu of the session the the (0.3) **computer** automatically archived (0.2) the lesson
  3. (0.5) patient (0.2) tutor as i said (0.5) **computers** don\_t go red in the face and shout (0.5) when the students get things wrong
- Same trend both in academic and IT corpus:  
anthropomorphism = THE COMPUTER IS A HUMAN BEING.

# Concluding Remarks

- ▶ Overall impression: nothing is clear and everything is allowed. There is no procedure specific to metaphor.
- ▶ It is strongly recommended to combine qualitative and quantitative approaches (combined methodologies) when using corpora to study metaphor and/or discourse.
- ▶ Metaphors are worthwhile to be studied because of their high frequency, communicative functions and their overall importance.