Learner Corpus Research 2011

Louvain-le-Neuve, 15-17 September 2011

If-conditionals in ICLE and the BNC: A language teaching success story?

Costas Gabrielatos

Lancaster University

c.gabrielatos@lancaster.ac.uk

Abstract

This paper aims to contribute to the methodological toolbox of "pedagogy-driven corpus-based research" (Gabrielatos, 2006), that is, research which is situated at the intersection of language description, pedagogical lexicogrammar, and pedagogical materials evaluation (e.g. Harwood, 2005; Hunston & Francis, 1998; Kennedy, 1992; Owen, 1993; Römer, 2004, 2005). The contribution of the present paper mainly lies in proposing a method of triangulating the corpus-based evaluation of lexicogrammatical information in EFL coursebooks, by way of examining a relevant corpus sample of learner written output.

More precisely, Gabrielatos (2006) compared the information and examples on *if*-conditionals in eleven coursebooks for advanced EFL learners with a random sample of 781 *if*-conditionals from the written BNC (Aston & Burnard, 1998) – using *BNCweb* (see Hoffmann et al., 2008). The analysis revealed that the common-ground typology – i.e. the information presented in all the coursebooks examined – accounted for just over one-quarter (27.8%) of the *if*-conditionals in the BNC sample. More importantly, even if the information given in all the sample coursebooks were collated to produce an inclusive typology, it would account for less than three-quarters (72.5%) of the *if*-conditionals in the BNC sample. Even lower proportions were revealed when the sample of coursebooks included both intermediate and advance coursebooks (Gabrielatos, 2003). The observed under-representation of the variety of *if*-conditionals in the coursebooks mainly hinged on the following:

- The coursebook typologies predominantly focused on conditionals with apodoses expressing degress of likelihood, ignoring or backgrounding conditionals with apodoses expressing deontic or volitional senses. Similalry, coursebooks ignored the type of conditionals termed "indirect" (Quirk et al., 1985), "speech act" (Sweetser, 1990), or "pragmatic" (Athanasiadou & Dirven, 1997) – e.g. Out of the corner of his eye he saw Hammond start forward. "But you promised ..." Spatz interrupted Hammond, his face hard. "I promised nothing, if you recall." [GUG 121].
- Patterns presented as 'exceptions' or 'special cases' in the coursebooks proved to be too frequent to be accurately described as such. For example, the coursebooks present Past tense marking with past time reference in protases as a special case – stressing its epistemic interpretation. However, in the BNC sample, one-third of Past tense marking in protases expressed past time.
- Modal marking in 'rules' and examples was predominantly by way of central and (less so) peripheral modals; lexical modal markers were ignored in both rules and examples.

On the basis of the above results, it was hypothesised in the present study that learner written production – when compared to similar texts in the written BNC – would be characterised by the following:

a) Under-representation of indirect conditionals.
b) Among direct conditionals, an over-representation of conditionals with apodoses expressing degrees of likelihood, and a corresponding under-representation of other types.
c) Lower proportion of Past tense marking with past time reference in protases.
d) Smaller extent of modal marking in protases.
e) Over-representation of central modals.

The study aimed to compare the *if*-conditionals in the random sample from the written BNC with those in a random sample from ICLE (Granger et al., 2002). However, ICLE only contains argumentative essays, whereas the written BNC is richer in text types. For reasons of comparability, only the instances from academic texts, essays and editorials in the BNC sample were considered – resulting in a sample of 195 *if*-conditionals. Therefore, a random sample of 200 instances was drawn from ICLE – using *CQPweb* (Hardie, forthcoming). Each *if*-conditional was annotated for its type, using the typology developed in Gabrielatos (2010), as well as the modal marker and the type of modality in the protases and apodoses.

Only hypotheses 'b', 'c' and, to some extent, 'd' were supported by the results – in the other two respects learner production in ICLE was comparable to that of the native speakers in the BNC. However, it would be premature to conclude that the explicit information in coursebooks has limited influence on learner production. ICLE contains the written production of learners having a variety of L1s, and coming from a variety of educational contexts, which can be expected to employ different pedagogical materials and/or instructional approaches. Therefore, the possibility cannot be discounted that the picture emerging from the present analysis may hide country-specific and/or L1-specific variation.

Motivation

Information in ELT materials a poor reflection of types of *if*-conditionals in the BNC (Gabrielatos, 2003, 2006).

Information in 11 advanced ELT coursebooks
vs.
Sample of 781 *if*-conditionals in the written BNC

- Basic: 13.8%
- Consensual: 27.8%
- Inclusive: 72.5%

Seems ok,
but is misleading

Format of ELT information
Zero, First, Second, Third, Mixed

Protasis	Apodosis
Tense - aspect marking Particular modal marker (usu. central modals)	

Construction	
Time reference	Degree of likelihood (truth/actuality/factuality)

'Special' cases

- Modal markers in the Protasis
  - *be to*
  - *could*
  - *should* (→ politeness)
  - *will* (→ insistence, willingness)
  - *would* (→ request)

- Modal markers other than central or marginal modals.

- *'If* + Past tense' with past time reference

Not included:
Indirect / Speech-Act / Pragmatic Conditionals
(Quirk et al., 1985; Sweetser, 1990; Athanasiadou & Dirven, 1997)

- Out of the corner of his eye he saw Hammond start forward. "But you promised ..." Spatz interrupted Hammond, his face hard. "I promised nothing, if you recall." [GUG 121]

- "Evidence is what the whole system is based on. If we cannot trust that, where are we?" [J10 2618]

- Frequency in written BNC sample: **10%**.

Not explicitly included: Conditionals with apodoses having non-epistemic functions

- Ability

  If I can live with them, so can everyone else. [FS9 2538]

- Obligation/Permission

  This is the best "bargain offer" pensioners have ever had, and any woman over 60 or man over 65 should take advantage of it if possible. [C8Y 946]

- Volition

  If anything can be salvaged from the tragedy it's hoped the publicity surrounding his death will help his work become more well known. [K21 3757]

Zero and First: two sides of the same coin

	General / Timeless	Specific / Future
Zero	The argument obviously generalises to show that, if there is a non-negative solution of (9.8) with <gap desc=formula>, then any new tableau obtained by pivoting in column j is efficient. [CA4 738]	If Bridges is right, this still does not avoid possible legal argument over the "reasonableness" of the contract between purchaser and provider, nor over how well contracts are complied with. [CR5 693]
First	If a Troll suffers harm his flesh will almost instantly re-grow. [CMC 250]	"If they charge the wrong man, it'll make a difference to him!" said Melissa dryly. [HNJ 1807]

Resulting hypotheses

Learner written production – when compared to NS texts – would be characterised by the following:

a) Under-representation of indirect conditionals.
b) Over-representation of conditionals with apodoses expressing degrees of likelihood.
c) Lower proportion of Past tense marking with past time reference in protases.
d) Smaller extent of modal marking in protases.
e) Over-representation of central modals in apodoses.

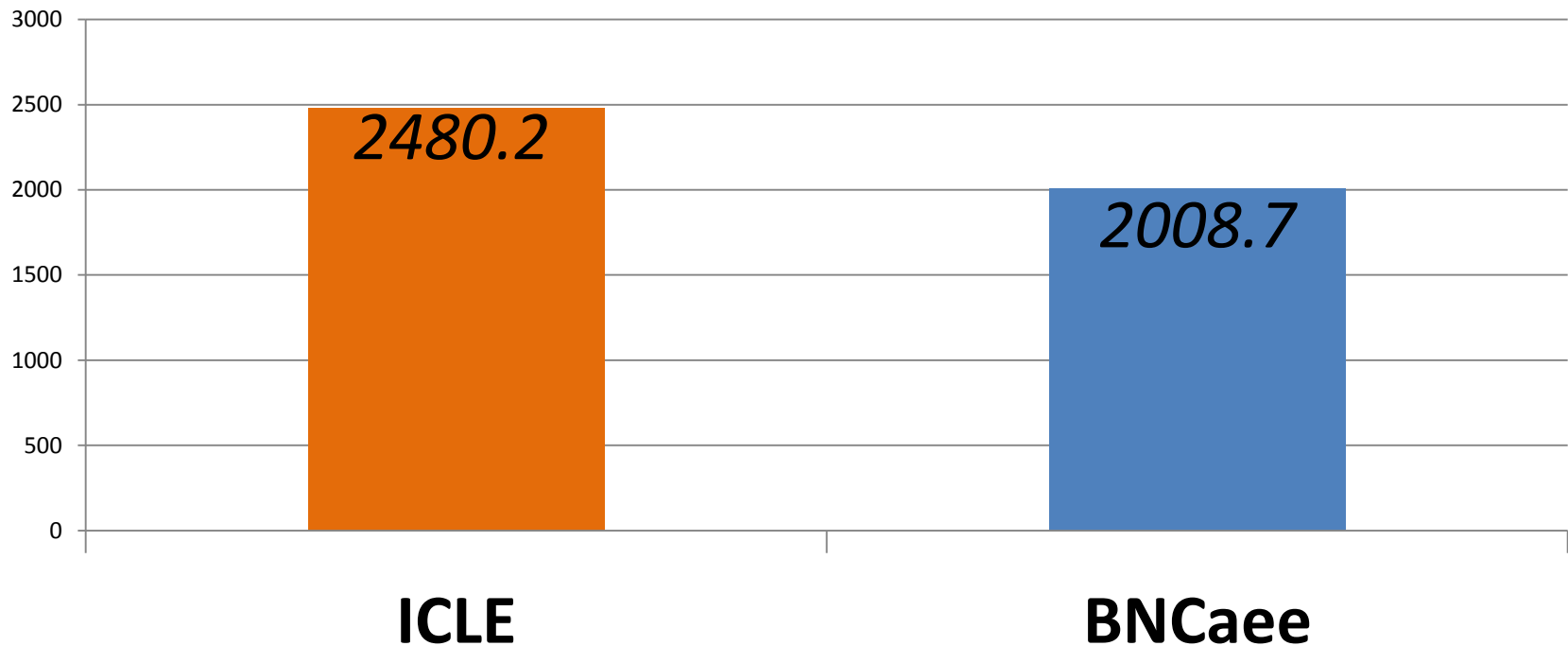• Plus, other relevant observations.

Triangulation
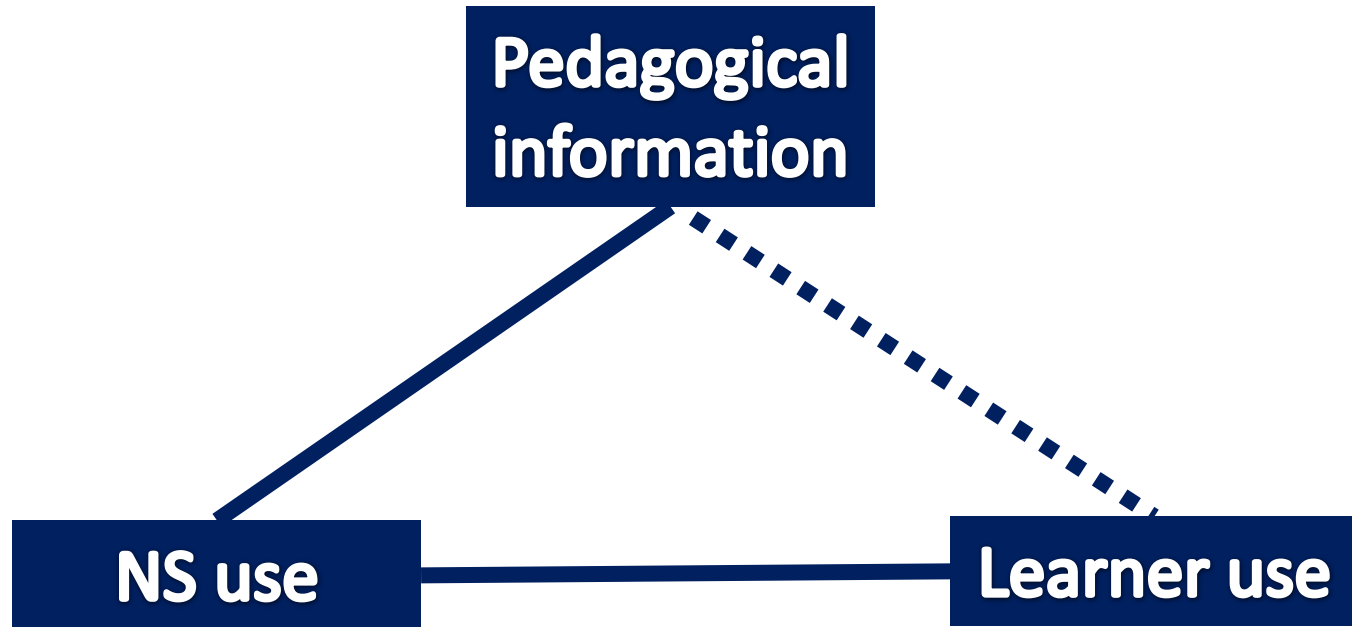
Corpus samples

- For comparability, same BNC sample used in previous study.

However

- ICLE : argumentative essays
  - → Random sample from whole BNC not appropriate.

- Instances from academic texts, essays and editorials

- BNCaee: 195
- ICLE: 190

Relative frequencies of *if*-conditionals

- ICLE has 23.5% more *if*-conditionals than BNCaee
  - LL = 249.88, *p*<10^{-17}
- Prominence in ELT materials → Learners over-use …
  - conditionals
  - *if*-conditionals (and under-use other conditionals)

Proportion of indirect conditionals

	IND	Constructions	*%IND*
ICLE	6	190	***3.2***
BNCaee	7	195	***3.6***

- Comparable proportions.

However

- Academic writing not the best context for IND.
  → Analysis of spoken sample (watch this space.)
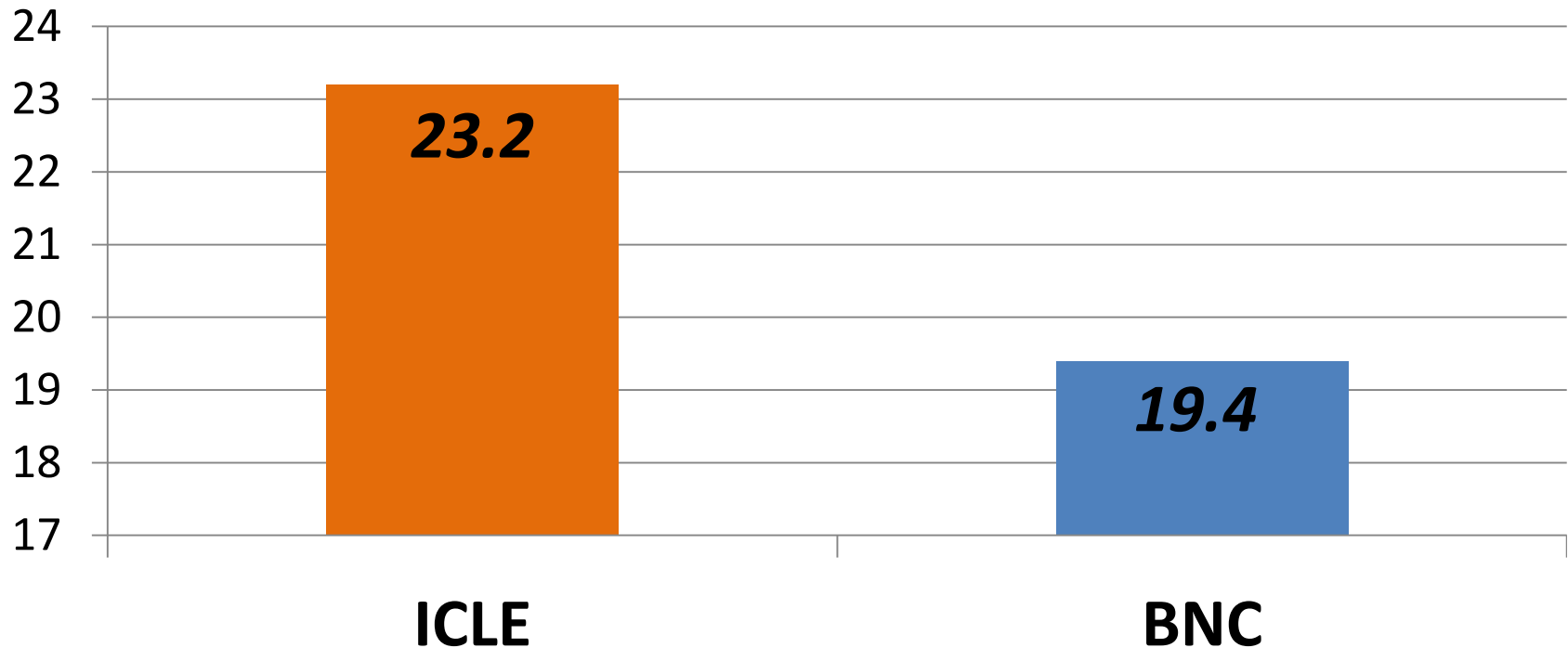
Sub-types of direct conditionals

	ICLE (n=184)		BNC (n=188)		*%Diff*	Stat. sig.
LK	131	*71.2*	120	*63.8*	*+11.6%*	not sig.
PP	21	*11.4*	17	*9.0*	*+26.7%*	not sig.
DD	29	*15.8*	50	*26.6*	***-40.6%***	**LL=6.59,** *p<0.05*
DN	1	*0.5*	1	*0.5*	*No diff.*	NA

LK + PP	152	*82.6*	137	*72.8*	***+13.5%***	**LL=5.1,** *p<0.05*

- ELT materials seem to have an effect on the DIR sub-types used by ICLE learners.

Zero conditional

%Zero

- ICLE has about 20% more Zero conditionals.
- Not stat. sig. due to small number of instances (44, 38) and small samples.

Past tense in protases:
temporal vs. modal use

	Temporal	*%Temporal*
ICLE (n=35)	10	*28.6*
BNC (n=42)	17	*40.5*

- BNC has 42% higher frequency of temporal uses of Past tense in protases.
  - Difference is not stat. sig. (LL=1.2), but this is due to the very small sub-sample (number of protases marked for Past tense).
  - For p<0.05 the same difference needs to be shown in a sample four times larger (watch this space.)

Proportion of modal marking in Protases
Two complementary metrics

Modal Density

Definition	Average number of modal markings per **clause**.
Expression	Number of modal markings per 100 clauses. (%)
Utility	Helps comparisons between samples by normalising for the complexity of the constructions in each.

(Gabrielatos, 2008, 2010)

Lexical Density:

- The average number of content words per clause (Halliday, 2004: 654-655).

- The percentage of the tokens in a text that are content words (Ure, 1971).

Modalisation Spread

Definition	Proportion of constructions that carry at least one modal marking.
Expression	Proportion (%) of modalised constructions.
Utility	Corrects for heavily modalised constructions in the sample.

<div align="right">(Gabrielatos, 2010)</div>

Spread:

- The proportion of corpus speakers who use a particular language item (Gabrielatos & Torgersen, 2009; Gabrielatos et al., 2010).

Proportion of modal marking in Protases

	Modalisations	Clauses	*MD*
ICLE	81	247	*32.8*
BNC	105	266	*39.5*

- BNC protases have 20.5% higher MD …
- but difference not sig. (LL=2.48)

	Modalised	Constructions	*MS*
ICLE	68	190	*35.8*
BNC	80	195	*37.9*

- BNC protases have 14.5% higher MS …
- but difference not sig. (LL=0.90)

Proportion of central modals in Apodosis
(can, could, may, might, must, shall, should, will, would)

	Central modals	Total modal markers	*%central modals*
ICLE	114	163	*69.9*
BNC	143	206	*69.5*

- Almost identical

What about the hypotheses?

a) Under-representation of indirect conditionals in ICLE.
b) Over-representation of conditionals with apodoses expressing degrees of likelihood in ICLE.
c) Lower proportion of Past tense marking with past time reference in ICLE protases.
d) Smaller extent of modal marking in ICLE protases.
e) Over-representation of central modals in ICLE apodoses.

- Plus, other relevant observations.

What about the hypotheses?

a) ~~Under-representation of indirect conditionals in ICLE.~~
b) **Over-representation of conditionals with apodoses expressing degrees of likelihood in ICLE.**
c) Lower proportion of Past tense marking with past time reference in ICLE protases.
d) Smaller extent of modal marking in ICLE protases.
e) ~~Over-representation of central modals in ICLE apodoses.~~

f) **Higher frequency of *if*-conditionals in ICLE.**
g) Higher frequency of Zero in ICLE.

Summary → Interpretation → Hypotheses

Some indications of effect of pedagogical information on learner output – but not across the board.

- What is taught is not necessarily what is learned.

- Ped. info 'diluted' in evidence from texts: pedagogical (e.g. coursebooks) or authentic (e.g. web. TV).

- Learners taught through materials diverging from ELT norms in some respects.

- Teachers adapted / supplemented ped. materials.

- Influence of particular L1.

- BNCaee not an appropriate reference corpus.

	NS	NNS
Expert	BNCaee	
Novice		ICLE

Further steps

→ Comparison with novice L1 writers (e.g. LOCNESS).

→ Comparison of NS and NNS spoken corpora.

→ Use of larger corpus samples.

→ Comparison between learners with different L1s.

→ Separate examination of different countries / educational contexts.

→ Examination of conditionals with other subordinators (e.g. *assuming, provided*)

If you have any questions,
I'd be happy to answer them

References

Athanasiadou, A. & R. Dirven (1997). Conditionality, hypotheticality, counterfactuality. In A. Athanasiadou & R. Dirven (Eds.) *On Conditionals Again* (pp. 61–96). Amsterdam: John Benjamins.

Gabrielatos, C. (2003). Conditional sentences: ELT typology and corpus evidence. Paper given at 36th Annual BAAL Meeting, University of Leeds, UK, 4-6 September 2003. [http://eprints.lancs.ac.uk/140/]

Gabrielatos, C. (2006). Corpus-based evaluation of pedagogical materials: *If*-conditionals in ELT coursebooks and the BNC. 7th *Teaching and Language Corpora* Conference, Paris, France, 1-4 July 2006. [http://eprints.lancs.ac.uk/882/]

Gabrielatos, C. (2008). If-conditionals, modality, and Schrödinger's cat. Paper presented at the RITL Research Group (Research in Theoretical Linguistics), Department of Linguistics and English Language, Lancaster University, 27 November 2008. [http://eprints.lancs.ac.uk/39788/]

Gabrielatos, C. (2010). A corpus-based examination of English if-conditionals through the lens of modality: Nature and types. PhD Thesis. Lancaster University, UK.

Gabrielatos, C. & Torgersen, E. (2009). A corpus-based sociolinguistic analysis of indefinite article use in London English. ICAME 30, Lancaster, UK, 27-31 May 2009. [http://eprints.lancs.ac.uk/26562/]

Gabrielatos, C., Torgersen, E., Hoffmann, S. & Fox, S. (2010). A corpus-based sociolinguistic study of indefinite article forms in London English. Journal of English Linguistics, 38(4), 297-334.

Granger, S., Dagneux, E. & Meunier, F. (2002). *The International Corpus of Learner English. Version 1.1*. Handbook and CD-ROM. Louvain-la-Neuve: Presses Universitaires de Louvain.

Hardie, A. (forthcoming). CQPweb – compining power, flexibility and usability in a coprus analysis tool.

Harwood, N. (2005). What do we want EAP teaching materials for? *Journal of English for Academic Purposes* 4(2), 149–161.

Hoffmann, S., Evert, S., Smith, N., Lee, D. & Berglund-Prytz, Y. (2008). *Corpus Linguistics with BNCweb − a practical guide.* Frankfurt am Main: Peter Lang.

Hunston, S. & Francis, G. (1998). Verbs observed: A corpus-driven pedagogic grammar. *Applied Linguistics* 19(1): 45-72.

Kennedy, G. (1992). Preferred ways of putting things with implications for language teaching. In J. Svartvik (Ed.) *Directions in Corpus Linguistics: Proceedings of the Nobel Symposium 82, Stockholm, 4-8 August 1991* (pp. 335-378). Berlin: Mouton de Gruyter.

Owen, C. (1993). Corpus-based grammar and the Heineken effect: Lexico-grammatical description for language learners. *Applied Linguistics* 14(2): 167-187.

Quirk, R., S. Greenbaum, G. Leech & J. Svartvik (1985). *A Comprehensive Grammar of the English Language*. London: Longman.

Römer, U. (2004). Textbooks: A corpus-driven approach to modal auxiliaries and their didactics. In J. McH. Sinclair (Ed.) *How to Use Corpora in Language Teaching* (pp. 185–199). Amsterdam: John Benjamins.

Römer, U. (2005). *Progressives, Patterns, Pedagogy*: *A corpus-driven approach to progressive forms, functions, contexts and didactics*. Amsterdam: John Benjamins.

Sweetser, E.E. (1990). *From Etymology to Pragmatics: Metaphorical and cultural aspects of semantic structure*. Cambridge: Cambridge University Press.