If-conditionals as modal colligations
A corpus-based investigation

Costas Gabrielatos
Lancaster University
c.gabrielatos@lancaster.ac.uk

This paper examines the case for treating if-conditionals as strong attractors of modality. A stronger claim to be examined is that if-conditionals, and if-constructions in general, can be seen as modal colligations. The main research questions are:

- Do if-conditionals contain a statistically significant higher frequency of modal expressions than average?
- Do if-conditionals show a statistically significant higher frequency of modal expressions compared to non-conditional if-constructions?

This examination is theoretically informed by three compatible notions: grammatical construction, colligation, and semantic preference. A grammatical construction is a “syntactic pattern which is assigned one or more conventional functions” (Fillmore, 1988: 36). Colligation was initially defined as the co-occurrence of grammatical categories (Firth, 1968: 181), and has recently been adapted to refer to the co-occurrence of lexis and grammatical categories (e.g. Hoey, 1997: 8). Semantic preference is the “relation between a lemma or word-form and a set of semantically related words.” (Stubbs, 2002: 65). These notions can combine and expand into the notion of semantic colligation: the mutual attraction holding between a grammatical construction (in this case, if-conditionals - see Fillmore, 1986) and a semantic category (in this case, modality - hence modal colligation).

The claim is tested through keyword comparisons of un-annotated corpora: a sample of 1,000 if-constructions from the written BNC, the written BNC Sampler, FLOB, all the if-sentences from the written BNC, and the non-conditional if-sentences from the sample. Further tests involve frequency comparisons of specific modal words between the manually annotated sample and the annotated versions of BNC, BNC Sampler and FLOB, as well as a collocational analysis of if in the written BNC. The paper will also comment on methodological issues arising from the keyword analysis, as well as issues pertaining to corpus annotation, quantitative analysis, the nature of if-conditionals, and the role of if.

References


**Hypotheses and research questions**

- *If*-conditionals are strong attractors of modality.
- *If*-conditionals can be regarded as modal colligations.
- Do *if*-conditionals contain a statistically significant higher frequency of modal expressions than average?
- Do *if*-conditionals contain a statistically significant higher frequency of modal expressions compared to non-conditional *if*-constructions?
Modality and modal expressions

- Modality is “concerned with the speaker’s attitude towards the factuality or actualisation of the situation expressed by the rest of the clause” (Huddleston & Pullum, 2002: 173).

- Accounts of modality seem to converge on modality expressing attitude towards actuality, factuality, likelihood, ability, potentiality and desirability (e.g. volition, obligation, permission).

- Modality can be expressed through a variety of formal means:
  - modal auxiliaries (e.g. may, ought to)
  - catenative verbs (e.g. need, want)
  - adverbs (e.g. possibly, probably)
  - the imperative
  - the past tense (in some contexts, e.g. conditionals)
  - constructions involving …
    - lexical verbs (e.g. it appears that …).
    - adjectives (e.g. it is likely that …; it is imperative that …),
    - nouns (e.g. there is a chance that …; we have an obligation to …).

Informing concepts

Semantic preference
- The “relation between a lemma or word-form and a set of semantically related words.” (Stubbs, 2001: 111)

Colligation
- Co-occurrence of grammatical categories. (Firth, 1968: 181)
- Co-occurrence of lexis and grammatical categories. (Stubbs, 2001: 112)
- “The grammatical company a word keeps.” (Hoey, 1997: 8)
Modal colligation

- A hybrid between colligation and semantic preference.
- In more general terms it could be termed 'semantic colligation'.
- The mutual attraction holding between a grammatical construction, if-conditionals, and "a set of semantically related words" (Stubbs, 2001: 111), or, more generally, a semantic category: modality.

Corpora

- Sample of 853 if-conditionals from the written BNC (Sample)
- The non-conditional if-sentences from the initial sample.
- All if-sentences (s-units) from the written BNC (If-BNC)
- The written BNC (BNCw)
- The written BNC Sampler (BNCSw)
- FLOB
Methodology

- Sample manually annotated for:
  - Form (tense/aspect marking, modal expression)
  - Meaning (modal notion, modality type)
  - Type of conditional (semantic/pragmatic relation between the two clauses)

- Frequency of modalisation in the conditional and main clauses in the Sample.

- Keyword analyses (LL ≥ 6.63, p ≤ 0.01):

- Establishing keyness of individual modals was a means to an end used as an indication of significant higher frequency of modal marking.

Sample: if-clause modalisation

<table>
<thead>
<tr>
<th>Category</th>
<th>Freq.</th>
<th>% (n=853)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modalised</td>
<td>280</td>
<td>32.8%</td>
</tr>
<tr>
<td>Unmodalised</td>
<td>570</td>
<td>66.8%</td>
</tr>
<tr>
<td>Elliptical (non-inferable)</td>
<td>3</td>
<td>0.4%</td>
</tr>
<tr>
<td>Total</td>
<td>853</td>
<td>100%</td>
</tr>
</tbody>
</table>

- One-third of if-clauses are modalised …
- … in addition to the modalisation through if.
- 1% have two or more modal markers.
Almost three-quarters of main clauses are modalised.
7% of all main clauses have two or more modal markers.

More than half of the clauses in the sample are modalised.
On average: one modalisation per if-conditional.
However, this may not be significantly higher than average.
Keyword analysis (1)

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Modal keywords</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive</td>
<td>Negative</td>
<td>Positive</td>
<td>Negative</td>
<td></td>
</tr>
<tr>
<td>Sample - BNCSw</td>
<td>27</td>
<td>0</td>
<td>3.09</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Sample - FLOB</td>
<td>21</td>
<td>1</td>
<td>2.77</td>
<td>1.39</td>
<td></td>
</tr>
</tbody>
</table>

Keywords include:

- All central modals
- Marginal modals (e.g. be able to, be unable to, need, want)
- Lexical verbs (e.g. comply, doubted, feel, know/knew, required, think/thinks)
- Adjectives (e.g. necessary, willing)
- Adverbs (e.g. probably, hopefully)
- Nouns (e.g. evidence, obligation)
- Other constructions (e.g. (be) liable (to))

Keyword analysis (2)

- *Is the apparent semantic attraction a characteristic of if-conditionals in general, or of the makeup of the if–conditionals in the sample?*

  - KW comparison: Sample - If-BNCw.
  - One positive (shall, 0.19%) and one negative (wants, 2.78%) modal keyword.
  - Sample not richer in modality.
Keyword analysis (3)

Would the comparison of a larger sample of *if*-conditionals with BNCSw and FLOB support the hypotheses?

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Modal keywords</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Positive</td>
<td>Negative</td>
<td>Positive %</td>
</tr>
<tr>
<td>Sample - BNCSw</td>
<td>27</td>
<td>0</td>
<td></td>
<td>3.09</td>
</tr>
<tr>
<td>Sample - FLOB</td>
<td>21</td>
<td>1</td>
<td></td>
<td>2.77</td>
</tr>
<tr>
<td><em>if</em>-BNCw - BNCSw</td>
<td>93</td>
<td>6</td>
<td></td>
<td>4.47</td>
</tr>
<tr>
<td><em>if</em>-BNCw - FLOB</td>
<td>63</td>
<td>9</td>
<td></td>
<td>3.92</td>
</tr>
</tbody>
</table>

- Higher proportion of modal keywords.
- Higher relative significance:
  - In Sample-BNCSw/FLOB, 25% of positive KWs were among the top 25% of all KWs.
  - In *if*-BNCw-BNCSw/FLOB, more than 50% of positive KWs were among the top 25% of all KWs.

Keyword analysis (4)

Is the attraction to modality a feature of conditionality, or of the collocational profile of the word *if*?

- KW comparison: conditional with non-conditional *if*-units in the sample.
- Only one positive modal keyword (*may*), representing 8.3% of all KWs.
- No negative modal keywords.
- Conditional and non-conditional *if*-sentences do not seem to differ in terms of modal load.
Manual keyword comparison of annotated corpora

- **Focus**: modality, *not* specific modal expressions.

- **Ideal**: totalling all modal expressions (lexical and grammatical) and comparing their frequency in the sample and reference corpora.

- **Feasible**: Keyness of central modals taken as a group.
  - Central modals in the Sample represent only 12% of modal types, but account for almost 60% of modal tokens.
  - Comparison of frequency of group of central modals in the Sample and the annotated reference corpora.

Relative frequencies of the group of central modals

<table>
<thead>
<tr>
<th>Sample</th>
<th>BNCw</th>
<th>BNCSw</th>
<th>Diff. %</th>
<th>LL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freq.</td>
<td>Freq./mil.</td>
<td>Freq.</td>
<td>Freq./mil.</td>
<td></td>
</tr>
<tr>
<td>562</td>
<td>22,295.39</td>
<td>1,176,108</td>
<td>13,474.40</td>
<td>65.5%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample</th>
<th>BNCSw</th>
<th>FLOB</th>
<th>Diff. %</th>
<th>LL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freq.</td>
<td>Freq./mil.</td>
<td>Freq.</td>
<td>Freq./mil.</td>
<td></td>
</tr>
<tr>
<td>562</td>
<td>22,295.39</td>
<td>15,008</td>
<td>13,868.42</td>
<td>60.8%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample</th>
<th>FLOB</th>
<th>Diff. %</th>
<th>LL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freq.</td>
<td>Freq./mil.</td>
<td>Freq.</td>
<td>Freq./mil.</td>
</tr>
<tr>
<td>562</td>
<td>22,295.39</td>
<td>12,994</td>
<td>12,731.43</td>
</tr>
</tbody>
</table>

\[ p \leq 10^{-14} \]
Problems with counting within constructions

- Discrepancies between Sample frequency counts in the automatic and manual KW analysis.
- Text portions not belonging to the construction.
  - Overestimation of sample size.
  - Underestimation of keyness.

Additional elements

- (1) Yes, I come from Lochaber, and the Lochaber people, if they were here, would be at one with the people of Breadalbane.
- (2) If the leg is cured while it is still attached, it is technically a gammon -- hence the confusion caused by the term "gammon ham".
- The elements not strictly belonging to the conditional account for 27.3% and 37.5% of the words in (1) and (2) respectively.
Embedded *if*-conditionals

- Why should the fact that D was engaged on causing damage to property at the time (even damage to D's own property) make his conduct into an offence punishable with life imprisonment when, if D were engaged on some other activity, it would not be punishable as such and would only amount to manslaughter if a death happened to be caused?
  
  - To maintain sample randomness, only the conditional sentence containing the *if* picked out by the *thin* function of BNCweb was taken into account and annotated

- As an academic critic and university teacher specializing in modern literature and literary theory, I spend much of my time these days reading books and articles that I can barely understand and that cause my wife (a graduate with a good honours degree in English language and literature) to utter loud cries of pain and nausea if her eye happens to fall on them. [A1A 208]

Discontinuous *if*-conditionals: Stem-List

**Abdomen: When to seek advice**

Urgently, Right now!
…………………………………………

If the stool is bloody, black or tar-like.

[B1R 681]
Conclusions and further steps

- Seen as a single group, *if*-conditionals seem to contain modality significantly more frequently than average. However ...

- Colloca tional analysis of *if*.

- Examination of different types of *if*-conditionals.

- Further examination of non-conditional *if*-sentences:
  - larger sample of
  - different types (e.g. indirect questions).

- Spoken language

- Different genres