Lexicogrammar: Lexical Grammar or Construction Grammar?

Two corpus-based case studies

Costas Gabrielatos
Edge Hill University
Motivation

Search for theoretical explanations of corpus-based results:

• Modal load of conditional and non-conditional structures.

• Lexicogrammatical patterns of (BE) interested (incl. modality).
Core premise

Theories are there to be tested, not applied – and definitely not consumed, parroted, worshipped, or brandished.

https://twitter.com/congabonga/status/95514338944753664?s=20
Lexicogrammar


• Lexis and grammar seen as “complementary perspectives” (1991: 32)


• “[I]f you interrogate the system grammatically you will get grammar-like answers and if you interrogate it lexically you get lexis-like answers” (1992: 64).
Lexical Grammar


• Sinclair (1991) posited the distinction between the *idiom principle* (exemplified by *collocation*) and the *open-choice principle* (words fill in particular syntactic positions).
  – The two are presented as operating alongside each other.
  – The idiom principle accounts for “the restraints that are not captured by the open-choice model” (1991: 115) – later formalised as *Lexical Grammar* (2004).

• Collocation is defined as “a purely lexical relation, non-directional and probabilistic, which ignores any syntactic relation between the words” (Stubbs, 2001: 64).
However ...

Halliday on Sinclair’s approach:
• [Sinclair] is “tunnelling through the system interrogating it lexically while moving further and further towards the grammatical end” (1992: 64) in order to identify aspects of language use that cannot be derived from a purely grammatical analysis (1966: 410).

Sinclair on Halliday’s approach:
• Lexicogrammar is “fundamentally grammar with a certain amount of attention to lexical patterns within the grammatical frameworks; it is not in any sense an attempt to build together a grammar and lexis on an equal basis.” (2004: 164).
However ...

No/little consideration of the open choice principle in subsequent studies on Lexical Grammar.

The *lexical item* (Sinclair, 1996: 75; Stubbs, 2009: 123-126) consists of a *core* (i.e. a word or phrase) and its ...

- collocates
- semantic preference (optional)
- semantic prosody
- colligations (optional)

These components are seen as belonging to the core (Hunston & Francis, 2000: 3, 49).

Lexical Grammar has “restored lexis in its rightful place at the centre of language description” (Hunston & Francis, 2000: 253).
The construct of colligation was redefined in a manner consistent with LG tenets.

**Original definition** (Firth, 1968: 181)

- “The statement of meaning at the grammatical level is in terms of word and sentence classes or of similar categories and of the interrelation of those categories in colligations. Grammatical relations should not be regarded as relations between words as such – between *watched* and *him* in ‘I watched him’ – but between a personal pronoun, first person singular nominative, the past tense.”

**Re-definition**

- “[T]he grammatical company a word keeps” (Hoey, 1997: 8).
- “[T]he relation between content and function words, and between words and grammatical categories” (Stubbs, 2002: 238).
However ...

The utility of collocation was expanded in a manner consistent with LG tenets.

Original conception

• Firth (1957: 195-196) proposed collocation as an approach to establishing meaning, distinguishing “meaning by collocation” from the "conceptual or idea approach to the meaning of words".

• Simply put, the proposal was that the meaning of words is defined by “the company they keep” (Firth 1957: 11).

Re-conceptualisation

• Grammar emerges from the interaction and patterning of lexis in discourse (Hoey 2005: 1; Sinclair 1991: 100)
Main features of Lexical Grammar:

• Lexis is (at) the core of language description.
• Grammar emerges from lexical patterning.
• Lexical and grammatical patterns belong to (lexical) cores.

⇒ Primacy of lexis.
⇒ Lexis and grammar are not treated “on an equal basis”.
Also ...

• In LG, collocation is defined as the co-occurrence of word-forms, as different forms of a word can have different sets of collocates (e.g. Sinclair, 1991: 53-56).

⇒ But this can be re-stated as ‘morphological marking affects collocation patterns’.

⇒ Collocation is not purely lexical, but is influenced by grammar.

⇒ Collocation is lexicogrammatical (Gabrielatos, 2018: 244)
Case Study 1

Modal load of conditional and non-conditional structures

# Modal Load

<table>
<thead>
<tr>
<th>Definition</th>
<th>Modal Density</th>
<th>Modalisation Spread</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average number of modal markings per clause.</td>
<td></td>
<td>Proportion of constructions that carry at least one modal marking.</td>
</tr>
</tbody>
</table>

| Expression | | Proportion (%) of modalised constructions. |
| Number of modal markings per 100 clauses. | |

| Utility | Helps comparisons by normalising for the complexity of the constructions in the sample. | Corrects for heavily modalised constructions in the sample. |

(Gabrielatos, 2006, 2010)
## Corpus Samples (BNCw)

<table>
<thead>
<tr>
<th>Code</th>
<th>Content</th>
<th>N s-units</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>if</em>-cnd</td>
<td>Conditionals with <em>if</em></td>
<td>959</td>
</tr>
<tr>
<td><em>assuming</em>-cnd</td>
<td>Conditionals with <em>assuming</em></td>
<td>727</td>
</tr>
<tr>
<td><em>in_case</em>-cnd</td>
<td>Conditionals with <em>in case</em></td>
<td>945</td>
</tr>
<tr>
<td><em>provided</em>-cnd</td>
<td>Conditionals with <em>provided</em></td>
<td>859</td>
</tr>
<tr>
<td><em>supposing</em>-cnd</td>
<td>Conditionals with <em>supposing</em></td>
<td>213</td>
</tr>
<tr>
<td><em>on_condition</em>-cnd</td>
<td>Conditionals with <em>on condition</em></td>
<td>205</td>
</tr>
<tr>
<td><em>unless</em>-cnd</td>
<td>Conditionals with <em>unless</em></td>
<td>989</td>
</tr>
<tr>
<td><em>even_if</em>-cc</td>
<td>Conditional-concessives with <em>even if</em></td>
<td>995</td>
</tr>
<tr>
<td><em>whether</em>-cc</td>
<td>Conditional-concessives with <em>whether</em></td>
<td>184</td>
</tr>
<tr>
<td><em>if</em>-q</td>
<td>Indirect interrogatives with <em>if</em></td>
<td>978</td>
</tr>
<tr>
<td><em>whether</em>-q</td>
<td>Indirect interrogatives with <em>whether</em></td>
<td>809</td>
</tr>
<tr>
<td><em>as if</em>-c</td>
<td>Structures of comparison with <em>as if</em></td>
<td>995</td>
</tr>
<tr>
<td><em>as though</em>-c</td>
<td>Structures of comparison with <em>as though</em></td>
<td>999</td>
</tr>
<tr>
<td><em>when</em>-t</td>
<td>Structures expressing time with the conjunction <em>when</em></td>
<td>902</td>
</tr>
<tr>
<td><em>whenever</em>-t</td>
<td>Structures expressing time with the conjunction <em>whenever</em></td>
<td>959</td>
</tr>
<tr>
<td><em>baseline</em></td>
<td>Sample from the whole BNCw</td>
<td>872</td>
</tr>
<tr>
<td><em>non-cnd</em></td>
<td>Non-conditional structures</td>
<td>856</td>
</tr>
</tbody>
</table>
ML of whole structures
ML of whole structures: clusters

Most conditionals cluster together … but not all.

Indirect interrogatives immediately cluster together (irrespective of subordinator)

The two structures with *if* are in completely different clusters!

The two structures with *whether* only cluster in step 10

Most conditionals cluster together … but not all.
The ML of bi-partite constructions may not reflect the semantic preference of *if* within the usual short collocation span of 4-5 words.

Examination of ML in its immediate co-text – i.e. the subordinate part.
ML of subordinate parts

[Graph with various subordinate conjunctions plotted on a grid with axes labeled MD and MS.]
ML of subordinate parts: clusters

When we look at the immediate co-text of *if* (sub. part), the ML of *if*-cnd and *if*-q is comparable.

These patterns do not support an explanation in terms of semantic preference.

But this is not the case when we look at the immediate co-text of *whether* (subordinate part).
Comparison of ML in subordinate and matrix parts

Subordinate parts

Matrix parts
ML ratio (subordinate/matrix): clusters

The two structures with *if* are in completely different clusters!

The two structures with *whether* are in completely different clusters!

Structures of the same type do not consistently cluster together!
• Semantic preference cannot, on its own, explain these patterns.

• Type of structure cannot, on its own, explain these patterns.

⇒ Both lexis and grammar seem to be involved.
Construction Grammar

Constructions are ...

• “Conventionalised pairings of form and function” (Goldberg, 2006: 1).

• “Symbolic units” with particular features pertaining to their form and meaning (Croft & Cruse, 2004: 257).
  – Form: morphological, phonological, lexical, syntactic properties
  – Meaning: semantics, (potential) pragmatic uses

CxG sees words (even morphemes) as constructions.

Complex constructions are made up of simple(r) constructions.

(Croft & Cruse, 2004: 258; Fillmore et al., 1988: 501; Fried & Östman, 2004: 18-21)
Accounting for the ML patterns: CxG

The observed ML patterns can be seen to result from the interaction of:

- the function of the construction itself
- the function of the subordinate part
- the function of the matrix part
- the type of link between subordinate and matrix parts
- the meaning of the subordinator

In this light, the semantic preference of a subordinator can be seen as part of the semantic component of a construction.
Conditionals: Two types of syntactic link

Direct: subordinate part is an adjunct

Indirect: subordinate part is a style disjunct

(Quirk et al., 1985: 1071-1072)

Direct
• If we can assemble a package of cash, stock options, and newly issued shares as a good inducement, I think we'll convince the key manager and he'll persuade the others to sell. [FPB 108]

Indirect
• He's not a bad sort for a brother if you know what I mean [AN7 3257]
• If antibiotics are likely to clear up the infection, why are we having this long discussion? [CH1 5292]
DIR and IND: ML of subordinate & matrix parts
The word *if* is not a ‘free agent’; it is part of a very small number of structures.

- On its own: conditionals (DIR, IND), indirect interrogatives.
- As part of a MWU: conditional-concessives (*even if*), comparison structures (*as if*).

In the written BNC, about 85% of *if* tokens are subordinators of conditionals.

⇒ A grammar-independent (bag of words) collocation analysis of *if* would ...

- provide a homogenised picture of its semantic preference.
- essentially mirror the collocations of *if* in (the subordinate parts of) conditionals.
Collocation: Grammatical Constraints (2)

• It would be better, it might even be bearable, if only he knew what had become of James. [A0N 2403]
  – *if only* = conditional + “exclamatory wish” (“intensified equivalent of *if*”) (Quirk et al., 1985: 842, 1092).

• Secondly, the increase in world oil prices has been responsible, if only in part, for the increase in prices of many of the products of Western economies. [K94 2062]
  – *if only* = concessive (elliptical/verbless subordinate clause)  
    → *although / albeit* (Quirk et al., 1985: 1004-1005, 1099).
Case Study 2

Complementation patterns of *(BE) interested*

(Gabrielatos, 2015, 2018)
None of the components of ‘be interested’ can be adequately defined without recourse to grammar.

• The seemingly lexical starting point – the word-form *interested* – must be defined as an adjective, rather than as the past tense of the verb *interest*.

• ‘BE’ must be defined ...
  ... as a copular verb ...
  ... in all its tense-aspect permutations.
Verb Collocates → Semantic Preference

**BE interested in** + **-ing Clause**
- No particular meaning group is more frequent than others.
  - Verbs in the complement seem to be topic-specific.

**BE interested** + **to-inf**
- More than half of verb collocates have meanings relating (directly or indirectly) to knowledge, or actions leading to knowledge (i.e. related to inquiry).
Knowledge-related verbs in complements of *BE interested + to-inf*

**Direct**

- *determine, discover, find out, know, learn, receive* (e.g. information), *share* (e.g. discovery), *study, understand*.

**Indirect**

- *analyse, assess, check, compare, contrast, discuss, examine, experience, explore, hear, identify, interview, listen, look, monitor, notice, observe, read, research, see, speak, study, talk, test, visit, watch, witness.*
### Proportion of knowledge-related verb collocates

<table>
<thead>
<tr>
<th>Expression</th>
<th>BNCw</th>
<th>BNCs</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>BE interested + to-inf</em></td>
<td>53.7%</td>
<td>57.1%</td>
</tr>
<tr>
<td><em>BE interested in + -ing Clause</em></td>
<td>7.6%</td>
<td>14.6%</td>
</tr>
</tbody>
</table>
Proportion of modalised BE in different complementation patterns of *BE interested*

<table>
<thead>
<tr>
<th>Pattern</th>
<th>BNCw %</th>
<th>BNCs %</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>BE interested in</em> + NP</td>
<td>12.8</td>
<td>16.7</td>
</tr>
<tr>
<td><em>BE interested in</em> + -ing Clause</td>
<td>14.9</td>
<td><strong>33.3</strong></td>
</tr>
<tr>
<td><em>BE interested in</em> + <em>wh</em>- Clause</td>
<td>14.3</td>
<td>15.4</td>
</tr>
<tr>
<td><em>BE interested</em> + to-inf</td>
<td><strong>36.4</strong></td>
<td><strong>57.1</strong></td>
</tr>
<tr>
<td><em>BE interested</em> ✓</td>
<td>12.5</td>
<td>21.5</td>
</tr>
</tbody>
</table>
Limitations of collocation as ‘purely lexical’

• A collocation analysis of the word-form *interested* would mainly return collocates of *interested* in its most frequent word class, and in the most frequent syntactic patterns the word is found.

• **Collocation, and the resulting semantic preference, are lexicogrammatical features.**
Conclusions and Suggestions (1)

• The starting point or focus (lexical or grammatical) should not mislead us to conclude that ...  
  ... the starting point is at the core of the patterns we observed.  
  ... the starting point should be treated as being primary.  
  ... any patterns observed can be explained in terms of lexis/grammar only.

• Whatever the focus of the analysis, patterns can only be fully defined if both lexical and grammatical aspects are taken into account.

• Studies may temporarily focus (more) on lexical or grammatical aspects, but these would need to be re-integrated.

⇒ Every instance of language use is lexicogrammatical.
Lexicogrammar and perspectives: an example

– Study 1 examines the frequency that a semantically-defined group of verbs is used in the progressive aspect.

– Study 2 examines the frequency that the progressive aspect is used with particular verbs.

– Despite their different starting points (lexis, grammar), both studies would be essentially examining the same lexicogrammatical item – in a complementary way.
Conclusions and Suggestions (2)

• Halliday’s (1992: 64) “tunnelling” metaphor may not be entirely useful, as it seems to imply both linearity and directionality in research.

• Lexicogrammatical research cannot be mono-directional: at any given point in the analysis, both grammar and lexis are involved.

• Whether the findings are perceived as lexis-like or grammar-like is a matter of perspective or theoretical orientation.

⇒ Description is theoretical.
References (1)

References (2)

References (3)

• Huston, S. (2018) Towards one thousand constructions: Rethinking the learner’s understanding of lexis and grammar. 13th Teaching and Language corpora Conference (TALC2018), University of Cambridge, 18-21 July 2018. [e-copy]