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 degrees of likelihood, ignoring or backgrounding conditionals with apodoses expressing deontic or volitional senses. Similarly, 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-condition 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

<table>
<thead>
<tr>
<th>Protasis</th>
<th>Apodosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tense - aspect marking</td>
<td>Particular modal marker (usu. central modals)</td>
</tr>
</tbody>
</table>

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
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

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

Pedagogical information

NS use

Learner use
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 $\rightarrow$ Learners over-use ...
  - conditionals
  - *if*-conditionals (and under-use other conditionals)
Proportion of indirect conditionals

<table>
<thead>
<tr>
<th></th>
<th>IND</th>
<th>Constructions</th>
<th>%IND</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICLE</td>
<td>6</td>
<td>190</td>
<td>3.2</td>
</tr>
<tr>
<td>BNCae</td>
<td>7</td>
<td>195</td>
<td>3.6</td>
</tr>
</tbody>
</table>

• Comparable proportions.

However

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

<table>
<thead>
<tr>
<th></th>
<th>ICLE (n=184)</th>
<th>BNC (n=188)</th>
<th>%Diff</th>
<th>Stat. sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>LK</td>
<td>131 71.2</td>
<td>120 63.8</td>
<td>+11.6%</td>
<td>not sig.</td>
</tr>
<tr>
<td>PP</td>
<td>21 11.4</td>
<td>17 9.0</td>
<td>+26.7%</td>
<td>not sig.</td>
</tr>
<tr>
<td>DD</td>
<td>29 15.8</td>
<td>50 26.6</td>
<td>-40.6%</td>
<td>LL=6.59, p&lt;0.05</td>
</tr>
<tr>
<td>DN</td>
<td>1 0.5</td>
<td>1 0.5</td>
<td>No diff.</td>
<td>NA</td>
</tr>
<tr>
<td>LK + PP</td>
<td>152 82.6</td>
<td>137 72.8</td>
<td>+13.5%</td>
<td>LL=5.1, p&lt;0.05</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>Temporal</th>
<th>%Temporal</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICLE (n=35)</td>
<td>10</td>
<td>28.6</td>
</tr>
<tr>
<td>BNC (n=42)</td>
<td>17</td>
<td>40.5</td>
</tr>
</tbody>
</table>

- 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

<table>
<thead>
<tr>
<th>Definition</th>
<th>Average number of modal markings per clause.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expression</td>
<td>Number of modal markings per 100 clauses. (%)</td>
</tr>
<tr>
<td>Utility</td>
<td>Helps comparisons between samples by normalising for the complexity of the constructions in each.</td>
</tr>
</tbody>
</table>

(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

<table>
<thead>
<tr>
<th>Definition</th>
<th>Proportion of constructions that carry at least one modal marking.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expression</td>
<td>Proportion (%) of modalised constructions.</td>
</tr>
<tr>
<td>Utility</td>
<td>Corrects for heavily modalised constructions in the sample.</td>
</tr>
</tbody>
</table>

(Gabrielatos, 2010)

**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

<table>
<thead>
<tr>
<th></th>
<th>Modalisations</th>
<th>Clauses</th>
<th>MD</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICLE</td>
<td>81</td>
<td>247</td>
<td>32.8</td>
</tr>
<tr>
<td>BNC</td>
<td>105</td>
<td>266</td>
<td>39.5</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>Modalised</th>
<th>Constructions</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICLE</td>
<td>68</td>
<td>190</td>
<td>35.8</td>
</tr>
<tr>
<td>BNC</td>
<td>80</td>
<td>195</td>
<td>37.9</td>
</tr>
</tbody>
</table>

- 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)

<table>
<thead>
<tr>
<th></th>
<th>Central modals</th>
<th>Total modal markers</th>
<th>%central modals</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICLE</td>
<td>114</td>
<td>163</td>
<td>69.9</td>
</tr>
<tr>
<td>BNC</td>
<td>143</td>
<td>206</td>
<td>69.5</td>
</tr>
</tbody>
</table>

- 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.
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.
<table>
<thead>
<tr>
<th></th>
<th>NS</th>
<th>NNS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert</td>
<td>BNCae</td>
<td></td>
</tr>
<tr>
<td>Novice</td>
<td></td>
<td>ICLE</td>
</tr>
</tbody>
</table>
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


