Empirical research in data-driven learning: A summary

Alex Boulton, CRAPEL–ATILF/CNRS, Université de Lorraine, France

alex.boulton@univ-lorraine.fr

Updated web supplement to:

Alex Boulton. 2010. Learning outcomes from corpus consultation. In María Moreno Jaén, Fernando Serrano Valverde & María Calzada Pérez (eds), *Exploring New Paths in Language Pedagogy: Lexis and corpus-based language teaching*. London: Equinox, p. 129-144.

The main paper as referenced above surveys 27 empirical studies of 'data-driven learning' which focus on **learning outcomes**, where DDL is taken to refer to any explicit use of corpora for foreign or second language learning. The studies surveyed are part of a larger collection of empirical evaluations of DDL that I have collected in recent years, hence the present supplement.

This supplement is divided into 3 sections:

1. a grid *overview* providing some basic information on each;
2. a short *summary* of each paper;
3. a list of *references* to the source articles.

NB Apologies in advance for inevitable inaccuracies – I would of course be grateful to the authors or others for any correction or further input. The information presented here reflects my own interpretation of the studies: given the constraints of space, it cannot pretend to be complete, and having been collated over time is likely to be a little idiosyncratic in places. My main objective is to highlight the work so far (the repeated complaint of lack of empirical research in DDL suggests that these papers are not necessarily very high profile), and to whet your appetite. So if a particular study interests you, I would strongly recommend that you read the original and do not rely on solely on my review.

This supplement will be updated as more studies come to light at CorpusCALL (http://corpuscall.eu/) and on my homepage (bit.ly/Midi7k) where you can also find pre-prints of many of my own papers. I would be grateful to anyone who could provide any leads to other empirical studies published in the area (in English or in other languages).

1. Overview

Key:

- **study:** the paper(s) reporting each study;
- **country (L1):** the country where it was conducted, along with the mother tongue of the majority of participants;
- **L2 (level):** the target language and current level of proficiency – from beginning or low to advanced, passing through (lower or upper) intermediate where the information can be derived;
- **context (speciality):** the type of institution where it was conducted (e.g. *uni 3* = 3rd year of university study), and the major field of study of most participants;
- **learners:** the number of learners involved (including any control group, indicated in brackets);
- **time:** the duration of the study in hours, weeks, semesters, etc.;
- **interface:** whether the participants used a hands-on concordancer, a CALL program including corpus data, or paper-based materials;
- **RQ (research question):** whether the study evaluates learners' behaviour (B), attitudes (A), using corpora as a reference resource (R), or learning outcomes (L);
- **design / instruments:** the main research instruments (including use of control items or populations);
- **data:** whether some kind of statistical analysis (S) is provided, or only raw figures and percentages, or no quantitative analysis at all (0);
- **aim:** the main point under study – usually a language item, but e.g. *corpus linguistics* refers to an introduction to CL for L2 learners.

study	country (L1)	L2 (level)	context (speciality)	learners	time	interface	RQ	design / instruments	data	aim
1. Baten et al. 1989	Netherlands (Dutch)	English (int+)	uni 2 (economics)	400	4 months	hands-on	B, A, L	project, oral exam, feedback	0	ESP vocabulary, lexical relations
2. Mparutsa et al. 1991	Zimbabwe (Shona?)	English (int+)	uni (economics, geology, philosophy)	variable	variable	hands-on	B A	informal feedback, questionnaires, class discussions, short reports	0	conceptual meaning
3. Stevens 1991	Oman (Arabic)	English (int?)	uni 1 (science)	54	2 sessions	paper	B	post tests	S	vocabulary
4. Ma 1994	Hong Kong (Chinese)	English (int?)	uni	18	10h	hands-on	B R	2 questionnaires, 2 interviews (some students only)	0	writing, error-correction
5. van Halteren 1994	Netherlands (Dutch)	English (adv)	uni 1/2 (L2)	?	2 sems	hands-on	B A	"poll"	0	syntax
6. Gan et al. 1996	Malaysia (Malay)	English (adv)	teacher training (L2)	48	10 wks (20h)	hands-on	A L	pre+post tests, control items, questionnaires	S	vocabulary
7. Aston 1997	Italy (Italian)	English (adv)	uni (translat)	?	?	hands-on	B	class discussions	0	corpus linguistics
8. Cobb 1997a, 1997b, 1999	Oman (Arabic)	English (low)	uni 1 (business)	11 (100s)	1 yr	program	L	pre+post tests & delayed, control pop, questionnaires, tracking	S	vocabulary
9. Johns 1997	UK (Malay)	English (adv)	uni, teachers (L2)	?	6h	program	B	diaries, written feedback, class discussions, tracking	raw n°s	various
10. Maia 1997	Portugal (Portuguese)	English (adv)	uni (translat)	12-16	1 yr	hands-on	B	project reports, class discussions, email feedback	0	translation, corpus linguistics
11. Bowker 1998	Ireland (English)	French (adv)	uni 3 (linguistics)	14	4h	hands-on	A R	translations, questionnaires	S	translation
12. Granath 1998, 2009	Sweden (Swedish)	English (adv?)	uni 1 (L2)	57	1 sem (27h)	hands-on	B A	questionnaires	raw n°s	syntax
13. Turnbull & Burston 1998	Australia (mixed)	English (adv)	uni 4+ (Japanese, managemt)	2	8 mths	hands-on	B A R	observations, interviews, questionnaires, class discussions, written evaluations	0	self-correction

14. Bowker 1999	Ireland (English)	French (adv)	uni 4 (translat)	5	2 wks (2h30)	hands-on	A R	translations, interviews	S	translation
15. Whistle 1999	UK (English)	French (int+)	uni 1/2 (L2)	?	1 term (6h?)	hands-on	B A	class discussions	0	set grammar points
16. Bernardini 2000	Italy (Italian)	English (adv)	uni 3 (translat)	6	8 wks	hands-on	B	class discussions	0	corpus linguistics
17. Seidlhofer 2000, 2002	Austria (German)	English (adv)	uni 1/2 (L2)	10	1 sem?	hands-on	A	informal feedback, teacher evaluations, class discussions, diary grids	0	corpus linguistics
18. Ciesielska-Ciupek 2001	Poland (Polish)	English (low)	high school	33	?	paper	A L	post tests & delayed, questionnaires	raw n°s	various
19. Cobb et al. 2001; Cobb 2006b	Canada (English)	French / German (int)	(adult) (?)	2 (*c*=1)	14h	program	L	pre+post tests	raw n°s	vocabulary
20. Horst & Cobb 2001; Horst et al. 2001	Canada (mixed)	English (int)	uni (mixed)	33	12 wks	program	B L	post tests, control items, rating WordBank entries, questionnaires	S	reading, vocabulary
21. Kennedy & Miceli 2001, 2002	Australia (English)	Italian (int)	uni 2 (L2, mixed)	17	7h	hands-on	B	interviews, questionnaires, class discussions, classroom observations, text revision	0	corpus linguistics
22. St. John 2001	UK (English)	German (beg)	uni 2 (linguistics)	1	20h?	hands-on	B A	task reports, interviews	0	corpus linguistics
23. Todd 2001	Thailand (Thai)	English (int–)	uni 4+ (science, engineer)	23	?	hands-on	R	self-correction	S	self-correction, vocabulary
24. Bernardini 2002	Italy (Italian)	English (adv)	uni 4 (translat)	10	1 sem (15h)	hands-on	B A	observations, class discussions, translations or oral reports on language point	0	translation, corpus linguistics
25. Fan & Xun-feng 2002	Hong Kong (Chinese)	English (adv)	uni 3 (translat)	21	1h40	program	A R	questionnaires, class discussions	raw n°s	reading
26. Hadley 2002	Japan (Japanese)	English (false beg)	uni 1 (economics)	25	1 mth (6h)	paper	A	questionnaires	raw n°s	phrasal verbs
27. Johns 2002	UK (French, English)	English, French (int?)	uni 2/3? (mixed)	16 (8 lang pairs)	3 days	paper	A	reciprocal translations, questionnaires	0	grammar/usage
28. Cheng et al. 2003	Hong Kong (Chinese)	English (adv)	uni 2 (L2)	29	2 wks	hands-on	B	project reports	raw n°s	corpus linguistics
29. Clerehan et al 2003	Australia (mixed)	English (mixed)	uni 1 (computing)	170	4 weeks	hands-on	A B	questionnaires	raw n°s	general
30. Curado Fuentes 2003, 2002	Spain (Spanish)	English (int)	uni 3 (business)	20 (*c*=10)	2 wks	hands-on	L	post tests (oral presentations), control pop, e-mail surveys	raw n°s	collocations, clusters, compounds
31. Lee & Liou 2003	Taiwan (Chinese)	English (int?)	high school	46	10 wks (8h)	hands-on	A L	pre+post tests, questionnaires	S	vocabulary
32. Sripicharn 2003	Thailand (Thai)	English (int+)	uni 2	40 (*c*=22)	4 months (30x15 mins)	paper	L B A	pre+post tests, questionnaires, interviews	S	vocabulary & grammar for writing
33. Sun 2003	Taiwan (Chinese)	English (int / adv)	uni 3/4 (L2/ling)	3	2h	hands-on	B	error-correction, think-aloud, teacher observations	0	error-correction
34. Sun & Wang 2003	Taiwan (Chinese)	English (int?)	high school	81 (*c*=40)	1h40	hands-on	L	pre+post tests, control pop	S	collocations
35. Chambers & O'Sullivan 2004	Ireland (English)	French (adv)	uni 4 (L2)	8	8h	hands-on	B A	self-correction, questionnaires	raw n°s	self-correction, lexico-grammar
36. Curado Fuentes 2004	Spain (Spanish)	English (int+)	uni 3 (tourism)	20 (*c*=10)	2 wks (5h+)	hands-on	L	oral reports, small group discussions	raw n°s	(semi-)technical language in spoken use
37. Gaskell & Cobb 2004	Canada (Chinese)	English (int–)	uni (mixed)	20	1 sem (45h)	hands-on	B A R	questionnaires, diary grids, tracking, writing	S	error-correction, grammar
38. Yoon & Hirvela 2004	USA (mixed)	English (int / adv)	uni 1/4+ (mixed)	22	20 wks? (90h?)	hands-on	B A	questionnaires, interviews, classroom observations, interviews with instructor	S	writing

39. Cargill & Adams 2005	Australia (mixed)	English (int+?)	uni 6+	17	1h30	hands-on	A	questionnaires post-course + 8 months later	raw n°s	writing
40. Chambers 2005	Ireland (English)	mixed (adv)	uni 2 (L2/ling)	14	9h	hands-on	B A	informal feedback, interviews, classroom observations, reports	0	corpus linguistics
41. Chan & Liou 2005	Taiwan (Chinese)	English (int?)	uni 1 (mixed)	32	5 wks (5h)	hands-on	A L	post tests & delayed, control items, questionnaires, tracking	S	verb-noun collocates
42. Frankenberg-Garcia 2005	Portugal (Portugues)	English (adv)	uni 4 (translat)	16	0h50	hands-on	B	questionnaires, translations	raw n°s	translation
43. Kaur & Hegelheimer 2005	USA (mixed)	English (int)	uni (?)	18 (*c*=9?)	4 wks?	hands-on	R L	post tests, control pop, writing	S	AWL vocabulary
44. Tian 2005a 2005b	Taiwan (Chinese)	English (int+)	uni 2/4 (mixed)	98 (*c*=48)	5 wks (10h)	paper	L	pre+post tests, control pop, mixed design	S	grammar, word usage, text type
45. Allan 2006	Ireland (mixed)	English (adv)	language centre (mixed)	18 (*c*=5)	12 wks	paper	B A L	post tests, control pop, informal feedback, questionnaires	S	vocabulary
46. Boulton & Wilhelm 2006; Boulton 2011b	France (French)	English (adv)	uni 4 (L2)	30	1 sem (3 yrs+)	hands-on	B	project papers	raw n°s	corpus linguistics
47. Götz & Mukherjee 2006	Germany (German)	English (adv)	uni 1/3 (L2)	36	1 sem	hands-on	B A	pre+post questionnaires, project reports, observations	S	corpus linguistics
48. Gut 2006	Germany (German)	English (adv)	uni 2/3 (L2)	21	1 sem (15h?)	program	B A	group presentations, questionnaires	raw n°s	phonetic properties of non-native speech
49. Koo 2006	USA (Korean)	English (int+ >adv?)	uni 6+ (mixed)	10	2 sessions (1 wk)	hands-on	B A R	tracking, stimulated recall, productions	raw n°s	paraphrasing
50. Koosha & Jafarpour 2006	Iran (Farsi ?)	English (adv)	uni 4 (L2)	200 (*c*=100)	1 sem	paper	L	pre+post tests, control pop	S	collocations of prepositions
51. Lee & Swales 2006	USA (Chinese)	English (adv)	uni 5+ (mixed)	4	13 wks (20h +)	hands-on	A	interviews, class discussions, project reports	0	corpus linguistics
52. Liou et al. 2006	Taiwan (Chinese)	English (adv?)	uni 1 (?)	varied	variable	program	B A R L	post tests & delayed, questionnaires, tracking	S	various pilot experimentation
53. O'Sullivan & Chambers 2006	Ireland (English)	French (adv)	uni 2 (L2/ling)	14	11h	hands-on	R	self-correction, questionnaires	raw n°s	self-correction
54. Schmied 2006	Germany (German)	English (adv)	uni (?)	?	(on-going)	program	B A	tracking, post tests, interviews	0	grammar
55. Boulton 2007	France (French)	English (int–)	uni 1 (engineer)	104 (*c*=51)	0h30	paper	A R	post tests	S	*will / going to*
56. Braun 2007	Germany (German)	English (int?)	high school	25 (*c*=13)	4 wks (16h)	program	A L	post tests, control pop, questionnaires, class discussions, teacher observations	S	lexical & grammatical points from coursebook
57. Cresswell 2007	Italy (Italian)	English (adv)	uni 3 (translat)	126 (*c*=65)	1 sem?	hands-on	B L	post tests, control pop, informal feedback, interviews, oral presentations, writing	S	connectors
58. Curado Fuentes 2007	Spain (Spanish)	English (int+)	uni (tourism)	20 (*c*=10)	5h	hands-on	L	post tests, control pop, email tasks	S	reading
59. Estling Vannestål & Lindquist 2007	Sweden (Swedish)	English (adv)	uni 1 (teachers, admin)	a) 37 (*c*=23); b) 35	1 sem each	hands-on	A L	pre+post tests, control pop, questionnaires, interviews	raw n°s	grammar & perceptions of grammar
60. Hafner & Candlin 2007	Hong Kong (Chinese)	English (adv)	uni 4 (law)	300 (9 case study)	1 yr	hands-on	B A	tracking, interviews	raw n°s	various
61. Hirata & Hirata 2007	Japan (Japanese)	English (int–)	uni (?)	69	4 wks (6h)	hands-on	A	questionnaires	raw n°s	reading, lexis
62. Huang & Liou 2007	Taiwan (Chinese)	English (int)	uni 1 (L2?)	38	12 wks (out of class)	program	L	pre+post tests, concordance tracking, questionnaires	S	vocabulary acquisition through reading
63. Lavid 2007	Spain (Spanish)	English (adv)	uni 4+ (linguistics)	?	1 sem?	hands-on	A	informal feedback, class discussions	0	mental transitivity
64. Sun 2007	Taiwan (Chinese)	English (adv?)	uni 4+ (technical)	20 (3 ints)	?	program	B A	tracking, questionnaires, interviews	S	writing

65. Yeh et al. 2007	Taiwan (Chinese)	English (int+)	uni 1 (L2)	19	4 wks (1h20)	hands-on	B A L	post tests & delayed, writing, questionnaires, tracking	S	adjectives
66. Belz & Vyatkina 2008, 2005a, 2005b	USA (English)	German (adv)	uni (?)	2	8 wks (4h)	hands-on, paper	B L	pre+post tests, observations, written production	raw n°s	modal particles / *da*-compounds
67. Boulton 2008a	France (French)	English (int–)	uni 1 (engineer)	113	0h30	paper	R	post tests	S	phrasal verbs
68. Boulton 2008b, 2010a	France (French)	English (int–)	uni 2 (architect)	62	1h	paper	A L	pre+post tests, control items, questionnaires	S	grammar / usage
69. Johns et al. 2008	Taiwan (Chinese)	English (int?)	high school	22 (*c*=11)	16 wks (48h)	program, hands-on, paper	A L	post tests, control pop, questionnaires, school exams	S	reading comp
70. Lin 2008	Taiwan (Chinese)	English (adv)	uni 3 (L2)	25	8 wks (13h20)	program	A L	pre+post tests, delayed writing, questionnaires	S	AWL vocabulary
71. Smith et al. 2008	Taiwan (mixed)	Chinese (int?)	internet volunteers (?)	25 (2 post-tests)	6 wks	hands-on	A L	pre+post tests, on-line discussions boards	raw n°s	word meaning, usage
72. Thorne et al. 2008	USA (mixed)	English (adv)	uni 4+ (ITAs)	53	?	paper	L	spoken usage	raw n°s	directives, speaking
73. Yoon 2008	USA (Korean)	English (adv)	uni 4+ (mixed)	6	10 wks (50h?)	hands-on, paper	B A R	observations, interviews, recall, tracking, assignments, reflection	0	writing
74. Boulton 2009a	France (French)	English (int–)	uni 1 (engineer)	132 (*c*=64)	0h30	paper	R L	post tests, control pop	S	connectors
75. Boulton 2009b	France (French)	English (int–)	uni 2 (architect)	34	12 wks (3h)	hands-on	A	questionnaires, Index of Learning Styles	S	learning styles
76. Chang & Sun 2009	Taiwan (Chinese)	English (int?)	high school	26	3h (4 wks)	hands-on	L A	pre+post tests, comparison groups, questionnaires	S	scaffolding in proof-reading of collocations
77. Chujo et al. 2009	Japan (Japanese)	English (beg)	uni 1 (engineer)	22 (*c*=25)	20 wks (10h)	hands-on	L A	pre+post tests, control pop, questionnaires	S	TOEIC vocabulary, grammar
78. Gilmore 2009	Japan (Japanese)	English (int)	uni 2	45	1h30	hands-on	R A	revisions, questionnaires	raw n°s	error-correction
79. Liu & Jiang 2009	USA, China (mixed)	English (int+)	uni 2/4 (EFL / ESL)	244	1 sem	hands-on	A	questionnaires	S	various
80. Schaeffer-Lacroix 2009	France (French)	German (int–)	junior high school	33 (final phase)	3 years (total)	hands-on	A B R	analysis of revisions, interviews	S	various
81. Tyne 2009	UK (English)	French (int?)	uni 3 (linguistics)	10	1 yr	hands-on	A	questionnaires	raw n°s	spoken corpora
82. Varley 2009	New Zealand (Chinese)	English (adv)	uni (L2)	19	1 sem (56h)	hands-on	B A	questionnaires, projects, reflective log, discussions	raw n°s	corpus linguistics, genre
83. Boulton 2010b	France (French)	English (int–)	uni 2 (architect)	29	1h	paper	B	Index of Learning Styles	S	learning styles
84. Chang 2010	Korea (Korean)	English (int / adv)	uni (mixed)	28	2h	hands-on	A	questionnaires	raw n°s	writing
85. Conroy 2010	Australia (mixed)	English (int+, adv)	uni (mixed)	165	1-4h	various	B R	questionnaires (53); some interviews, logs, writing	raw n°s	writing, error-correction
86. Kennedy & Miceli 2010	Australia (English)	Italian (int)	uni (L2)	3	0h45	hands-on	A B R	revisions, interviews, questionnaires	0	creative writing
87. Landure & Boulton 2010	France (French)	English (low)	uni	43	2h	hands-on	B R	revisions, questionnaires	raw n°s	error-correction
88. Moreno Jaén 2010	Spain (Spanish)	English (int+)	uni 3- (L2)	21	3 weeks	hands-on	L A B	pre+post tests, questionnaires	S	collocations
89. Okamoto 2010	Japan (Japanese)	English (low)	uni 3/4 (IT, engineer)	15	15 weeks	hands-on	R	questionnaires	raw n°s	writing
90. Park & Kinginger 2010	US (Chinese)	English (adv)	uni 1 (business)	1	1h49	hands-on	B	tracking, screen recordings, retrospection	raw n°s	writing
91. Rapti 2010	Greece (Greek)	English (int)	high school (age 13-15)	28 (*c*=14)	5 months (1h40 / wk)	paper (& hands-on)	A L	questionnaires, pre+post tests, interviews, observation	raw n°s	grammar (verbs)

92. Wu et al. 2010	New Zealand (mixed)	English (int?)	uni	9	2h	program	R	revisions, tracking	raw n°s	error-correction
93. Acar et al. 2011	Japan (Japanese)	English (?)	uni (engineer)	8	1 week	hands-on	R	writing (pre/post)	raw n°s	writing, error-correction
94. Al-Lawati 2011	Oman (Arabic)	English (int–)	uni 1 (L2?)	25	6h (7 wks)	paper	B	think-aloud	0	strategy use
95. Boulton 2011a	France (French)	English (int–)	uni 2 (architect)	59 (*c*=25)	12 wks (3h)	hands-on	L	post tests & delayed, control pop	S	noticing skills
96. Chang & Kuo 2011	Taiwan (Chinese)	English (int)	uni (computing)	23	1 semester	program	R A	pre-post assignments, questionnaires	S	writing, structure
97. Charles 2011	UK (mixed)	English (adv)	uni 4+ (mixed EAP)	44	6 wks (12h)	hands-on	A	questionnaire	raw n°s	discourse analysis, rhetorical functions
98. Chen 2011	Taiwan (Chinese)	English (int)	uni (?)	51	0h40	hands-on	R A	translation, interviews	S	translation, collocates
99. Gao 2011	Taiwan (Chinese)	English (int+)	uni 2 (engineer)	21	(1 class?)	hands-on	R A	pre+post tests, questionnaires	S	translation, error-correction
100. Kaszubski 2011	Poland (Polish)	English (int?)	uni (corpus linguistics / EAP)	12	own time / 1 sem	hands-on	B A	tracking, questionnaires	raw n°s	pursuing autonomous queries
101. Kettemann 2011	Austria (German)	English (int+)	uni 1 (L2)	10	1 wk (own time)	paper	A B	tasks, questionnaires	0	cultural awareness
102. Kettemann & Marko 2011	Austria (German)	English (adv?)	uni	17	6x1h	hands-on	B A	diaries, questionnaires, reports	0	critical discourse analysis
103. Landure 2011	France (French)	English (mixed)	unemployed (computing)	44	2x2h	hands-on	R B L A	observations, questionnaires, post-test	raw n°s	reading, vocabulary
104. Pérez-Paredes et al. 2011	Spain (Spanish)	English (adv)	uni 2 (L2)	37 (*c*=15)	2x1h	hands-on	B	tracking	S	grammar, writing
105. Philip 2011	Italy (Italian)	English (adv)	uni	41 (*c*=10)	1 session	hands-on	B R	observation, controlled writing	0	writing, comparing resource use
106. Rodgers et al. 2011	Ireland (Irish)	French (int?)	uni 2/3 (biotechnology)	19	20h (10 wks)	hands-on	A	questionnaires, interviews	raw n°s	grammar, topic vocabulary
107. Smith 2011	Taiwan (Chinese)	English (int)	uni 1 (humanities)	19 (/90)	1 sem	hands-on	A B	comments	raw n°s	corpus building
108. Boulton 2012a	France (French)	English (adv)	uni 3 (L2)	12	1 sem	hands-on	A R	translations, exams, questionnaires	S	translation, meta-awareness
109. Boulton 2012b	France (French)	English (int–)	uni 2 (architecture)	40	3h (10 wks)	paper, hands-on	L A	test, questionnaires	S	paper- vs computer-based DDL
110. Chang 2012	USA (Chinese)	English (adv?)	uni 6+ (soc sciences)	7	5h (5 wks)	hands-on	B L	pre+post tests, self-analysis	S	stance & move in writing
111. Charles 2012	UK (mixed)	English (adv)	uni 4+ (mixed)	50	6 wks (12h)	hands-on	A	questionnaires		discourse + grammar for EAP writing
112. Chujo & Oghigian 2012	Japan (Japanese)	English (beg)	uni (engineer)	62 (*c*=23)	2 sem (30h)	paper, hands-on	L A	pre+post tests	S	noun & verb phrases, vocabulary
113. Frankenberg-Garcia 2012	Portugal (Portuguese)	English (adv)	uni 2 (tourism)	48 (*c*=12)	1 class	paper	R	post-tests	S	lexis: production & reception
114. Geist & Hahn 2012	Germany (German)	English (int)	high school	53	3h45 (5 wks)	hands-on	B	diaries, observation, productions diaries, observation, productions	0	writing
115. Osolsobě & Vališová 2012	Czech Rep (mixed)	Czech (int– > adv)	uni	20	90 mins (1 class)	hands-on	A B	observation, questionnaires	raw n°s	word forms
116. Geluso Forthcoming	Japan (Japanese)	English (int)	uni 1 (communication)	25	2h (3 classes)	hands-on	R	productions	S	formulaic sequences, writing

2. Short summaries

- **Baten et al. 1989.** After an initial introduction, 400 second-year Dutch economics students of English were left to their own devices to select a business topic and texts, choose 'keywords' for comparison with concordances in the library Brown Corpus and produce diagrammatic representations of the relationships over a 4-month period. Their written work and a prepared oral exam on it showed remarkable sophistication in corpus use, though spontaneous conversation with 20 students showed transfer of keywords but not so much of collocates. Feedback is overwhelmingly positive regarding motivation and the possibilities for individualisation and autonomy in contextualised vocabulary work; negative points relate mainly to the limitations of the corpus and availability of the concordancer.

- **Mparutsa et al. 1991.** This paper reports on 3 case studies in different disciplines with variable number of learners ranging from a single volunteer (for peer tutoring) to 27 students sharing 4 computers. The focus is generally on conceptual meaning of specialist terminology from small textbook corpora; initial tasks were inevitably highly controlled as some learners had no previous experience of computers, achieving remarkable freedom in later sessions. Students claim greater confidence and show remarkable enthusiasm, even opting for extra sessions.

- **Stevens 1991.** Stevens compared students' performance in tests using a traditional gap-fill text or multiple concordances, i.e. for testing purposes only. Although there was no control group, each participant completed half the questions using traditional gap-fills, half using concordances. The concordance questions were answered significantly better in the second session, perhaps suggesting the importance of training (although all students had received some training in advance of the first session). Various explanations are discussed, but it seems that multiple concordances do elicit better scores, and perhaps contribute to the learning process.

- **Ma 1994.** 18 part time students followed a 10-week course in computing in Hong Kong, including this 10-hour English module with an aim to writing computer manuals. The corpus was created specially for the course: 52K words of start-up/installation chapters from 14 manuals, as well as a smaller corpus (size unknown) of complete "documentation files". The students' writing was marked for errors to check against the corpus. Data consisted of detailed "concordance diaries"; two questionnaires (half-way through and at the end) plus interviews with randomly selected students. The search queries are analysed and compared against the other data; despite some success, the main conclusion is that left to their own devices, students do not use their time efficiently, formulating inappropriate queries and underusing some functions, and concentrate on language use for error-correction only.

- **van Halteren 1994.** Students use a parsed treebank corpus to explore English syntax for linguistic purposes, according to course guidelines. Students claim they think they are doing more and are more actively involved than in the earlier traditional programme, but are mixed in whether they are learning more or not about syntax, partly as they have less time for explanations.

- **Gan et al. 1996.** As part of their pre-service training, 48 future teachers learned vocabulary through both concordancing and traditional instruction, with 5x 2-hour classes on each approach. Although some were unfamiliar even with a computer keyboard, only 2 hours of training was provided. The pre+post-test design showed significant improvement favouring the experimental treatment, and questionnaires showed a preference for the experimental teaching, especially working in pairs, though the majority thought a combination of both methods would be best.

- **Aston 1997.** An initial trial suggested that learners seek to find generalised language rules in corpus data, but do not pick up on patterns and tendencies. A second trial encouraged learners to look first for non-linguistic information in medical research articles. Small groups of learners successfully pursued their own questions in browsing mode, reported their findings clearly to others, and derived procedures for tackling texts in the future.

- **Cobb 1997a, 1997b, 1999.** This large study describes an in-house program in Oman including concordance data for vocabulary learning to help with reading. Several hundred 1st-year business students used the software, based on 2,000 frequent words in English, with the aim of maximising vocabulary learning; control groups used word lists and dictionaries. The main analysis focuses on one group of 11 lower-level learners: their average gain was 430 words in 3 months using the VLT, with the experimental group outperforming the control group on weekly quizzes. Overall, the experimental group rated the materials highly, and did better on definitional knowledge and transfer to new texts, both short and long-term.

- **Johns 1997.** News and popular science texts are available as 'remedial grammar' for international students using Context incorporating quizzes and pre-selected 10-line concordances. The first 50 case histories are reported, followed by feedback from Malay teachers. The quiz approach is found to encourage students to select the 'difficult'

option, and to attempt the quiz without looking at the evidence, leading to a high rate of abandonment. The teachers highlight how important it is to persevere, but note the limits of computer-correction (only specified answers may be considered correct), though they remembered many details even 2 months later.

- **Maia 1997.** The students in this translation course compiled thematic corpora (monolingual, comparable) for language learning and translation purposes, using various resources. Students chose from a list of themes for their project corpus with a view to creating translation glossaries, and the qualitative analysis details each.

- **Bowker 1998.** 14 students on a translation course (French/English) had already been introduced to corpus use; they translated 2 specialist texts into their L1, one using an L1 ESP corpus, the other using only traditional reference sources. Analysis of their translations showed the corpus-aided translations contained significantly fewer comprehension errors, and fewer production errors of all types (though not all significantly so). Learners were generally positive to corpus use (for comprehension, finding and choosing terms and confirming choices), though some prefer the familiarity of traditional grammars and dictionaries; there was some frustration at the time spent using corpora.

- **Granath 1998, 2009.** In the first paper, 70 students analysed printed concordances, performed pre-set queries, chose and performed their own research in the lab. Questionnaires show that most learners appreciated the more closed exercises, but did not see the point of the hands-on work. No short-term effect was found on their linguistic performance, although it is suggested that language awareness may have increased, producing longer-term benefits. The follow-up paper reports on the same course until 2006: although time in the computer room is considerably reduced to only 2 hours, students' reactions have become more positive, in particular with increasing numbers finding it easier and useful (though the majority still find it not very useful), and expecting to use corpora in the future.

- **Turnbull & Burston 1998.** 2 students enrolled for a master's degree in Australia submitted course assignments, and errors were then highlighted using MS Word revision. They were shown how to use concordancers to correct their errors; the small scale of the study means that each learner's concordancing activities can be reported in depth. Radically different motivations and learning styles meant that one (field independent) student used the concordancer frequently and effectively, and felt she learned a great deal, while the other (field dependent) showed less interest and was less successful, and found it a waste of time.

- **Bowker 1999.** 5 MA translation students translated a scientific text from French into their L1, using either traditional resources (anything they could find on campus) or a corpus. 2 weeks later, having received no feedback, they translated the same text again, but using the other resource type. Analysis of the translations shows that corpus resources resulted in significantly fewer errors, while interviews reveal that student translators tend to have blind faith in dictionaries and their general language knowledge, but few are willing to use corpora spontaneously.

- **Whistle 1999.** More or less detailed worksheets were designed to help learners through hands-on concordancing to explore each of the assigned grammar points. The programme was suspended after one term as it was apparently unpopular with a majority of students, especially because formulating their own search queries was time-consuming, laborious and boring; they would have preferred prepared concordances. Benefit was found to be limited to students with a more solid understanding at the start.

- **Bernardini 2000.** The learners were introduced to corpora, and required to analyse set questions on their own out of class. Reporting back, their conclusions were generally found to be correct, and their enthusiasm is clear as they reported using the BNC for other courses too. There was however some frustration, especially as they had difficulty generating and refining search strings.

- **Seidlhofer 2000, 2002.** To improve on an earlier, disappointing course, this cohort of students examined a corpus of texts of their own writings. Following a number of introductory sessions, students explored each other's texts using WordSmith Tools. Despite the small corpus of very short texts, learners were able to discover many of the elements of corpus linguistics and compare against larger corpora. Allowing learners greater personal involvement in the corpus and in the questions to explore led to far greater enthusiasm and motivation.

- **Ciesielska-Ciupek 2001.** A group of 33 secondary-school learners of mixed levels supplemented their course book with internet materials and concordance print-outs. Tests on language items covered were positive, and maintained 4 weeks later with no further revision. The students rated these resources as a whole increasingly highly in terms of interest and motivation, preferring them to traditional materials, although they found them difficult.

- **Cobb et al. 2001; Cobb 2006b.** One adult intermediate learner of French read a short story by Maupassant 3 times (twice with sound) using R-READ, an interface which allows automatic dictionary look-up as well as concordances from the entire Maupassant corpus (1m words) by clicking on any item, and creation of WordBank entries on-line; another

read a German novella unassisted. The learner was tested using the VLT before the experiment and after each reading. Test **results** show dramatic increase in the number of known words i the assisted format (153 definitely or possibly known), substantially more than for the unassisted (only 11 new words entering the 'possibly known' category), and the differences cannot be attributed to time on task.

- **Horst & Cobb 2001; Horst et al. 2001.** Dictionary use and concordancing for vocabulary learning was compared among 33 intermediate-level Asian and Romance speakers who needed English for university study in Canada. Learners could choose either resource in creating collaborative Word Banks on line, with the resulting entries generally being deemed very satisfactory in terms of both definitions and examples. Vocabulary gains were significant but generally small, which is attributed to problems in using a general vocabulary test to measure learning of specific items. A questionnaire showed that concordancers were used less frequently than bilingual, monolingual or computer dictionaries, but that concordancing was the strongest predictor of vocabulary learning.

- **Kennedy & Miceli 2001, 2002.** Learners proceeded from controlled activities, to correcting their own work, to exploring their own questions; most found it useful, although there was some discouragement in working with complex data. 10 students' work was examined in detail, and difficulties outlined in the 4 main stages of: formulating the question; devising a search strategy; observing the examples found and selecting relevant ones; drawing conclusions. The project is still seen as useful, with further training as the main way forward.

- **St. John 2001.** The single learner in this study is claimed to be a beginner, although an exemplary linguistics student who liked grammar and had already been learning German for a year at university. The student attempted 9 of 17 set tasks using corpora and showed considerable creativity, autonomy and was generally successful in their completion; motivation and confidence (and success) increased over the course.

- **Todd 2001.** 2 errors were indicated in reports submitted by 23 postgraduate students, who used the internet to generate concordances and induce rules. In 80% of cases their rules matched the concordances; 70% also matched traditional reference sources. For 20/23 cases, their corrections matched their rules, resulting in 18 valid corrections. Items with many meanings or patterns of usage were found to be the most difficult; more surprisingly, adjectives were found easier than verbs, in turn easier than nouns. Overall, the study lends strong support to learners' ability to induce patterns from self-generated concordances.

- **Bernardini 2002.** After the theoretical introduction, the 10 volunteers were keen to try hands-on work. This started with controlled activities using the BNC, becoming freer and taking more autonomy with specific corpora and then translation/parallel corpora. Class discussion following oral presentation of language points revealed mainly positive reactions (in control, autonomising, relevant, transferable, etc.), with the few negative reactions relating to lack of time.

- **Fan & Xun-feng 2002.** Students were tested on comprehension of 2 legal texts, assisted by a bilingual corpus allowing them to compare translations or click for concordances of key items. Questionnaires and videoed interviews show positive reactions, with all but one student finding the concordances and bilingual hyperlinks at least moderately useful. Comprehension scores were mixed, however, showing that bilingual corpora present their own difficulties.

- **Hadley 2002.** 25 Japanese university students, allegedly beginners in English, spent 4 classes working on the COBUILD Sampler for phrasal verbs; the teacher monitored as they completed the exercises on their own. Questionnaires showed they found grammar important but unappealing before the experiment, but were subsequently motivated by the authentic nature and discovery processes involved in the concordance activities, even if these were found difficult.

- **Johns 2002.** In 3 sessions on consecutive days, British engineering students of French and French students of various subjects, all at Birmingham University, were paired to analyse worksheets for particular problems: translations of *dont*, French conjunctions requiring the subjunctive, and translations of phrasal verbs with *out*. Evaluation by means of a questionnaire was generally positive, with suggestions for improvement.

- **Cheng et al. 2003.** 2 weeks of this 2-semester course were devoted to corpus linguistics; here they were required to design and write up a short research project for assessment. The variety of the topics covered attests to their enthusiasm; their own evaluations show that the vast majority found the work interesting and useful. However, 3/4 admitted encountering difficulties in various stages, including conceiving their project, formulating queries, selecting and analysing data, making sense of truncated concordances, etc.

- **Clerehan et al. 2003.** A general on-line resource including on-line dictionaries and a concordancer was introduced to undergraduates in one hour, after which they were free to use it as they saw fit. The data include 170 surveys returned from 240 first-year computing students four weeks later, 57% of them non-native speakers of English. Less

than 20% of the students said they used the site more than 5 times, and 60% did not use the concordancer at all, probably as it had not been demonstrated to them. Of those who did, 5% did not find it helpful; 45% neutral (interpreted as "of some use"); 50% helpful or very helpful; the figures for local and international students are comparable. Simply making resources available is clearly not enough; but if even just one in five of the total population found a concordancer useful with no training at all, this might be enough to justify its inclusion.

- **Curado Fuentes 2003, 2002.** 2 groups of 10 learners prepared a 5-minute oral presentation on a business topic, the experimental group being trained in corpus use for this; the language focus was on clusters, collocations and compounds. Analysis of the videoed presentations shows the experimental group making rather more mistakes, but also considerably more effective use of the target points. A follow-up survey suggests the learners find the corpus tools more useful for semi-technical than technical items, grammar or discourse elements.

- **Lee & Liou 2003.** 46 high school students in Taiwan were introduced to concordancing using the VLC web concordancer. 50 minutes of class time over each of 10 weeks was given over to completing activities and exploring web concordances for vocabulary. Students were divided into 3 levels according to 3 separate measures; the same tools in post-test showed all levels improving, with the differences between them being reduced (although the differences were not significant). An unidentified pre-experiment questionnaire was designed to distinguish those with a preference for inductive and (the majority) deductive approaches to language learning; again, there was no significant difference in post-test performance, but those with a preference for induction performed better. A post-test questionnaire showed that both types were open to concordancing, with the major problems residing in technical aspects and user-unfriendly interface.

- **Sripicharn 2003.** 18 upper-intermediate students of English in Thailand used 30 paper-based DDL activities for 12 minutes of each class of a 4-month writing course. They scored slightly higher than the control group in the post-test, though the difference was not significant – perhaps because pre-test scores were already high, and the control and experimental situations not sufficiently different. Questionnaires show considerable enthusiasm for the approach to vocabulary in particular even in this essentially deductive environment; 6 in-depth interviews with each of 6 students on specific units allow insights to the processes involved.

- **Sun 2003.** After 30 minutes' introduction to corpus techniques, the learners were given about an hour to correct errors in a list of 8 sentences. They used the corpus to correct these under teacher guidance, a think-aloud protocol recording their actions. 4 factors influencing success are identified: prior knowledge of the language point (learners used the corpus successfully to confirm intuitions, but had more difficulty in exploring new points); cognitive skills (4 main stages: comparing, grouping, differentiating, inferring); teacher intervention (guidance was essential where prior knowledge was limited); concordancer skills (teacher guidance was again often essential, though presumably could be reduced with experience).

- **Sun & Wang 2003.** 81 second-year high-school students in Taiwan studied collocations using either an inductive approach from concordances, or deductively in a control group using traditional methods. Only 4 items were tested: 2 easy and 2 difficult collocation patterns. While the control group was provided with rules, the experimental group searched for the target patterns to induce their own rules; performance was measured by asking the learners to correct errors in sentences before and after the experiment. The experimental group showed greater improvement in the collocations tested, significantly so for the apparently easier items.

- **Chambers & O'Sullivan 2004.** 8 MA students of French wrote an assigned 600-word commentary with access to dictionaries, and errors were underlined. After a 6-hour introduction to corpus techniques, they were allowed 2 hours to use concordancers to improve their text. Improvement was noted in all areas, surprisingly even on areas (such as gender) which could have been checked more quickly in traditional resources. This indicates the motivation of the learners involved: reactions were generally positive and potentially useful, with a majority claiming they would use corpora in the future. Negative reactions were also gathered.

- **Curado Fuentes 2004.** Upper-intermediate students followed a third-year course for tourism based on spoken corpora of professional and academic English. The two tasks consisted of oral reports (monologues, prepared but no notes) and discussions (groups, spontaneous). The experimental group showed considerably more 'effective' use of structures that occurred in the corpus (37 vs. 18), semi-technical collocations (18 vs. 8), phrase clusters (15 vs. 10) and technical noun+noun combinations (4 vs. 0) in Task A, as well as more 'effective' marking elements (15 vs. 3) in Task 2; they also showed "greater confidence than the control group in the spontaneous speech task" (p22). A similar number of 'errors' can be attributed to the experimental group's faster pace of delivery and longer presentations.

- **Gaskell & Cobb 2004.** Errors were indicated on students' essays on this writing course: for the first 4 essays, concordances were prepared for each student for each of 5 errors and made available on-line; after this students

were expected to create their own concordances for self-correction. Questionnaire responses are positive: learners feel they have improved their writing skills and grammar on the points covered (though do not necessarily attribute this to concordances). Tracking shows the students did make use of the pre-cast concordances, but were less keen on making their own; these generally resulted in an appropriate revision, even for their own concordances. Over 1/3 of the learners became independent concordance-users and claimed they would continue in the future. Learners may well have integrated specific language points, but error types did not significantly decrease over the course as a result of concordancing; recommendations include longer training and a longer time-scale for such results to be seen.

- **Yoon & Hirvela 2004.** Following an extensive 4-week introduction to corpus techniques, learners mainly worked in their own time, with feedback during class. A detailed questionnaire was complemented by hour-long interviews with one positive and one negative student in each group. Learners were generally positive, the overwhelming majority claiming they would use corpora in the future. They found corpora most useful for lexical usage and phrases (preferring dictionaries for meaning), as well as writing (the skill most focused on in class). The intermediate learners seemed more favourable than the advanced group, perhaps as they had received more guidance, considered crucial here. The study also reports some negative reactions, mostly of frustration, the perceived time-wasting, and difficulty of interpreting truncated concordances.

- **Chambers 2005.** Students are from a variety of degrees, hence corpora cover several languages. Evaluation is based on learners' research reports; 5 case studies are provided. The final section of the report requires students to evaluate the activity. The reactions are generally positive, particularly for the large number of 'examples', although there were also strong views on the disadvantages; these tend to limit the role of corpus studies compared to grammar books and dictionaries, draw attention to the need for more training and practice, and categorise the analysis process as tedious and time-consuming

- **Chan & Liou 2005.** Following an introduction to corpus techniques, 32 Taiwanese college students were given 5 web-based treatments of verb-noun collocations, 3 inductively using TOTALrecall (a Chinese-English bilingual concordancer), 2 without; each unit was accompanied by 1 hour of various on-line exercises. An immediate post-test showed significantly greater improvement for the items using concordancing; a delayed post-test 10 weeks later showed a decrease, although the results were still significantly higher than for the pre-test, with the experimental treatment still significantly better. It is argued that an inductive approach takes time to produce its maximum effect. There was no significant correlation between level and post-treatment scores; a questionnaire showed generally favourable attitudes, which did correlate with performance.

- **Frankenberg-Garcia 2005.** Advanced student translators translated a short text into the L2 in the computer room, completing a grid to show resources used (various dictionaries, corpora, the internet, etc.); these were compared against the translations themselves for each item. Look-ups were generally perceived as successful, even though many still contained errors. Preferred references were term banks and dictionaries, accounting for 3/4 of all look-ups; monolingual and bilingual corpora accounted for about 1/10. However, translations showed that the preferred references were used with least success, while corpora and internet search engines were useful in about 2/3 of cases.

- **Kaur & Hegelheimer 2005.** 18 ESL students in the US were introduced to concordancing over 2 weeks, and completed a vocabulary pre-test on 30 AWL items. They then completed vocabulary exercises on the target items (cloze and sentence-building) in class, and a writing exercise out of class; the control group was allowed access to an on-line dictionary, while the experimental group also had access to the BNC written section using LexTutor. The vocabulary results show the concordance group performing better, though not significantly so, with no apparent correlation between results and concordance use (they used the dictionaries more frequently). On the other hand, they used the target items significantly more frequently and more accurately in the written assignment.

- **Tian 2005a, 2005b.** Students from various disciplines enrolled in an English news media class in Taiwan were divided into 2 large classes, 2 hours per week over 5 weeks to study grammar points, word usage, and text type. The experimental group were given print-outs of the target items from on-line news sources, along with basic guidelines for induction; the control group were given rules and examples from grammars and media textbooks. The different question types for each category in the pre- and post-tests may compromise the results, alleged to show that both corpus work and traditional teaching are successful, with corpora significantly more useful for work on grammar and text type. Proficiency was not a significant factor as measured by a local test for any of the tasks; students were only considered high or low, although a presumably normal distribution may be responsible for evening out differences.

- **Allan 2006.** The 18 students were enrolled for an exam preparation course for the CAE in Ireland. The experimental group worked with printed concordances from a 100K-word newspaper corpus over twelve weeks, with tasks assigned out of class. Questionnaires show considerable variation in the number of concordance tasks completed and the time

spent on them, though most rated them useful and interesting, and intended to use them in the future. The VKS for self-assessment showed the experimental group making greater gains than the control of 5 students – and not just on the items taught, which is taken to show that the benefits of corpus consultation can extend to other items.

- **Boulton & Wilhelm 2006; Boulton 2011b.** These French learners were enrolled for the linguistics option of a distance MA in English, assessed on the basis of a 20-page research paper based on an aspect of corpus linguistics of their choice. Their papers form the basis for this study, revealing choices of corpora and software, procedures employed, popular topics; in particular, learners were encouraged to describe obstacles encountered and solutions adopted. Difficulties of distance learning are not to be underestimated, but completion rates are higher than in some other options as informal feedback suggests students appreciate the autonomy to work on a topic of their own choosing in relation to their other studies or personal interests.

- **Cargill & Adams 2005.** Following a pilot session with applied linguistics students, 17 agricultural science students and researchers were introduced to ConcApp and a small corpus of 5-8 discipline-specific research articles for self-help with academic writing in a single 90-minute session. Questionnaires showed highly favourable responses, especially for creating their own corpus and intention to use in the future. However, a second questionnaire 8 months later elicited only 2 responses (for 10 valid email addresses), both of which were highly negative. Possible reasons are discussed, including timing (as these students were then in the experimental, not writing phase of their research), and lack of time to build their own corpus. Also, corpus-building had not been part of the training session, the original corpus was not available outside the university, and ConcApp is now no longer free. The main conclusion is that regular use is essential, and the researchers plan kibbitzer-like techniques in future one-to-one consulting sessions.

- **Götz & Mukherjee 2006.** Students conducted (1 of) 9 different corpus projects in a linguistics seminar: 1x literary analysis, 2x learner corpora, 3x spoken vs written English, 3x on-line environments and e-learning. Most perceived the experimental activities positively (if not enthusiastically) in terms of interest, usefulness and fun, although few thought they had learned much language. The less positive results are attributed to insufficient introduction to corpus consultation, although the students were allowed considerable autonomy.

- **Gut 2006.** 9 weeks of introduction and exercises were completed by project work on the LeaP corpus: 12 hours of NNS German and English sound-aligned with PRAAT, PoS, spectrogram, intonation, etc. Comparison with NS speech allows detailed analysis of NNS pronunciation. Students claimed they learned more from the traditional approach, which they also preferred to the corpus work, although they found this to be interesting. In the course as a whole they claimed they learned a lot about NNS accents which would be useful for their teaching, although only 1/10 claimed their own accent had improved.

- **Koo 2006a.** Data collected from video tracking and stimulated recall allow a detailed analysis of the use 10 Korean post-graduate students in the US made of corpora, dictionaries and other reference tools. Following a short introduction, participants were allowed a week to experiment with the corpus (New York Times 2000-02) before the main task which required them to paraphrase a newspaper article. Corpus uses include collocations, definitions, context, and parts of speech; dictionary uses include definitions, parts of speech, and sample sentences; the two were frequently combined, leading to greatest success. Participants particularly appreciated multiple examples of everyday language use, and frequency, patterns and collocations. They are claimed to have gained in responsibility for language learning, confidence in writing, independence and awareness of nature of authentic language.

- **Koosha & Jafarpour 2006.** 200 university students of English in Iran took part in this large-scale test for prepositional collocations over 15 classes, half as a control with traditional instruction, half in an experimental group using concordance print-outs. Cloze tests allowed comparisons before and after as well the experimental condition as between the 2 groups. The experimental group scored significantly higher in the use of the target language.

- **Lee & Swales 2006.** A detailed plan is provided of this 15-week course designed to help research students write academic papers. Work covered a number of language points decided by the teachers, though learners compiled corpora of their own writing and texts in their field, and had freedom of choice in the final project, presented as a conference paper. The high drop-out rate (12 students originally signed up) for this optional course is attributed to time constraints. Students were largely successful, and appreciated learning how to use tools to make them increasingly independent of native speakers, and language specific to their needs; this is reflected in the fact that most of them bought WordSmith Tools for their own future use.

- **Liou et al. 2006.** This paper describes the 3 main tools in the CANDLE project: TOTALrecall (a 40m-word Chinese-English parallel concordancer based on the cultural magazine Sinorama), Tango (a bilingual collocation aid based on the same corpus) and Collocation Checker (compares learner productions against the corpus). Various modules are then outlined, with brief reporting of empirical evidence from pilot tests. In particular, 16 graded texts were given to

38 first-year students over 12 weeks to read using these tools: the post-tests show significant improvement, the questionnaires are positive, and a sophisticated tracking system shows how the tools were used. A suite of tools to encourage inductive learning using various exercise types reports significant improvement for 32 first-year students in the post-test, but results from the delayed post-test and the questionnaires are not reported. In a third pilot, over 1/2 of errors indicated were successfully corrected using the tools with positive feedback. It is clear that most of these are only pilot studies of the tools rather than rigorous experimental analyses of corpus consultation, and basic information is often missing.

- **O'Sullivan & Chambers 2006.** As with Chambers and O'Sullivan (2004), learners first wrote an assigned 600-word text in the L2 with access to traditional resources, and errors were highlighted. They are then trained in corpus use over 9 hours, following which they are allowed 2 hours to use concordances to correct their original texts. Reactions were positive although slightly less so than for the previous MA students, with most finding corpora helpful (especially for lexico-grammar) and intending to use it again, although negative reactions are also noted (especially from students who had missed several training sessions).

- **Schmied 2006.** Learners access the Chemnitz Internet Grammar, which includes an English-German translation corpus, with various tools including links to the internet; learners can also input their own texts. The programme allows both deductive and inductive learning, with procedures outlined, exercises incorporating automatic feedback. The evaluation presented is minimal in this paper, the main finding being that the learners involved have extreme difficulty with an inductive approach. Other research by the same author (Schmied 2004) shows that there is no significant difference in performance between learners who follow set tasks and those with a more open-ended brief.

- **Boulton 2007.** Over 100 learners were tested on a single grammar point, use of *will* and *going to*, before and after the test condition. The experimental groups were given sets of raw concordances for each item, the control groups sheets of grammar explanations to study for 10 minutes; they were then used as an aid to completing the second test. Both groups scored significantly higher in the second test, although there was no significant difference between them. This is taken to indicate that learners can benefit from corpus consultation (with no instruction whatsoever) as much from traditional deductive teaching.

- **Braun 2007.** 16 classes in a German high school participated over 4 weeks, integrating corpus analysis to the ongoing course. 2 interviews were chosen from the ELISA corpus; the first was studied using traditional techniques, the second with corpus techniques for the experimental group, but not the control group. The corpus group rated these activities as more useful, and scored higher on the follow-up computer-based tasks; however, there was no overall difference on the final test on the unit as a whole.

- **Cresswell 2007.** The experimental group were introduced to corpora and required to investigate the meaning and usage of various connectors. About half attempted an inductive approach, starting with the data and comparing back to traditional references, the others starting with the references in a deductive approach. Learners of both types (especially inductive) generally succeeded in the task, with some qualification, although their findings proved difficult to convey to others. Furthermore, the overt knowledge was not found to translate well into use, as the experimental group performed only very slightly better than the control group on use of connectors in essays.

- **Curado Fuentes 2007.** The experimental group spent 3 hours exploring register with 6 tourist adverts in the corpus, before using a concordancer for 2 hours on the rest. Various exercises submitted by email show positive results. In the reading comprehension post-test, the experimental group performed significantly better than the control group, who had worked with text book and texts in a traditional manner.

- **Estling Vannestål & Lindquist 2007.** In the first trial (T1), the experimental group used corpus techniques, compared to the regular grammar book for the control. In each case, initial stages were conducted using concordance print-outs; then the queries were given; then students were asked to formulate their own queries for the questions set. Exercises were done in pairs between classes, followed by team-teaching of work done. Both trials used questionnaires to measure perceptions towards grammar, but only T1 measured performance from initial diagnostic test and final exam (i.e. not related to corpus work): learning outcomes were the same for both groups. Attitudes towards corpora were generally favourable in both trials, with many students saying they might use corpora in future (if not for teaching). Attitudes towards grammar were slightly worse in the experimental group after T1, though the T2 interviews highlighted technical problems, and the need for substantial training and support.

- **Hafner & Candlin 2007.** Logs of 2 cohorts of around 150 law students are supplemented by interviews with 9 participants, all advanced learners and regular computer users. Training resulted in initial benefits with some 'adopters' (4/9), but these disappeared over time as learners resorted to practices familiar in their law studies, tending to focus on full texts rather than corpus-like consultation (e.g. concordances). Indeed, some found the tools to

be constraining as they focus on language rather than content, while others also prefer more familiar tools (especially internet searches).

- **Hirata & Hirata 2007.** This paper involves students using keywords to browse the web for texts to compile their own corpus by simply copying into a text file, which they then explored at text level and using a concordancer, thus allowing them to 'authenticate' the texts. The 69 participants were involved in a university reading program designed to enhance their lexical skills and text awareness. Though they had only lower intermediate levels in English, and some had never used a search engine before even in their own language, they rate the course fairly positively in terms of satisfaction, understanding of the target items, and enhanced ability to induce meaning. This is taken to show that basic procedures can be beneficial with a minimum of training even for lower level learners with, it is claimed, longer-term benefits.

- **Huang & Liou 2007.** 16 news magazine texts were chosen and sequenced for lexical difficulty (ave 95% of 'known' words) and recycling; learners accessed these from home and responded to general comprehension questions. Target items were unknown words as tested before the experiment, subsequent learning via the VKS. Target words were highlighted red on first occurrence, then green, with on-line help including bilingual Chinese-English concordances and glosses. The use of concordances was not explicitly tested, but rather 'incidental' vocabulary learning, which was shown to be limited although the students were generally favourable to the programme overall; learning improved as the words were met more frequently, although some words encountered up to 15 times were still no learned. On-line tracking shows the concordances were used for already familiar words but not for target items; more training is recommended.

- **Lavid 2007.** Following an introduction to relevant areas of corpus linguistics, students were divided into 4 groups to investigate an area of mental transitivity; this involved selecting concordances, comparing L1 and L2 usage, and comparing their results with dictionaries. The students were enthusiastic about working with authentic language, but also expressed some frustration; they felt it would have been virtually impossible without considerable teacher guidance.

- **Sun 2007.** The Scholarly Writing Templates program was developed for postgraduate students who need to write articles in English, each student inputting articles from their specialist field. 20 Taiwanese students used the SWT to help with article structure as well as language use via corpus searches; while previous publishing experience was not seen to affect use, tracking shows that lower levels tended to be more accepting of proposals for both article structure and language use, while more advanced learners were more critical. Learning outcomes were not examined, although the paper reports more effective writing, and the feedback was generally positive, especially among those currently writing papers, and particularly for information structure.

- **Yeh et al. 2007.** Taiwanese students in their first year of English at university were introduced to a collocations concordancer based on the 40m-word Chinese-English bilingual corpus Sinorama. Over 4 weeks, 5 units (20 minutes each) focused on one overused adjective and 5 or 6 synonyms, the focus being on adjective-noun collocation. The initial induction exercises (quite hard in the traditionally deductive environment) were completed by a series of on-line exercises (substitution, gap-fill, translation). Immediate post-tests showed significant improvement, maintained 8 weeks later in a delayed post-test. Independently-assessed written production also improved, especially but not exclusively for the target items. Only 1/10 students held negative attitudes after the course, and 3/4 considered it effective.

- **Belz & Vyatkina 2008, 2005a, 2005b.** American learners of German and German learners of English participated in CMC activities, their emails and chats providing input to Telekorp, a 1.4m-word bilingual developmental learner corpus. An unspecified number of students participated over several years (the earlier papers), the main analysis focusing on 2 American learners. Initial CMC exchanges were analysed for use of German modal particles and *da*-compounds, with learners' own productions and Telekorp extracts forming the basis for some inductive activities (on paper or hands-on) alongside deductive presentations. Analysis of these learners' productions before and after the interventions show the target items being used more frequently and appropriately.

- **Boulton 2008a.** 131 learners were presented with concordance print-outs for common phrasal verbs to analyse with no prior training in order to test their ability to detect patterns in raw data. Pre- and post-tests showed significant improvement, although it was surprising that this was even greater for the phrasal than non-phrasal uses of the verbs; the suggestion is that participants concentrated more on what they perceived as difficult (the PVs), encouraged by the inductive approach. Although the higher level students performed better all round, the lower levels improved proportionally more; as all groups were at most lower intermediate, the conclusion seems to be that lower level learners can derive some benefit from raw concordances even with minimal training.

• **Boulton 2008b, 2010a.** 62 mainly lower-intermediate French learners of English were given a 5-minute introduction to concordancing, then in groups they worked on printed materials for 5 language items using corpora (inductively in pairs, feeding back to the whole group), and 5 others using dictionary entries and traditional teaching. The grammar/usage points had been collected from their own written productions earlier in the year, and featured in an earlier pre-test as well as a post-test the following week. The post-test shows significant improvement (unlike for 5 untreated items); although the improvement was greatest for the experimental treatments, the difference between treatments was not significant. The lower level students narrowed the gap using corpora while the more advanced ones maintained their advantage using the traditional approach. A questionnaire completed after the experimental condition showed very positive reactions.

• **Johns et al. 2008.** High school students in Taiwan were introduced to a novel in 16 weekly 3-hour extra sessions: the control group merely to read it, while the experimental group used various CALL programs (mainly gap-fill and bilingual sentence-shuffling) and (bilingual) concordancing tools designed by the authors, as well as concordance print-outs. The experimental group performed significantly better in the post-test for reading comprehension, and had double the reading speed; they also performed significantly better on the end-of-term exam, suggesting improvement extending beyond the specific tasks covered. Questionnaires showed very positive attitudes, with preferences for different tools, and greater motivation than the control group.

• **Lin 2008.** An 8-week course on AWL vocabulary was developed for 25 third-year English majors in Taiwan. In each lesson the words were first presented explicitly (by peer-teaching in the final lessons) and then in context; students then had to access to an on-line Moodle resource incorporating concordances to explore collocates. Very high scores in the VLT pre-test did not allow for breadth of knowledge to increase, but depth improved significantly on the VKS. A VocabProfile analysis of student essays showed a substantial increase in productive use of the target items, declining only slightly in the final delayed post-test 4 weeks later. Questionnaires showed fairly positive attitudes to the course overall.

• **Smith et al. 2008.** 25 students were recruited on-line for this experiment to test the utility of SketchEngine for learners of Chinese outside the classroom. Basic background information was collected and participants completed forced-choice 16-item pre-tests and post-tests on collocational and pragmatic context. Following a brief introduction in English, participants were asked to use the tools as much as possible over the coming 6 weeks. The context inevitably leads to certain difficulties, not least the fact that only 2 students completed the post-test; statistical information is thus not possible (one student increased score, the other had full marks in the pre-test). Feedback from the discussion boards suggest SketchEngine is popular, although there are problems in parsing Chinese, and one student wondered what it contributed above and beyond normal concordances; this feedback may only be from the 2 students who completed the post-test.

• **Thorne et al. 2008.** This pilot study of a training programme for international teaching assistants used transcripts from MICASE and the in-house ITAcorp, consisting of 300K words from 115 ITAs doing role plays of office hours and presentations. The contrastive approach was to raise awareness of language in relation to roles in spoken discourse. Inspired by Vygotskyan and Gal'perin, the authors distance themselves from 'traditional' DDL by making use of complex materialisations and detailed instructions to guide the learners in their analysis of corpus extracts – often longer stretches of discourse, though concordances were also used – to identify directives, discuss and identify patterns. None of the students had used the target structure before (you [HEDGE] want to…); none of the control group (n=13) used it afterwards either, but 10 of the two experimental groups (n=40) did in "varied and accurate ways" (p.277), and also used more other constructions than the control.

• **Yoon 2008.** A 10-week EAP course at an American university introduced the COBUILD Sampler to help advanced students with their writing. A variety of tools is used to assess the behaviour and reactions of 6 of the 14 students, mainly with Chinese or Korean L1. The participants used the tool for writing or checking their own productions out of class, with the teacher subsequently preparing handouts based on their emailed reports. Students reported increased confidence and autonomy in writing and improved writing procedures, as well as increased language awareness especially for checking existing knowledge. Corpus consultation was mainly perceived as a useful additional technique, more favourably received as time went on, although the paper reports mixed success depending on a variety of individual variables especially concerning past and on-going writing experiences.

• **Boulton 2009a.** The students were divided into 4 groups, 2 receiving traditional information types on the target language (grammar/usage books, bilingual dictionaries), 2 from corpus-based sources (KWICs, full-sentence data from newspapers via WebCorp). An immediate post-test using the information sheets was used to analyse how well they could use the different information types by comparison with the pre-test; recall was tested 10 days later. Both corpus groups performed significantly better in test 2, which is taken to mean that lower-level learners can identify patterns

from concordances and apply them to new contexts (especially noteworthy as the KWIC group scored higher than the full-sentence group). Scores in the third test were also higher than the pre-test showing that learning happened, although no difference was found between groups, suggesting that corpus data is as useful as traditional references.

- **Boulton 2009b.** This paper follows on from Boulton (2010b) with lower-intermediate consulting the BNC for specific language points in short sessions during class. Final questionnaires are rather disappointing: previous years had been more favourably disposed to paper-based materials. Reactions are also compared against learning styles measured using the ILS, showing the more Visual learners in particular were better disposed to corpus consultation, though correlations were not particularly high, perhaps due to the instrument used. Other areas for individual variation include motivation and attitude.

- **Chang & Sun 2009.** Corpora were introduced to 26 high school students in Taiwan to help with proof-reading of V+prep collocations in sets of 10 sentences. 15 minutes only were allowed to introduce concordancing; in subsequent weeks, the students (a) practiced for 30 minutes, (b) completed a task with or without scaffolding prompts for keyword selecting, output analysis, rule formation, evaluation; (c) completed a task without scaffolding prompts. The results show significant results all round: the corpus helped improved proof-reading on this point, scaffolding was useful and led to greater confidence in responses, and its benefits remained even when it was removed. Questionnaire feedback was largely positive, and items the students did or did not like about corpus work and scaffolding are discussed.

- **Chujo et al. 2009.** These 22 first-year Japanese engineering students are considered 'beginners' despite 6+ years of English (TOEIC average = 378). The experimental group spent their 90-minute class over 20 weeks working in pairs following detailed worksheets for hands-on work consultation of a Japanese-English parallel concordancer of newspaper texts (30 minutes), followed by class feedback, normative explanations, homework exercises and final feedback. Pre- and post-tests showed significant improvement on the TOEIC vocabulary and 5 of the 6 grammar areas covered ($p<0.01$); a control group not using corpora improved significantly on the vocabulary and two grammar areas only (however, this group was studying vocabulary and listening, not vocabulary and grammar). Questionnaire responses are generally positive about the course as a whole, though the DDL worksheets scored lowest (53% finding them useful) compared to the subsequent grammar explanations (74%), homework exercises (84%) and teacher feedback (84%).

- **Gilmore 2009.** 45 second-year intermediate-level Japanese students needing English for EAP writing received a 30-minute introduction to the BNC and BoE online, then spent an hour browsing it themselves. They then used the corpora out of class to improve a previous written assignment with errors highlighted. Four native-speaker reviewers then rated each version, and found the new version more 'natural' in 61% of cases, equivalent in 33%, less natural in only 6%. Feedback shows 95% of students found the corpora useful, with 84.5% preferring the BoE as more user-friendly, 15.5% preferring the BNC as larger.

- **Liu & Jiang 2009.** 160 intermediate-level EFL students at a Chinese university and 76 ESL students at 2 US universities (higher levels, including an MA TESOL course) followed a course introducing corpora in language learning during one semester. The instructors were all new to corpus linguistics and were trained by the researchers before teaching their courses. The classes are described, based on lesson plans and student productions; the main results are from a post-course questionnaire. The results are generally positive, as the learners appreciate the discovery approach and feel the courses helped their grammar and language awareness generally. The ESL results were more positive than the EFL ones for a variety of reasons discussed; there was also a significant correlation with the enthusiasm of the various teachers.

- **Schaeffer-Lacroix 2009.** In this three-year study, four large groups of German learners at a *collège* in France, aged 10-15, accessed specially-compiled tourist corpora as a writing resource, mainly via AntConc. In the main phase, 33 learners used the corpus as an aid to correction for writing projects; the output is analysed statistically and qualitatively (especially for revisions subsequent to corpus consultation), and interviews with learners provide further data. The corpus input is found in particular to aid collocation, conformity to genre-based norms, linguistic forms (e.g. compounding, morphology and punctuation), and contribute to learner autonomy, encouraging a "researcher" mentality to language learning; it was less successful for syntax.

- **Tyne 2009.** This sociolinguistics course was designed to encourage sensitivity to variation in L2 French. Following a 2-hour introduction to PRAAT, 3rd year students were required to record and transcribe 2 short extracts for further work. The course was offered over 3 years, with 65 students offering 'remarkably positive' feedback; 10 students in the final year answered a more detailed questionnaire, again with positive results concerning transcription, the use of corpora, and 9/10 claiming it had helped their French. These learners were clearly very aware of the issues involved,

but also claimed similar activities would have been useful at earlier stages, although some criticism leaves open the question of learning styles and cultures.

- **Varley 2009.** The course introduces English majors and minors to corpus study, beginning with closed activities leading up to individual projects where they explore genre through existing corpora or create their own. Initial questionnaires showed considerable enthusiasm, though this declined slightly in the reflective logs during the course, and in 5 subsequent interviews. Most receptive were the non-specialist students who had specific motivations, and who were often less successful with traditional teaching; age was not an important factor.

- **Boulton 2010b.** This study is a follow-up to Boulton (2008b, 2010a) where students used printed materials using either a traditional approach or DDL. Following the experiment, the students now completed the Index of Learning Styles in order to determine if there was any correlation between learning styles and a) their reactions to the DDL treatment, or b) their learning outcomes. Unsurprisingly, the strongest style for these architecture students is Visual; those who scored strongest on this scale were also the most receptive to DDL (r=0.44). There was also some correlation between the Active (r=0.31) and sequential (r=0.21) students and learning outcomes. However, the correlations are for the most part fairly low. This could be due to aspects of the experimental design (such as the small sample size: n=29), but might suggest that DDL is not limited to particular learning styles.

- **Chang 2010.** 28 of the 43 participants returned questionnaires following an introduction to corpus use, including general on-line corpora and home-made specialised corpora: introduction, demonstration and tasks. The questionnaires show that the participants found the corpora useful and would use them in the future (average >5 on a 6-point Likert scale), with a slight preference for general corpora despite their writing needs (for undergraduate, postgraduate and doctoral studies). Though the students had not used corpora before, over two thirds had used the Internet in similar ways, but found corpora more useful.

- **Conroy 2010.** This study introduced concordancing and advanced Google techniques in 1 to 4 hours to 165 students at an Australian university from a range of disciplines and courses. The 53 questionnaires show that all found the training 'very' (68%) or 'somewhat useful' (32%), and 87% wanted further training. In particular, 89% found using concordancers 'very rewarding' or 'somewhat useful'. 47% of students already used the internet for language learning, increasing to 53% two weeks after training, with 58% saying they would continue later; the equivalents for corpus use were 4%, 28% and 36%. Some students said they preferred the internet to the 1m-word Brown corpus; 45 errors were highlighted in the work of 3 students, who chose to use Google to correct 22 (15 successful), a corpus for one (successful), and other tools for the remaining 22 (11 successful). The study concludes that students can usefully be introduced to corpora, but that Google-assisted language learning may be easier for some to assimilate and thus more profitably promoted.

- **Kennedy & Miceli 2010.** Intermediate specialist learners of Italian at undergraduate level in Australia used the CWIC corpus as a "resource" for creative writing in this one-semester apprenticeship. The case-study of 3 students focused on use of the corpus and bilingual dictionaries to revise an earlier text for 45 minutes via "pattern-hunting" (looking for content or language ideas), "pattern-defining" (checking usage) and translation (finding an Italian equivalent of an idea in English). The results, along with a 45-minute interview and end-of-course questionnaires, show very different uptake among the three students; successful corpus use is linked to "trial and error" and "making it their own"; a number of principles are inferred.

- **Landure & Boulton 2010.** In most studies using corpora for revision, the learners' attention is drawn to the errors by the teacher. In this study, the students first translated 10 sentences using different tools: their own knowledge, dictionaries, and on-line translators; they were then encouraged to use the corpora when they themselves detected a difference in their other results. The revisions were generally positive, suggesting that low-level learners can draw some immediate benefit from careful corpus use even with no prior instruction. The reactions were overwhelmingly positive for these non-English majors, and resulted in considerable collaboration, unusual for them.

- **Moreno Jaén 2010.** Pre- and post-tests were compared for 21 students (3rd- to 5th-year English majors in 7 different Andalusian universities) using an on-line collocations module (ADELEX). Concordancers were used far more by the end of the course, though still less than dictionaries (though in both pre- and post-test, over half of students used no resources at all). The course was found to be effective with a significant improvement in overall results and for 5 or the 7 individual tasks set for identifying, using and producing collocations. Questionnaires also showed favourable reactions about collocation learning and concordancing techniques. However, it is argued that 3 weeks is not enough for all students to become autonomous corpus users.

- **Okamoto 2010.** Third- and fourth-year computer science students (n=21) opted for this 15-week course to help with professional writing in Japan; though many of their courses are in English, they still are of low proficiency (TOEIC

maximum 350). They began with the BNC and MICASE before moving on to AntConc to analyse at least three research articles they chose themselves, presenting their (mainly lexicogrammar) results to the class. 15 questionnaires were collected at the end of the course, suggesting that the students were "on the whole… positive" despite their L2 level, intending to continue with corpora in the future. Difficulties were mainly technical, especially in the use of Macintosh computers.

- **Park & Kinginger 2010.** This case study, of a single advanced learner in the US using a corpus for writing, tracks all corpus queries and records the computer screen; the participant later reviews these and comments on her thoughts. The results show that the writing process is divided into 'transactions' of specific questions, which can be described as interrelated narratives. 194 searches are conducted in 109 minutes for 118 transactions (26 of which are complex, featuring several related points); analysis shows that planning, writing and revising occur simultaneously and not in discrete steps.

- **Rapti 2010.** Following a pilot study, 13- to 15-year-old learners followed either a traditional or DDL approach to grammar (verbs) over a 5-month period. Although no statistical analysis is provided, the experimental group's test performance was more stable, with more students improving their score than in the control group. Questionnaires and interviews reveal mixed reactions to the use of concordances (ease, usefulness, motivation), but on the whole DDL was preferred to a traditional approach to grammar. Most classes were based on prepared hand-outs, the 2 hands-on classes encountering a number of difficulties.

- **Wu et al. 2010.** Following a 2-hour introduction to the tools, 9 students wrote an IELTS-style essay in standard conditions; this was annotated for "errors" which the students then attempted to revise using the software in 2 hours. Their searches were logged and the results compared. In total, 67% of errors were successfully corrected (12% were unchanged, the others unsuccessfully changed). The improvement is attributed to the frequency-based nature of the lists, as well as the reduction of confusing data.

- **Acar et al. 2011.** The students were given 20 minutes of instruction showing them how to use inverted commas in Google and instructed to check 4-word blocks of their writing on general topics over a 1-week period: less than 100 hits was given as a sign they should try changing prepositions, articles, etc. (>1000 = OK). Though most of the sentences still contained errors afterwards, 24% of them showed improvement in clarity or grammatical accuracy (min 16%, max 31%), a promising result from this entirely autonomous procedure with no teacher highlighting of errors.

- **Al-Lawati 2011.** A 'corrected' corpus of advanced student essays was used to devise paper-based grammar exercises for lower-intermediate students in Oman. Following an introductory session, students were recording working in groups of 4 over several weeks devising and testing hypotheses against the corpus data. A classification system based on Oxford (1990) and O'Malley & Chamot (1990) showed the main strategies used were association / elaboration, deductive reasoning, using linguistic clues, paying selective attention, and monitoring. This is taken to suggest that these are inherently "data-driven learning strategies and implies that students should be given training in how to use these strategies" (p314).

- **Boulton 2011a.** Lower-intermediate students of English in the experimental groups spent the last 10-20 minutes of class time over 12 sessions using the BYU-BNC for language points arising in class; no training was provided, but detailed instructions were given for each activity. At the end of the year, both experimental and control groups were given a short text to read for 5 minutes, then tested on whether they had noticed a number of language points (focus on form and on meaning) entirely unrelated to any work conducted during the year. The results show the experimental group performing better in noticing than the control group, though the difference did not quite reach statistical significance.

- **Chang & Kuo 2011.** Corpus and genre-analysis were combined in an on-line course based on 60 research articles in computer science. The 23 intermediate-level students in the pilot study were positive in most questionnaire feedback at the end of the semester, though the platform itself was viewed negatively. Two pre- and post-course assignments were analysed for students' ability to detect moves and organise information in introductions, and to apply data-commentary patterns to their own work. 65% of the first assignment were inappropriately organised before the course, but all were well structured with effective use of connecting expressions in the second. The second task also showed improvement, though this was not statistically significant, a fact attributed to different task conditions.

- **Charles 2011.** This course aimed to combine corpus consultation and discourse analysis via consultation of an ESP corpus of published native-speaker PhDs (300K materials science, 190K politics). 49 graduates with different L1s and different disciplines spent an hour a week over 6 weeks investigating rhetorical functions: comparing 2 extracts then exploring given items further in expanded contexts, leading to pair discussion and class feedback. The feedback is generally positive (averaging 82% agree or strongly agree with the questions); the paper discusses in detail the less

favourable responses, including difficulty of software use. It is argued that corpus and discourse can be combined in a way that is "both feasible and pedagogically valid" (p. 40).

- **Chen 2011**. This study presents WebCollocate which extracts collocates from 160m words of POS-tagged Project Gutenberg texts; a more modern Wikipedia version is under construction. Two groups of intermediate-level college students in Taiwan translated 30 sentences with problem collocates, then were introduced to the tool (29 students) or the VLC concordancer (n=22) in about 10 minutes, then allowed 30 minutes to improve their translations. There was no significant difference before, but the manually graded products were significantly better in the experimental group post-treatment. Coupled with interviews with some students, it seems that learners can use collocates lists extracted from a PoS-tagged corpus more easily than simply searching for collocates in raw concordances.

- **Gao 2011**. 21 first-year upper-intermediate L2 engineering students in Taiwan were briefly introduced to a Chinese-English parallel concordancer. Pre-/post-test translation tasks showed significant improvement as the participants managed to detect errors in their own earlier translations, especially for word choice, combinations and collocations. Subsequent questionnaires revealed overwhelmingly positive reactions to the parallel corpus as a complement to dictionaries, with students intending to continue using it afterwards.

- **Kaszubski 2011**. Pilot tests are outlined using IFAConc for EAP. In the first, 12 corpus linguistics students were introduced to the software in 30 minutes and required to complete two homework assignments: questionnaire responses are generally favourable, but a split emerges in preference for pursuing set questions or their own queries. In the second, 14 first-year writing students at lower levels went from training examples to autonomous consultation: tracking histories suggests considerable individual variation, with 6 'adopters' who continued after the study, 5 'minimalist' users and 3 'refusers' – categories which correlate with ability and final course grades. The importance of feedback via the annotation function is also highlighted in the questionnaires.

- **Kettemann 2011**. The teacher used WST with a 140K corpus of self-proclaimed 'Emo' texts (articles, blogs, forum threads, poems, lyrics) to create 4 tasks based on data on a decreasing cline of processing: (a) verbs following *I*, semantically grouped with frequencies given; (b) keywords compared to COLT; (c) most frequent content words; (d) unsorted concordances of *alone* / *lonely* / *on my own*. Ten data sets were collected from 40 first-year students who responded to questions about the data in their own time during the week, a qualitative analysis of which shows considerable self-investment and motivation. The conclusion is "optimistic" about the use of (language) corpora for other domains, here cultural studies.

- **Kettemann & Marko 2011**. Adopting a Critical Discourse Analysis to corpus use on Christian Fundamentalism, the authors compiled a 600K-word corpus on Intelligent Design with 17 students. After a brief introduction to CDA and WordSmith, the participants spent at least an hour a week over 6 weeks exploring this corpus. Research diaries show that after about 3 weeks they had become "relatively competent WordSmith users" (p.35), but tended then not to stretch to new techniques, though they mostly preferred working with their own data rather than the corpus provided. Different students showed different ways of working with corpora, but many went for the instant gratification of word lists rather than in-depth concordance analysis. Questionnaires are largely favourable; the final project reports generally encouraging.

- **Landure 2011**. This study introduces 44 unemployed trainees of varying ages to CoCA and the BNC to help with specialist computer texts. Language level is mixed (mainly lower-intermediate), as is level of education (¼ not having completed high school). Small group work and individual consultation change the dynamic of the class, increasing autonomy and collaboration as well as confidence as the students are required to 'discover' how the corpora work and what they can be used for. Two articles were worked on, the first using corpora, the second bilingual dictionaries; though the latter were found to be more efficient for immediate consultation, retention was better for corpora (68% vs 48%) a week later when gap-filling the same articles with the items previously searched for. Questionnaires show that only half found the corpus work easy, but 84% rated it useful and 80% claimed they would use corpora in their future careers.

- **Pérez-Paredes et al. 2011**. Second-year English students in Spain were given on-line grammar activities in two 1-hour sessions on a previously seen text, each involving an observation phase followed by corpus searches (BNC) and rewrite exercises. Unlike the control group (n=15), the experimental group (n=22) had explicit guidance for corpus use. Computer tracking showed that although the EG had less time (all but one accessed the guidance), they used the Internet more, used more different web sites, and completed more activities, though the only significant difference was in the corpus use. The authors argue that tracking is essential to see what happens at a finer level of granularity at the level of individual, and underline the importance of training.

- **Philip 2011**. Advanced students wrote a short story using key words in the same form and order as given. The main research question is based on post-hoc, apparently subjective classification of learners into small groups according to their preferred resource: 6 corpus users, 3 Googlers, 3 dictionary users and 20 mixed-resource users (compared to a control of 10 dictionary users) – all familiar resources for these Italian advanced students of English. Analysis of the final texts show that corpus use is definitely effective, but at least in part because they tend to be used by 'good' students – those who are motivated to discover and experiment with a new tool to add to their arsenal. A number of characteristics of other users are discussed.

- **Rodgers et al. 2011.** 200 articles (140K words) from 'popular' sources constitute a biotechnology corpus for 23 second- or third-year students with an option in French in Ireland. Following two introductory sessions, students spent 10 weeks following worksheets for grammar and topic-oriented tasks using WST. 19 questionnaires and interviews showed largely favourable reactions to learning both language and content, and all but 2 students expressly claimed they would like to continue the following year. However, learning styles and preferences are salient: some found it "more personal", "easy" and "quick to use", and cite the advantages of contextualisation, while others listed these same points as disadvantages (p.403-404).

- **Smith 2011**. 90 first-year undergraduates in Taiwan were exposed to corpora over a semester before choosing a project: 33 chose general websites, 49 chose some aspect of SketchEngine, 19 chose to build a corpus using WebBootCat, the focus of this study, mainly because it would be relevant to their field of study. 3 of the 19 had some difficulty in understanding what corpus linguistics is all about, a number followed the instructions on how to tweak their corpus without really understanding. 36 comments were collected, only 3 of which were negative (1 claiming it was not enjoyable / useful, 2 indicating they would not continue with corpora after course), the others generally finding the construction of their corpus enjoyable and useful and led to a sense of ownership, that they learned English in the process, might continue with corpora after the course, and felt they had acquired reusable skills. It is argued that the process of corpus building is as important as the explicit outcomes (if not more so).

- **Boulton 2012a.** Distance students majoring in English were encouraged to use monolingual on-line corpora (CoCA and the BNC) in non-literary translation from French during on-line exams. Minimal training was provided, with students having to explore the corpora in advance at home based on a limited theoretical introduction, a series of guides, and examples of past practice. The exam required them to choose three short extracts (3-5 words) and demonstrate how they used the corpora in translation. The paper discusses the wide range of search techniques and functions used, and how effectively they are able to think through the processes in the exams themselves. While these actual results are encouraging, a questionnaire shows the students lack confidence in their own abilities despite recognising the potential of corpus use in translation; further scaffolding and guidance is clearly necessary.

- **Boulton 2012b.** DDL was briefly introduced to 40 second-year architecture students, who then experienced a variety of DDL activities on problem lexicogrammar points in short sessions over 10 weeks, alternating between paper-based and computer-based activities. A test in the final session gave a slight but not significant advantage to paper-based activities, though questionnaires showed the students had a slight preference for computer-based activities. They were generally receptive to DDL as a whole, but this did not seem to correlate with learning outcomes. A correlation was found between proficiency and outcomes from the paper-based treatment, but this can be interpreted as meaning computer-based DDL is open to all levels even among these lower-intermediate learners.

- **Chang 2012.** 15 introductions to research papers in the social sciences were annotated for move structure and stance to help 7 Chinese doctoral students in the US in their writing. Following 2 orientation sessions, for 3 weeks the participants wrote and revised article introductions using the corpus. A post-test shows more overt move structures and greater awareness, as well as improved stance overall (though less so in more nuanced cases); this correlates significantly with performance. The most frequently used processes were 'make sense', 'explore' and 'reason/analyse'; less frequent were 'make inference' and 'verify'; least used was 'predict/hypothesise' (though there is no significant correlation with performance). Surprisingly, the students preferred to work a discourse rather than sentence level.

- **Charles 2012.** Five groups of 18 students from mixed disciplines were introduced to corpus work over 6 weeks, compiling and, in some cases, cleaning their own selections of 10 or more research articles to help with their own EAP writing. 50 questionnaires were collected, showing that 80% or more found the corpus compilation easy, AntConc easy to use to find answers to their own questions, and the whole process interesting. This is supported by the claim that 58% use their corpus at least once a week, with 86% expecting to add more files in future. Difficulties mainly concern the use of 'quick and dirty' corpora, though this may be 'adequate' for most purposes.

• **Chujo & Oghigian 2012.** Beginning level students of English used mainly computer-based DDL from a parallel Japanese-English newspaper corpus to study vocabulary, noun phrases and verb phrases over two semesters. Pre- and post-tests show them making significant gains in almost all areas, while the non-DDL group only gained on the vocabulary tests and one of the 4 VP question types, none of the NP ones. However, the non-DDL group had been concentrating on vocabulary and listening, and improved their TOEIC Bridge scores over the year more than the DDL groups. The following year, a comparable group (25 students) did most activities on paper: open-ended questions suggest differing preferences for a variety of reasons; translation was found to be less necessary when working with edited paper-based materials. A mix of paper-based and computer-based DDL is therefore mooted.

• **Frankenberg-Garcia 2012.** This experimental study was designed to test whether dictionary definitions would help with lexical meaning (reception: multiple choice translation) and usage (production: error-correction) more or less than specially-chosen single or multiple (3-item) corpus 'examples'. The 48 Portuguese students of English divided into groups gave clear results: in comprehension, definitions and multiple examples are equally useful, single items and no input (control group) significantly less so; in production, multiple examples are significantly better than single examples, which are significantly better than definitions or no input. Interestingly, substantial variation was found between individuals in the corpus groups, unlike the dictionary and control groups.

• **Geist & Hahn 2012.** 53 learners aged 14-16 were introduced to BNC Baby (4m words) and Xaira and then required to write a cover letter for a job advertisement. Of the 6 functions demonstrated, the most used was 'simple search'; 'sort' not at all. Technical aspects were not a problem: the class with previous experience of corpus use was quickest to master the technical aspects, but actually demonstrated less effective corpus use overall. Age-related issues include irrelevant corpus use just to satisfy the teacher (on-line bilingual dictionaries were preferred), and distractions from games and the Internet, as well as motivation – highly variable between students.

• **Osolsobě & Vališová 2012.** Groups of novice corpus users of mixed nationalities were introduced to large corpora of written and spoken Czech and encouraged to use them inductively to work out lemmas, gender/case, frequencies, and collocates. The first 3 were deemed easy and beneficial, while the last was more complex due to software use and presentation of results (especially for a highly-inflected language like Czech), but feedback was generally positive. Similar activities were conducted with 10 native Czech speakers, who also appreciated the approach (suggesting combining it with dictionary work), but were frustrated with annotation errors.

• **Geluso Forthcoming.** As many students are likely already using Google for linguistic purposes, Geluso exploits this as a way in to DDL with just a brief introduction and 3 short sessions. 25 Japanese students searched for formulaic sequences using double quote marks to improve self-selected 'questionable phrases' in their own writing, revising according to frequency. 4 raters compared the 167 phrases before and after, the results showing significant increase in 'naturalness' (though there was no control group revising in other ways).

3. References to source articles

Acar, Adam, Joe Geluso & Tadaki Shiki. 2011. How can search engines improve your writing? *CALL-EJ*, 12(1): 1-10. http://callej.org/journal/12-1/Acar_2011.pdf, accessed 28/03/11.

Allan, Rachel. 2006. Data-driven learning and vocabulary: Investigating the use of concordances with advanced learners of English. *Centre for Language and Communication Studies, Occasional Paper*, 66. Dublin: Trinity College Dublin.

Al-Lawati, Nasrin. 2011. Learning strategies used and observations made by EFL Arab students while working on concordance-based grammar activities. *Arab World English Journal*, 2(4): 302-322.

Aston, Guy. 1997. Involving learners in developing learning methods: Exploiting text corpora in self-access. In P. Benson & P. Voller (eds), *Autonomy and Independence in Language Learning*. London: Longman, p. 204-214.

Baten, Lut, Anne-Marie Cornu & L.V. Engels. 1989. The use of concordances in vocabulary acquisition. In C. Laurent & M. Nordman (eds), *Special Language: From Humans Thinking to Thinking Machines*. Clevedon: Multilingual Matters, p. 452-467.

Belz, Julie A. & Nina Vyatkina. 2005a. Learner corpus analysis and the development of L2 pragmatic competence in networked intercultural language study: The case of German modal particles. *Canadian Modern Language Review*, 62(1): 17-48.

Belz, Julie A. & Nina Vyatkina. 2005b. Computer-mediated learner corpus research and the data-driven teaching of L2 pragmatic competence: The case of German modal particles. *CALPER Working Papers*, 4, April.

Belz, Julie A. & Nina Vyatkina. 2008. The pedagogical mediation of a developmental learner corpus for classroom-based language instruction. *Language Learning & Technology*, 12(3): 35-52. http://llt.msu.edu/vol12num3/belzvyatkina.pdf, accessed 02/04/11.

Bernardini, Sylvia. 2000. Systematising serendipity: Proposals for concordancing large corpora with language learners. In L. Burnard & T. McEnery (eds), *Rethinking Language Pedagogy from a Corpus Perspective*. Frankfurt: Peter Lang, p. 225-234.

Bernardini, Sylvia. 2002. Exploring new directions for discovery learning. In B. Kettemann & G. Marko (eds), *Teaching and Learning by Doing Corpus Analysis*. Amsterdam: Rodopi, p. 165-182.

Boulton, Alex. 2007. DDL is in the details… and in the big themes. In M. Davies, P. Rayson, S. Hunston & P. Danielsson (eds), *Proceedings of the Corpus Linguistics Conference: CL2007.* http://ucrel.lancs.ac.uk/publications/CL2007/, accessed 14/05/08.

Boulton, Alex. 2008a. Looking for empirical evidence for DDL at lower levels. In B. Lewandowska-Tomaszczyk (ed.), *Corpus Linguistics, Computer Tools, and Applications: State of the Art*. Frankfurt: Peter Lang, p. 581-598.

Boulton, Alex. 2008b. DDL: Reaching the parts other teaching can't reach? In Ana Frankenberg-Garcia (ed.), *Proceedings of the 8th Teaching and Language Corpora Conference*. Lisbon: Associação de Estudos e de Investigação Cientifíca do ISLA-Lisboa, p. 38-44.

Boulton, Alex. 2009a. Testing the limits of data-driven learning: Language proficiency and training. *ReCALL*, 21(11): 37-51.

Boulton, Alex. 2009b. Corpora for all? Learning styles and data-driven learning. In M. Mahlberg, V. González-Díaz & C. Smith (eds), *Proceedings of 5th Corpus Linguistics Conference*. http://ucrel.lancs.ac.uk/publications/cl2009/, accessed 11/10/10.

Boulton, Alex. 2010a. Data-driven learning: Taking the computer out of the equation. *Language Learning*, 60(3): 534-572.

Boulton, Alex. 2010b. Consultation de corpus et styles d'apprentissage. *Cahiers de l'APLIUT*, 29(1): 98-115.

Boulton, Alex. 2011a. Language awareness and medium-term benefits of corpus consultation. In A. Gimeno Sanz (ed.), *New Trends in Computer-Assisted Language Learning: Working Together*. Madrid: Macmillan ELT, p. 39-46.

Boulton, Alex. 2011b. Bringing corpora to the masses: Free and easy tools for language learning. In N. Kübler (ed.), *Corpora, Language, Teaching, and Resources: From Theory to Practice*. Bern: Peter Lang, p. 69-96.

Boulton, Alex. 2012a. Beyond concordancing: Multiple affordances of corpora in university language degrees. In D. Macaire & A. Boulton (eds), *Languages, Cultures and Virtual Communities. Procedia: Social and Behavioral Sciences*, 34: 33-38. http://www.sciencedirect.com/science/journal/18770428/34, accessed 09/05/12.

Boulton, Alex. 2012b. Hands-on / hands-off: Alternative approaches to data-driven learning. In J. Thomas & A. Boulton (eds), *Input, Process and Product: Developments in Teaching and Language Corpora*. Brno: Masaryk University Press, p. 153-169.

Boulton, Alex & Stephan Wilhelm. 2006. Habeant corpus – they should have the body: Tools learners have the right to use. *ASp*, 49-50: 155-170.

Bowker, Lynne. 1998. Using specialised monolingual native-language corpora as a translation resource: A pilot study. *Meta*, 43(4): 631-651. http://www.erudit.org/revue/meta/1998/v43/n4/002134ar.pdf, accessed 15/03/06.

Bowker, Lynne. 1999. Exploring the potential of corpora for raising language awareness in student translators. *Language Awareness*, 8(3-4): 160-173.

Braun, Sabine. 2007. Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora. *ReCALL*, 19(3): 307-328.

Cargill, Margaret & Ray Adams. 2005. Learning discipline-specific research English for a world stage: A self-access concordancing tool? *Higher Education in a Changing World: Proceedings HERDSA*, p. 56-92. Sydney 2005. http://conference.herdsa.org.au/2005/pdf/refereed/paper_202.pdf, accessed 08/03/11.

Chambers, Angela. 2005. Integrating corpus consultation in language studies. *Language Learning & Technology*, 9(2): 111-125. http://llt.msu.edu/vol9num2/pdf/chambers.pdf, accessed 21/10/08.

Chambers, Angela & Íde O'Sullivan. 2004. Corpus consultation and advanced learners' writing skills in French. *ReCALL*, 16(1): 158-172.

Chan, Tun-pei & Hsien-Chin Liou. 2005. Effects of web-based concordancing instruction on EFL students' learning of verb-noun collocations. *Computer Assisted Language Learning*, 18(3): 231-251.

Chang, Ching-Fen & Chih-Hua Kuo. 2011. A corpus-based approach to online materials development for writing research articles. *English for Specific Purposes*, 30: 222-234.

Chang, Ji-Yeon. 2010. Postsecondary EFL students' evaluations of corpora with regard to English writing. *SNU Journal of Education Research*, 19: 57-85. http://s-space.snu.ac.kr/bitstream/10371/72997/1/vol19_3.pdf, accessed 11/04/11.

Chang, Peichin. 2012. Using a stance corpus to learn about effective authorial stance-taking: A textlinguistic approach. *ReCALL*, 24(2): 209-236.

Chang, Wen-Li & Yu-Chih. Sun. 2009. Scaffolding and web concordancers as support for language learning. *Computer Assisted Language Learning*, 22(4): 283-302.

Charles, Maggie. 2011. Using hands-on concordancing to teach rhetorical functions: Evaluation and implications for EAP. In A. Frankenberg-Garcia, L. Flowerdew & G. Aston (eds), *New Trends in Corpora and Language Learning*. London: Continuum, p. 26-43.

Chen, Hao-Jan Howard. 2011. Developing and evaluating a web-based collocation retrieval tool for EFL students and teachers. *Computer Assisted Language Learning*, 24(1): 59-76.

Cheng, Winnie, Martin Warren & Xu Xun-feng. 2003. The language learner as language researcher: Putting corpus linguistics on the timetable. *System*, 31(2): 173-186.

Chujo, Kiyomi, Laurence Anthony & Kathryn Oghigian. 2009. DDL for the EFL classroom: Effective uses of a Japanese-English parallel corpus and the development of a learner-friendly, online parallel concordancer. In M. Mahlberg, V. González-Díaz & C. Smith (eds), *Proceedings of 5th Corpus Linguistics Conference*. http://ucrel.lancs.ac.uk/publications/cl2009/, accessed 01/06/10.

Chujo, Kiyomi & Kathryn Oghigian. 2012. DDL for EFL beginners: A report on student gains and views on paper-based concordancing and the role of L1. In J. Thomas & A. Boulton (eds), *Input, Process and Product: Developments in Teaching and Language Corpora*. Brno: Masaryk University Press, p. 170-183.

Ciesielska-Ciupek, Maria. 2001. Teaching with the internet and corpus materials: Preparation of the ELT materials using the internet and corpus resources. In B. Lewandowska-Tomaszczyk (ed.), *PALC 2001: Practical Applications in Language Corpora.* Frankfurt: Peter Lang, p. 521-531.

Clerehan, Rosemary, Giselle Kett & Renee Gedge. 2003. Web-based tools and instruction for developing it students' written communication skills. *Proceedings of Exploring Educational Technologies,* 16-17 July, Monash University. http://www.monash.edu.au/groups/flt/eet/full_papers/clerehan.pdf, accessed 28/03/11.

Cobb, Tom. 1997a. *From Concord to Lexicon: Development and Test of a Corpus-Based Lexical Tutor*. Montreal: Concordia University. http://www.er.uqam.ca/nobel/r21270/webthesis/Thesis0.html, accessed 15/03/06.

Cobb, Tom. 1997b. Is there any measurable learning from hands-on concordancing? *System*, 25, 3: 301-315.

Cobb, Tom. 1999. Breadth and depth of lexical acquisition with hands-on concordancing. *Computer Assisted Language Learning*, 12(4): 345-360.

Cobb, Tom. 2006. Internet and literacy in the developing world: Delivering the teacher with the text. *Educational Technology Research and Development*, 54(6): 627-645.

Cobb, Tom, Chris Greaves & Marlise Horst. 2001. Can the rate of lexical acquisition from reading be increased? An experiment in reading French with a suite of on-line resources. Translation of 'Peut-on augmenter le rythme d'acquisition lexicale par la lecture? Une expérience de lecture en français appuyée sur une série de ressources en ligne. In P. Raymond & C. Cornaire (eds), *Regards sur la Didactique des Langues Secondes*. Montreal: Editions Logique, p. 133-153. http://www.er.uqam.ca/nobel/r21270/cv/BouleE.htm, accessed 23/07/06.

Conroy, Mark A. 2010. Internet tools for language learning: University students taking control of their writing. *Australasian Journal of Educational Technology*, 26(6): 861-882. http://ascilite.org.au/ajet/ajet26/conroy.html, accessed 08/03/11.

Cresswell, Andy. 2007. Getting to 'know' connectors? Evaluating data-driven learning in a writing skills course. In E. Hidalgo, L. Quereda & J. Santana (eds), *Corpora in the Foreign Language Classroom*. Amsterdam: Rodopi, p. 267-287.

Curado Fuentes, Alejandro. 2002. Exploitation and assessment of a business English corpus through language learning tasks. *ICAME Journal*, 26: 5-32. http://gandalf.aksis.uib.no/icame/ij26/curadofuen.pdf, accessed 26/02/06.

Curado Fuentes, Alejandro. 2003. The use of corpora and IT in a comparative evaluation approach for oral business English. *ReCALL*, 15(2): 189-201.

Curado Fuentes, Alejandro. 2004. The use of corpora and IT in evaluating oral task competence for tourism English. *CALICO Journal*, 22(1): 5-22.

Curado Fuentes, Alejandro. 2007. A corpus-based assessment of reading comprehension in English. In E. Hidalgo, L. Quereda & J. Santana (eds), *Corpora in the Foreign Language Classroom*. Amsterdam: Rodopi, p. 309-326.

Eriksson, Andreas. 2012. Pedagogical perspectives on bundles: Teaching bundles to doctoral students of biochemistry. In J. Thomas & A. Boulton (eds), *Input, Process and Product: Developments in Teaching and Language Corpora*. Brno: Masaryk University Press, p. 196-212.

Estling Vannestål, Maria & Hans Lindquist. 2007. Learning English grammar with a corpus: Experimenting with concordancing in a university grammar course. *ReCALL*, 19(3): 329-350.

Fan, May & Xu Xunfeng. 2002. An evaluation of an online bilingual corpus for the self-learning of legal English. *System*, 30(1): 47-63.

Frankenberg-Garcia, Ana. 2005. A peek into what today's language learners as researchers actually do. *International Journal of Lexicography*, 18(3): 335-355. http://ijl.oxfordjournals.org/cgi/rapidpdf/eci015v1?ijkey=03dnuAUsgGOEE&keytype=ref, accessed 21/11/08.

Frankenberg-Garcia, Ana. 2012. Learners' use of corpus examples. *International Journal of Lexicography*: advance access.

Gan, Siowck-Lee, Freddie Low & Noran Fauziah bte Yaakub. 1996. Modeling teaching with a computer-based concordancer in a TESL preservice teacher education program. *Journal of Computing in Teacher Education*, 12(4): 28-32.

Gao, Zhao-Ming. 2011. Exploring the effects and use of a Chinese-English parallel concordancer. *Computer Assisted Language Learning*, 24(3): 255-275.

Gaskell, Delian & Tom Cobb. 2004. Can learners use concordance feedback for writing errors? *System*, 32(3): 301-319.

Geist, Monika & Angela Hahn. 2012. Using a corpus for written production: A classroom study. In J. Thomas & A. Boulton (eds), *Input, Process and Product: Developments in Teaching and Language Corpora*. Brno: Masaryk University Press, p. 124-136.

Geluso, Joe. Forthcoming. Phraseology and frequency of occurrence on the web: Native speakers' perceptions of Google-informed second language writing. *Computer Assisted Language Learning* [iFirst article].

Gilmore, Alex. 2009. Using online corpora to develop students' writing skills. *ELT Journal*, 63(4): 363-372.

Götz, Sandra & Joybrato Mukherjee. 2006. Evaluation of data-driven learning in university teaching: A project report. In S. Braun, K. Kohn & J. Mukherjee (eds), *Corpus Technology and Language Pedagogy: New resources, New Tools, New Methods.* Frankfurt: Peter Lang, p. 49-67.

Granath, Solveig. 1998. Using corpora in teaching English syntax to EFL students at the university level. In L. Burnard (ed.), *Proceedings of Teaching and Language Corpora (TaLC98)*. Oxford: Keble College, p. 87-92.

Granath, Solveig. 2009. Who benefits from learning how to use corpora? In K. Aijmer (ed.), *Corpora and Language Teaching*. Amsterdam: John Benjamins, p. 47-65.

Gut, Ulrike. 2006. Learner speech corpora in language teaching. In S. Sabine Braun, K. Kohn & J. Mukherjee (eds), *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*. Frankfurt: Peter Lang, p. 69-86.

Hadley, Gregory. 2002. Sensing the winds of change: An introduction to data-driven learning. *RELC Journal*, 33(2): 99-124.

Hafner, Christoph A. & Christopher N. Candlin. 2007. Corpus tools as an affordance to learning in professional legal education. *Journal of English for Academic Purposes*, 6(4): 303-318.

Hirata, Yoko & Yoshihiro Hirata. 2007. Independent research project with web-derived corpora for language learning. *JALT CALL Journal*, 3(3): 33-48.

Horst, Marlise & Tom Cobb. 2001. Growing academic vocabulary with a collaborative on-line data-base. In B. Morrison, D. Gardner, K. Keobke & M. Spratt (eds), *ELT Perspectives on IT and Multimedia: Selected Papers from the ITMELT Conference 2001*. Hong Kong: Hong Kong Polytechnic University, p. 189-225. http://elc.polyu.edu.hk/conference/papers2001/cobb.htm, accessed 01/10/07.

Horst, Marlise, Tom Cobb & Ioana Nicolae. 2001. Expanding academic vocabulary with an interactive on-line database. *Language Learning & Technology*, 9(2): 90-110. http://llt.msu.edu/vol9num2/pdf/horst.pdf, accessed 17/02/06.

Huang, Hung-Tzu & Hsien-Chin Liou. 2007. Vocabulary learning in an automated graded reading program. *Language Learning & Technology*, 11(3): 64-82. http://llt.msu.edu/vol11num3/pdf/huangliou.pdf, accessed 28/11/07.

Johns, Tim. 1997. Contexts: The background, development and trialling of a concordance-based CALL program. In A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (eds), *Teaching and Language Corpora*. Harlow: Addison Wesley Longman, p. 100-115.

Johns, Tim. 2002. Data-driven learning: The perpetual challenge. In B. Kettemann & G. Marko (eds), *Teaching and Learning by Doing Corpus Analysis*. Amsterdam: Rodopi, p. 107-117.

Johns, Tim, Lee Hsingchin & Wang Lixun. 2008. Integrating corpus-based CALL programs and teaching English through children's literature. *Computer Assisted Language Learning*, 21(5): 483-506.

Kaszubski, Przemsław. 2011. IFAConc: A pedagogic tool for online concordancing with EFL/EAP learners. In A. Frankenberg-Garcia, L. Flowerdew & G. Aston (eds), *New Trends in Corpora and Language Learning*. London: Continuum, p. 81-104.

Kaur, Jagdish & Volker Hegelheimer. 2005. ESL students' use of concordance in the transfer of academic word knowledge: An exploratory study. *Computer Assisted Language Learning*, 18(4): 287-310.

Kennedy, Claire & Tiziana Miceli. 2001. An evaluation of intermediate students' approaches to corpus investigation. *Language Learning & Technology*, 5(3): 77-90. http://llt.msu.edu/vol5num3/pdf/kennedy.pdf, accessed 17/02/06.

Kennedy, Claire & Tiziana Miceli. 2002. The CWIC project: Developing and using a corpus for intermediate Italian students. In B. Kettemann & G. Marko (eds), *Teaching and Learning by Doing Corpus Analysis.* Amsterdam: Rodopi, p. 183-192.

Kennedy, Claire & Tiziana Miceli. 2010. Corpus-assisted creative writing: Introducing intermediate Italian learners to a corpus as a reference resource. *Language Learning & Technology*, 14(1): 28-44. http://llt.msu.edu/vol14num1/kennedymiceli.pdf, accessed 28/02/10.

Kettemann, Bernhard. 2011. Tracing the emo side of life: Using a corpus of an alternative youth culture discourse to teach cultural studies. In A. Frankenberg-Garcia, L. Flowerdew & G. Aston (eds), *New Trends in Corpora and Language Learning*. London: Continuum, p. 44-61.

Kettemann, Bernhard & Georg Marko 2011. Data-driving critical discourse analysis. In N. Kübler (ed.), *Corpora, Language, Teaching, and Resources: From Theory to Practice*. Bern: Peter Lang, p. 19-48.

Koo, Kyosung. 2006. *Effects of Using Corpora and Online Reference Tools on Foreign Language Writing: A Study of Korean Learners of English as a Second Language*. http://ir.uiowa.edu/cgi/viewcontent.cgi?article=1250&context=etd, accessed 02/04/12.

Koosha, Mansour & Ali Akbar Jafarpour. 2006. Data-driven learning and teaching collocation of prepositions: The case of Iranian EFL adult learners. *Asian EFL Journal Quarterly*, 8(4): 192-209. http://www.asian-efl-journal.com/December_2006_EBook.pdf, accessed 09/08/07.

Landure, Corinne 2011. Le data-driven learning: Apprendre et enseigner a contre-courant. *Mélanges Crapel*, 32: 163-178.

Landure, Corinne & Alex Boulton. 2010. Corpus et autocorrection pour l'apprentissage des langues. *ASp*, 57(1): 11-30.

Lavid, Julia. 2007. Contrastive patterns of mental transitivity in English and Spanish: A student-centred corpus-based study. In E. Hidalgo, L. Quereda & J. Santana (eds), *Corpora in the Foreign Language Classroom.* Amsterdam: Rodopi, p. 237-252.

Lee, Chuen-Yi & Hsien-Chin Liou. 2003. A study of using web concordancing for English vocabulary learning in a Taiwanese high school context. *English Teaching and Learning*, 27(3): 35-56.

Lee, David & John Swales. 2006. A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes*, 25: 56-75.

Lin, Ming-Chia. 2008. Building a lexical syllabus on Moodle with web concordancers for EFL productive academic vocabulary. *Proceedings of WorldCALL 2008*. Fukuoka: Fukuoka University. http://www.ntnu.edu.tw/acad/docmeet/97/a4/a405-1.pdf, accessed 07/11/10.

Liou, Hsien-Chin, Jason S. Chang, Hao-Jan Chen, Chih-Cheng Lin, Meei-Ling Liaw, Zhao-ming Gao, Jyi-Shing Roger Jang, Yuli Yeh, Thomas C. Chuang & Geeng-Neng You. 2006. Corpora processing and computational scaffolding for an innovative web-based English learning environment: The CANDLE project. *CALICO Journal*, 24(1): 77-95.

Liu, Dilin & Ping Jiang. 2009. Using a corpus-based lexicogrammatical approach to grammar instruction in EFL and ESL contexts. *Modern Language Journal*, 93: 61-78.

Ma, Bruce K.C. 1994. Learning strategies in ESP classroom concordancing: An initial investigation into data-driven learning. In L. Flowerdew & A. Tong (eds), *Entering Text*. Hong Kong: Hong Kong University of Science and Technology Language Centre, p. 197-214.

Maia, Belinda. 1997. Making corpora: A learning process. In G. Aston, L. Gavioli & F. Zanettin (eds), *Proceedings of Corpus Use and Learning to Translate*. http://www.sslmit.unibo.it/cultpaps/paps.htm, accessed 23/04/06 (via http://web.archive.org/).

Moreno Jaén, Maria. 2010. Developing university learners' collocational competence: An empirical corpus-based investigation. In M. Moreno Jaén, F. Serrano Valverde & M. Calzada Pérez (eds), *Exploring New Paths in Language Pedagogy: Lexis and Corpus-Based Language Teaching*. London: Equinox, p. 229-243.

Mparutsa, Cynthia, Alison Love & Andrew Morrison. 1991. Bringing concord to the ESP classroom. In T. Johns & P. King (eds), *Classroom Concordancing. ELR Journal*, 4: 115-134.

Okamoto, Kiyomi. 2010. Incorporating corpora into English language teaching for undergraduate computer science and engineering students with limited proficiency. *Proceedings of Professional Communication Conference (IPCC)*, p. 152-156.

O'Sullivan, Íde & Angela Chambers. 2006. Learners' writing skills in French: Corpus consultation and learner evaluation. *Journal of Second Language Writing*, 15(1): 49-68.

Osolsobě, Klára, & Pavlína Vališšová. 2012. Using data-driven methods in teaching Czech as a foreign language. In J. Thomas & A. Boulton (eds), *Input, Process and Product: Developments in Teaching and Language Corpora*. Brno: Masaryk University Press, p. 184-195.

Park, Kwanghyun & Celeste Kinginger. 2010. Writing/thinking in real time: Digital video and corpus query analysis. *Language Learning & Technology*, 14(3): 31-50. http://llt.msu.edu/issues/october2010/parkkinginger.pdf, accessed 15/10/10.

Pérez-Paredes, Pascual, María Sánchez Tornel, Jose María Alcaraz Calero & Pilar Aguada Jiménez. 2011. Tracking learners' actual uses of corpora: Guided vs non-guided corpus consultation. *Computer Assisted Language Learning*, 24(3): 233-253.

Philip, Gill. 2011. '…and I dropped my jaw with fear': The role of corpora in teaching phraseology. In N. Kübler (ed.), *Corpora, Language, Teaching, and Resources: From Theory to Practice*. Bern: Peter Lang, p. 49-68.

Rapti, Nikoletta. 2010. *A Study of Classroom Concordancing in the Greek Context: Data-Driven Grammar Teaching and Adolescent EFL Learners*. Unpublished PhD thesis. Nottingham: University of Nottingham. http://etheses.nottingham.ac.uk/1472/1/Thesis_-_Nikoletta_Rapti.pdf, accessed 24/04/12.

Rodgers, Ornaith, Angela Chambers & Florence LeBaron. 2011. Corpora in the LSP classroom: A learner-centred corpus of French for biotechnologists. *International Journal of Corpus Linguistics*, 16(3): 392-358.

Schaeffer-Lacroix, Eva. 2009. *Corpus Numériques et Production Ecrite en Langue Etrangère : Une Recherche avec des Apprenants d'Allemand*. Unpublished PhD thesis. Paris: Sorbonne Nouvelle Paris 3. http://tel.archives-ouvertes.fr/docs/00/43/90/95/PDF/Eva_Schaeffer-Lacroix_these_tome_1_CNU_V0.pdf, accessed 13/06/10.

Schmied, Joseph. 2006. Corpus linguistics and grammar learning: Tutor versus learner perspectives. In S. Braun, K. Kohn & J. Mukherjee (eds), *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods.* Frankfurt: Peter Lang, p. 87-106.

Seidhofer, Barbara. 2000. Operationalizing intertextuality: Using learner corpora for learning. In L. Burnard & T. McEnery (eds), *Rethinking Language Pedagogy from a Corpus Perspective*. Frankfurt: Peter Lang, p. 207-223.

Seidhofer, Barbara. 2002. Pedagogy and local learner corpora: Working with learner-driven data. In S. Granger, J. Hung & S. Petch-Tyson (eds), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins, p. 213-234.

Smith, Simon. 2011. Learner construction of corpora for general English in Taiwan. *Computer Assisted Language Learning*, 24(4): 291-316.

Smith, Simon, Alice Chen & Adam Kilgarriff. 2008. A corpus query tool for SLA: Learning Mandarin with the help of SketchEngine. In B. Lewandowska-Tomaszczyk (ed.), *Corpus Linguistics, Computer Tools, and Applications: State of the Art*. Frankfurt: Peter Lang, p. 673-686.

Sripicharn, Passapong. 2003. Evaluating classroom concordancing: The use of corpus-based materials by a group of Thai students. *Thammasat Review*, 8(1): 203-236. http://www.thammasatreview.tu.ac.th/tu_doc/2003-Volume8-No1/8[8].evaluating.pdf, accessed 09/05/12.

St. John, Elke. 2001. A case for using a parallel corpus and concordancer for beginners of a foreign language. *Language Learning & Technology*, 5(3): 185-203. http://llt.msu.edu/vol5num3/pdf/stjohn.pdf, accessed 14/02/06.

Stevens, Vance. 1991. Concordance-based vocabulary exercises: A viable alternative to gap-filling. In T. Johns & P. King (eds), *Classroom Concordancing. ELR Journal*, 4: 47-61.

Sun, Yu-Chih. 2003. Learning process, strategies and web-based concordancers: A case-study. *British Journal of Educational Technology*, 34(5): 601-613.

Sun, Yu-Chih. 2007. Learner perceptions of a concordancing tool for academic writing. *Computer Assisted Language Learning*, 20(4): 323-343.

Sun, Yu-Chih & Li-Yuch Wang. 2003. Concordancers in the EFL classroom: Cognitive approaches and collocation difficulty. *Computer Assisted Language Learning*, 16(1): 83-94.

Thorne, Steven L., Jonathon Reinhardt & Paula Golombek. 2008. Mediation as objectification in the development of professional academic discourse: A corpus-informed curricular innovation. In J. Lantolf & M. Poehner (eds), *Sociocultural Theory and the Teaching of Second Languages*. London: Equinox, p. 256-284.

Tian, Shiauping. 2005a. Data-driven learning: Do learning tasks and proficiency make a difference?' *Proceedings of the 9th Conference of the Pan-Pacific Association of Applied Linguistics.* Tokyo: Waseda University Media Mix Corp, p. 360-371. http://www.paaljapan.org/resources/documents.html, accessed 08/03/10.

Tian, Shiauping. 2005b. The impact of learning tasks and learner proficiency on the effectiveness of data-driven learning. *Journal of Pan-Pacific Association of Applied Linguistics*, 9(2): 263-275.

Todd, Richard Watson. (2001). Induction from self-selected concordances and self-correction. *System*, 29(1): 91-102.

Turnbull, Jill & Jack Burston. 1998. Towards independent concordance work for students: Lessons from a case study. *ON-CALL*, 12(2): 10-21.

Tyne, Henry. 2009. Corpus oraux par et pour l'apprenant. In A. Boulton (ed.), *Des Documents Authentiques Oraux aux Corpus: Questions d'apprentissage en didactique des langues. Mélanges CRAPEL*, 31: 91-111. http://revues.univ-nancy2.fr/melangesCrapel/, accessed 02/02/09.

van Halteren, Hans. 1994. Syntactic databases in the classroom. In A. Wilson & T. McEnery (eds), *Corpora in Language Education and Research*. *UCREL Technical Papers*, 4: 17-28.

Varley, Steve. 2009. I'll just look that up in the concordancer: Integrating corpus consultation into the language learning environment. *Computer Assisted Language Learning*, 22(2): 133-152.

Whistle, Jeremy 1999. Concordancing with students using an 'off-the-web' corpus. *ReCALL*, 11(2): 74-80.
Wu, Shaoqun, Ian H. Witten & Margaret Franken. 2010. Utilizing lexical data from a web-derived corpus to expand productive collocation knowledge. *ReCALL*, 22(1): 83-102.
Yeh, Yuli, Hsien-Chen Liou & Yi-Hsin Li. 2007. Online synonym materials and concordancing for EFL college writing. *Computer Assisted Language Learning*, 20(2): 131-152.
Yoon, Hyunsook. 2008. More than a linguistic reference: The influence of corpus technology on L2 academic writing. *Language Learning & Technology*, 12(2): 31-49. http://llt.msu.edu/vol12num2/yoon.pdf, accessed 15/03/09.
Yoon, Hyunsook & Alan Hirvela. 2004. ESL student attitudes toward corpus use in L2. *Journal of Second Language Writing*, 13(4): 257-283.