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

Halliday (1966, 1991, 1992)

- Lexis and grammar seen as "complementary perspectives" (1991: 32)

- Lexicogrammatical continuum (1991).

- "[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, 1996, 2004)

- 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).

However …

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)

So …

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

(Gabrielatos, 2006, 2007, 2010, 2011, 2013, forthcoming)
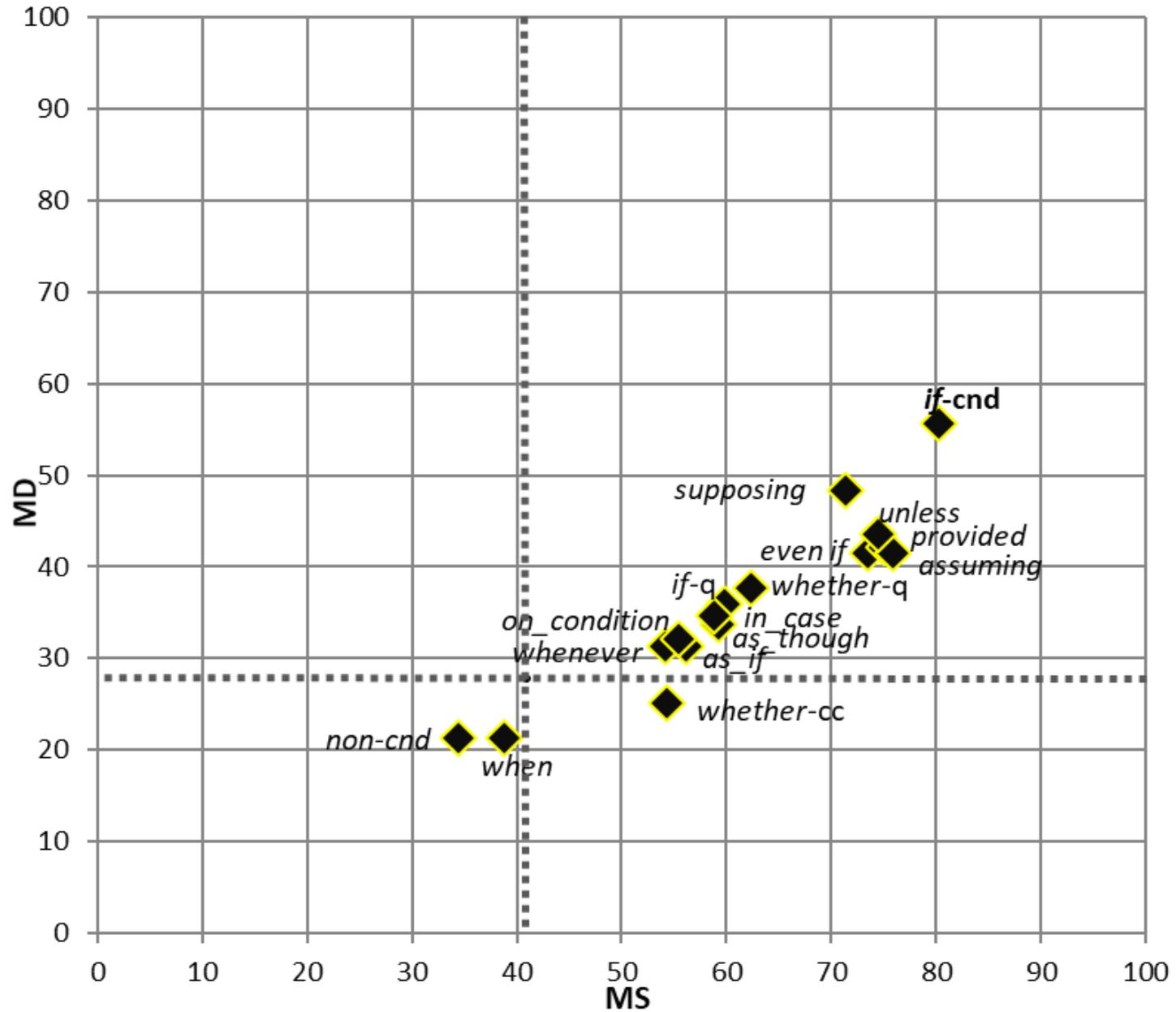
Modal Load

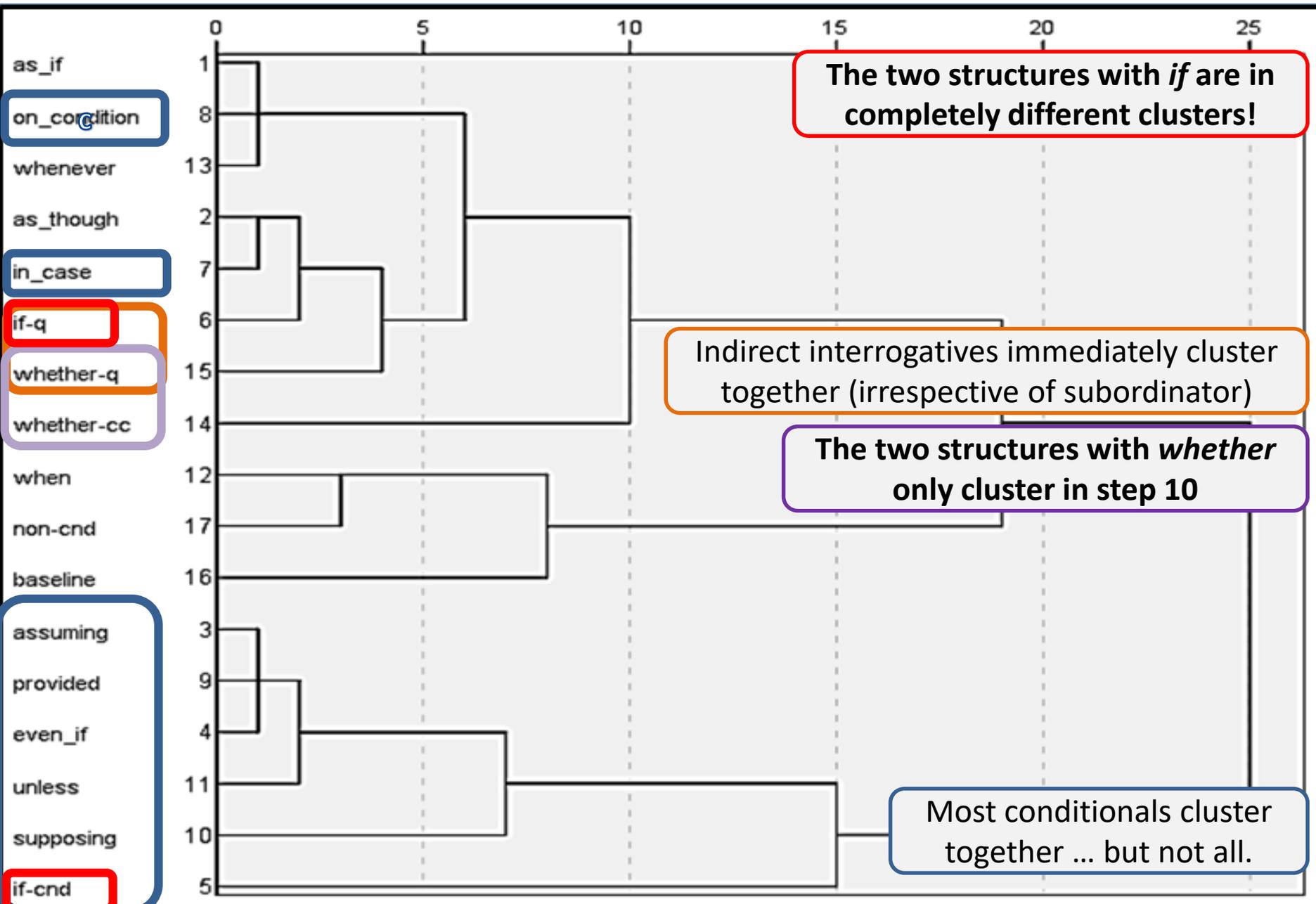|	Modal Density	Modalisation Spread
Definition	Average number of modal markings per clause.	Proportion of constructions that carry at least one modal marking.
Expression	Number of modal markings per 100 clauses.	Proportion (%) of modalised constructions.
Utility	Helps comparisons by normalising for the complexity of the constructions in the sample.	Corrects for heavily modalised constructions in the sample.

(Gabrielatos, 2006, 2010)

Corpus Samples (BNCw)

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

ML of whole structures

ML of whole structures: clusters

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

Indirect interrogatives immediately cluster together (irrespective of subordinator)

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

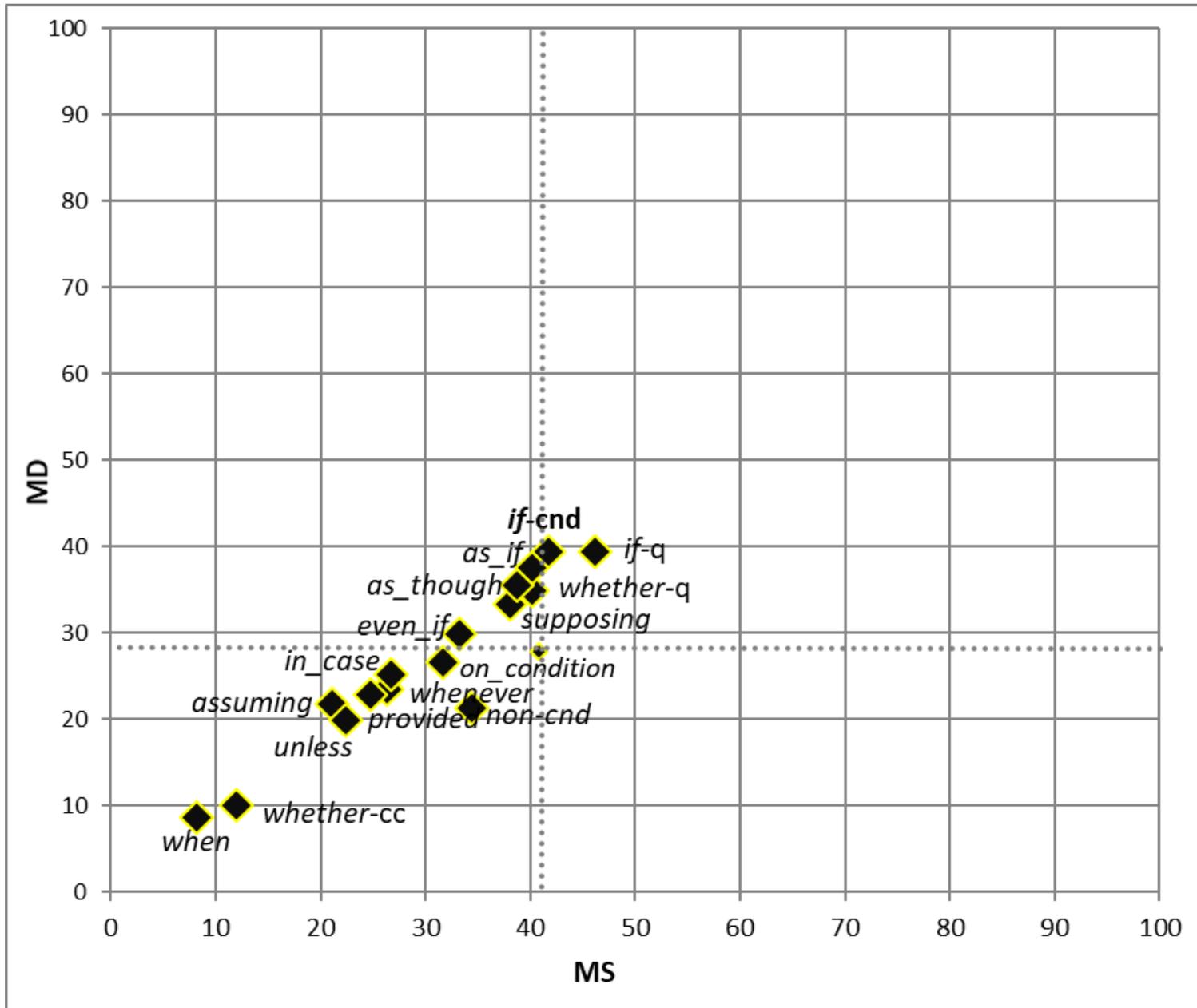Most conditionals cluster together … but not all.

The ML of bi-partite constructions may not reflect the semantic preference of *if* within the usual short collocation span of 4-5 words
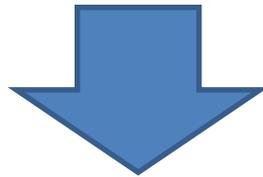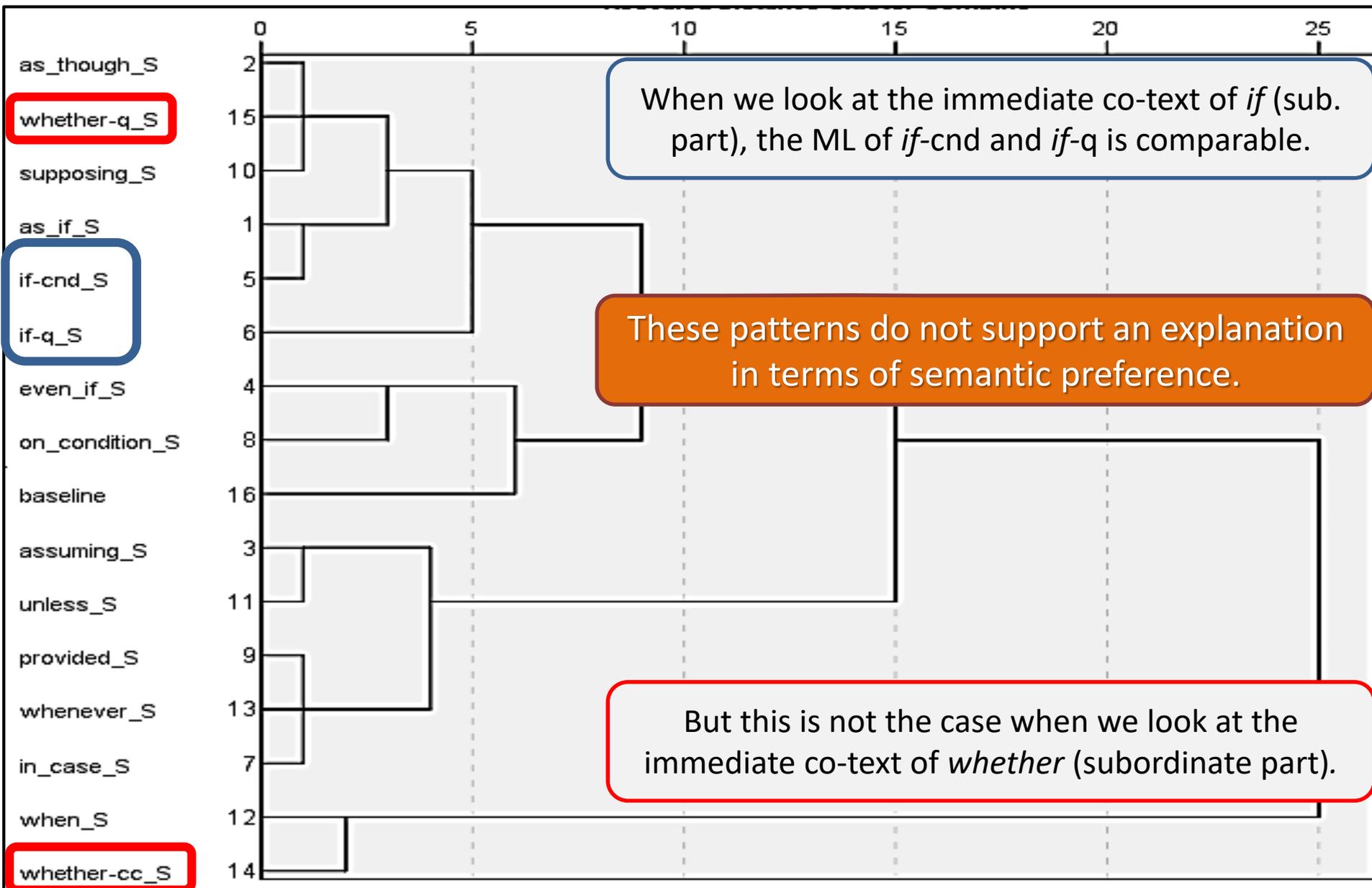
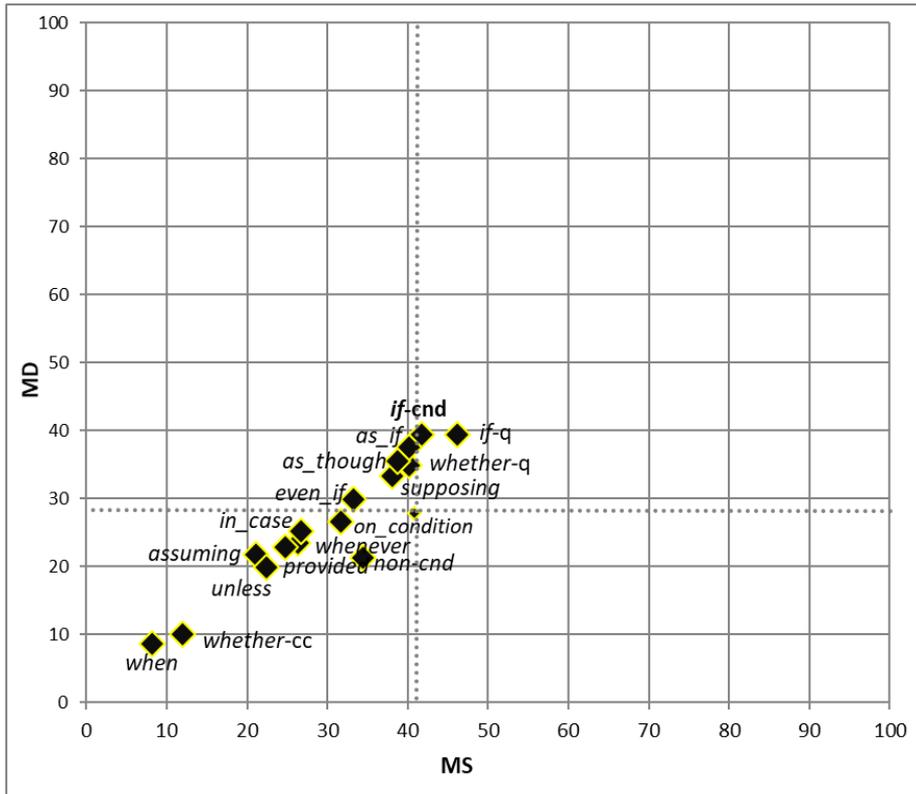Examination of ML in its immediate co-text – i.e. the subordinate part

ML of subordinate parts

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).
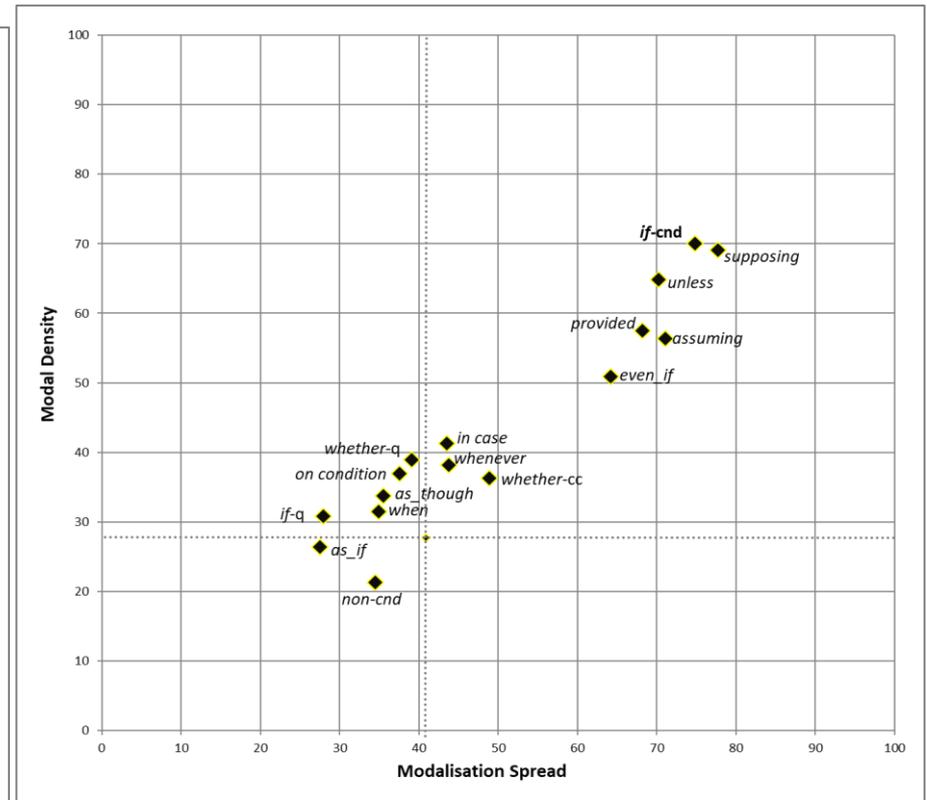
Comparison of ML in subordinate and matrix parts

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!
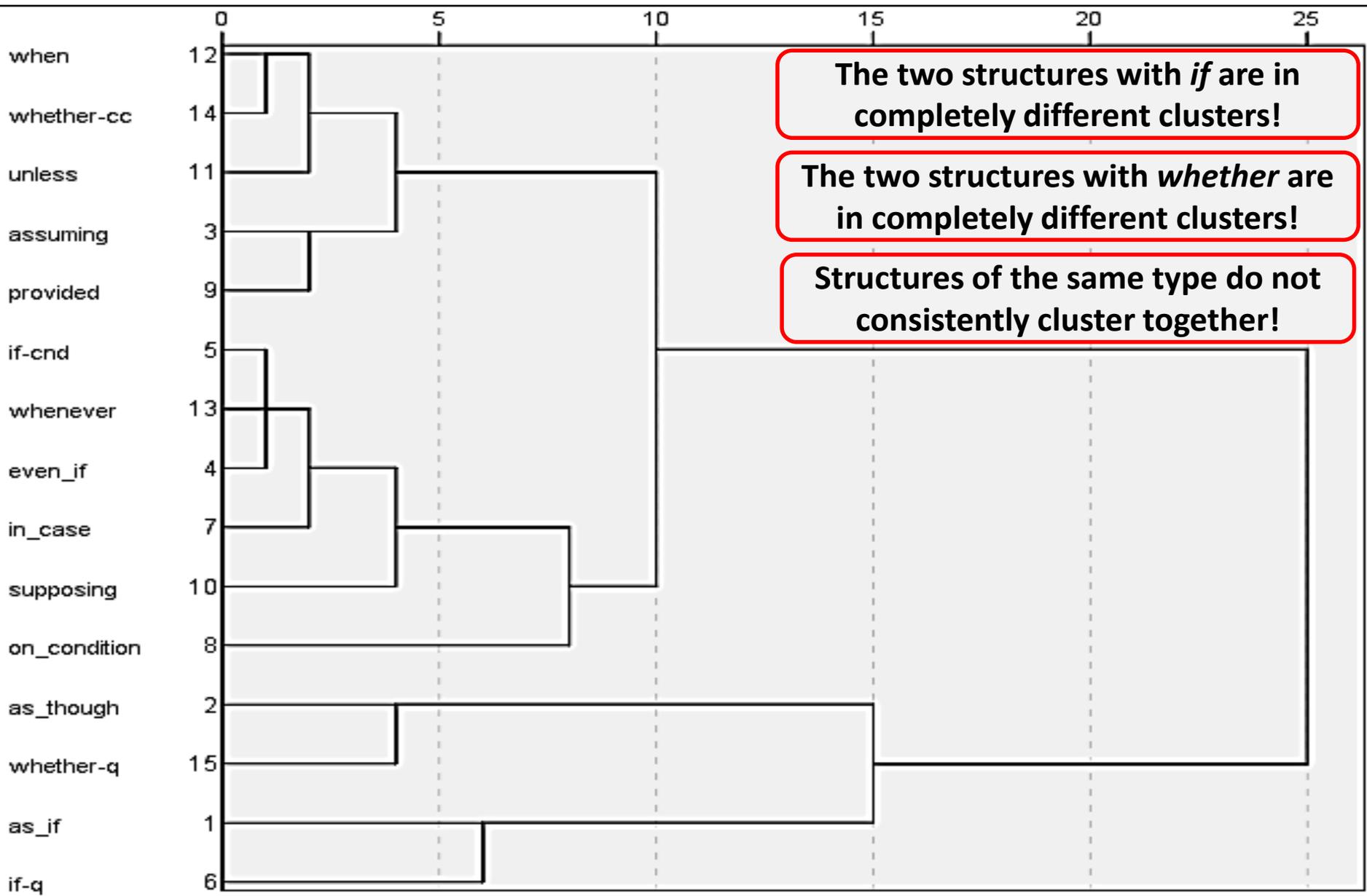
- Semantic preference cannot, on its own, explain these patterns.

- 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
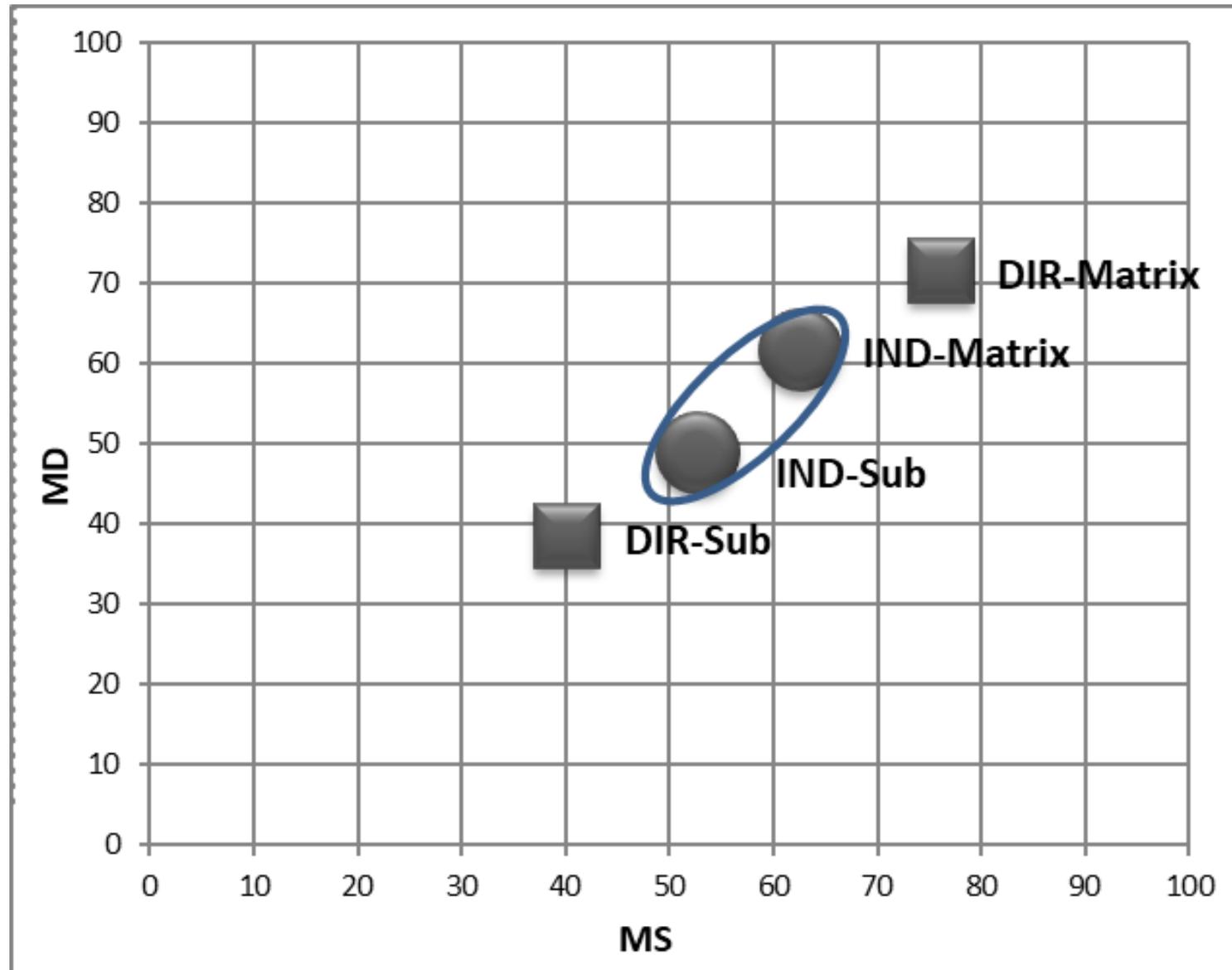
Collocation: Grammatical Constraints (1)

- 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)

(BE) interested

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

	BNCw	BNCs
BE *interested* + *to*-inf	53.7%	57.1%
BE *interested in* + *-ing* Clause	7.6%	14.6%

Proportion of modalised BE
in different complementation patterns
of *BE interested*

	BNCw %	BNCs %
BE interested in + NP	12.8	16.7
BE interested in + -ing Clause	14.9	**33.3**
BE interested in + *wh*- Clause	14.3	15.4
BE interested + *to*-inf	**36.4**	**57.1**
BE interested ⊘	12.5	21.5

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)

- Croft, W. & Cruse, D.A. (2004) *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- Fillmore, C.J., Kay, P. & O'Connor, M.C. (1988) Regularity and idiomaticity in grammatical constructions: the case of *let alone*. *Language*, 64(3), 501-538.
- Firth, J.R. (1957) Modes of meaning. *Papers in Linguistics 1934-1951*. London: Oxford University Press.
- Firth, J.R. (1968) A synopsis of linguistic theory. In Frank R. Palmer (ed.), *Selected Papers of J.R. Firth 1952-59,* 168-205. London: Longmans.
- Fried, M. & Östman, J-O. (2004) Construction Grammar: A thumbnail sketch. In Mirjam Fried & Jan-Ola Östman (eds.), *Construction Grammar in a Cross-language Perspective,* 11-86. Amsterdam: John Benjamins.
- Gabrielatos, C. (2006) *If*-conditionals as modality attractors. *Corpus Linguistics Research Group* (CRG), Lancaster University, 20 March 2006.
- Gabrielatos, C. (2007) If-conditionals as modal colligations: A corpus-based investigation. In M. Davies, P. Rayson, S. Hunston & P. Danielsson (eds.), *Proceedings of the Corpus Linguistics Conference: Corpus Linguistics 2007*. Birmingham: University of Birmingham. [e-copy]
- Gabrielatos, C. (2010) *A corpus-based examination of English if-conditionals through the lens of modality: Nature and types*. PhD Thesis. Lancaster University. [e-copy]

References (2)

- Gabrielatos, C. (2011) Construction Grammar vs. Lexical Grammar: A case study of the modal load in *if*-conditionals. International Symposium On The Sociology Of Words: Lexical Meaning, Combinatorial Potential and Computational Implementation. LACELL, University of Murcia, Spain, 1-2 December 2011. [e-copy]
- Gabrielatos, C. (2013) Using corpus analysis to compare the explanatory power of linguistic theories: A case study of the modal load in *if*-conditionals. *Corpus Linguistics 2013*, 23-26 July 20013, Lancaster University. [e-copy]
- Gabrielatos, C. (2018) The lexicogrammar of *BE interested*: description and pedagogy. In Hoffmann, S., Sand, A., Arndt-Lappe, S. & Dillmann, L.M. (eds.) *Corpora and Lexis.* Language and Computers, Vol. 81. Leiden: Brill. 240-276.
- Halliday, M.A.K. (1966) Lexis as a linguistic level. In Bazell, C.E., Catford, J.C., Halliday, M.A.K. & Robins, R.H. (eds.) *In memory of F. R. Firth*. London: Longman. 148–62.
- Halliday, M.A.K. (1991) Corpus studies and probabilistic grammar. In Aijmer, K. & Altenberg, B. (eds.), *English Corpus Linguistics: Studies in honour of Jan Svartvik*. London: Longman. 30-40.
- Halliday, M.A.K. (1992) Language as system and language as instance: the corpus as a theoretical construct. In Svartvik, J. (ed.), *Directions in corpus linguistics: proceedings of the Nobel Symposium 82, Stockholm, 4–8 August 1991*. Berlin: Mouton de Gruyter. 61–77.

References (3)

- Hoey, M. (1997) From concordance to text structure: New uses for computer corpora. In Patrick James Melia & Barbara Lewandoska (eds.), *PALC '97: Practical Applications in Language Corpora,* 2-23. Łódź: Łódź University Press.
- Hoey, M. (2005) *Lexical Priming: A new theory of words and language*. London: Routledge.
- 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]
- Hunston, S. & Francis, G. 2000. *Pattern Grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins.
- Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. (1985) *A Comprehensive Grammar of the English Language*. London: Longman.
- Sinclair, J. (1991) *Corpus Concordance Collocation*. Oxford: Oxford University Press.
- Sinclair J. (1996) The search for units of meaning. *Textus,* 9(1), 75-106.
- Sinclair, J. (2004) *Trust the Text: Language, corpus and discourse*. London: Routledge.
- Stubbs, M. (2001) *Words and Phrases: Corpus studies of lexical semantics*. Oxford: Blackwell.
- Stubbs, M. (2002) Two quantitative methods of studying phraseology in English. *International Journal of Corpus Linguistics,* 7(2), 215-244.
- Stubbs, M. (2009) The search for units of meaning: Sinclair on empirical Semantics. *Applied Linguistics,* 30(1), 115–137.