

TALC2018

University of Cambridge, 20 July 2018

**Pedagogy-driven
corpus-based
lexicogrammar**

Costas Gabrielatos
Edge Hill University

Motivation

PedMats (especially grammars and coursebooks) often provide partial, inaccurate, or misleading information.

PedMats tend to be informed by previous PedMats and/or RefMats – not CL research.

Each time I examined a language feature with a TALC focus, I realised that existing theoretical constructs or classifications needed to be revised.

- *If*-conditionals and modality (Gabrielatos, 2003, 2006, 2007, 2010, 2011, 2013; Gabrielatos & Kormos, 2014)
- Lexicogrammatical patterns of (BE) *interested* – and modality (Gabrielatos, 2015, 2018)

Focus

- TALC and ...
 - pedagogical materials
 - reference materials
 - theoretical frameworks / constructs
 - primary research on L1
- Lexicogrammar: conceptions
 - Lexicogrammar and Lexical Grammar
 - Compartmentalisation in RefMats and PedMats
- Pedagogical lexicogrammars

TALC: the cast

- L1 corpora (different types)
- L2 corpora (different levels and types)
- RefMats
 - reference grammars (e.g. Quirk et al. 1985, Biber et al. 1999)
 - dictionaries for L1 users
- PedMats

Information: coverage, presence, absence, prominence, accuracy

 - pedagogical grammars
 - dictionaries for L2 users
 - syllabuses: coursebooks / textbooks
 - syllabuses: language tests/exams

TALC research aims and strands

TALC aims to contribute to

- FL/SL/translation learning and teaching (DDL)
- SLA
- Pedagogical materials
- Language testing

TALC research examines

- Learner Language
 - frequency and accuracy (errors)
 - at different levels of proficiency and/or time points
 - In different contexts of use (e.g. classroom interaction, testing)
 - in relation to L1 use, RefMats, PedMats
 - in relation to the learners' L1
- PedMats in comparison to L1 use
- DDL: corpora as teaching/learning materials

Interrelated aspects shared by all strands

- Application of theory
- Reliance on RefMats

Theoretical orientation influences

⇒ Unit of analysis

⇒ Classifications

⇒ Decisions on what constitutes an error.

Who is the arbiter of accuracy?

- PedMats?
- RefMats?
- L1 use? ⇒ CL studies?

Why not just consult RefMats?

- Most do not provide frequency information
 - Those who do are not comprehensive in their coverage (e.g. Biber et al., 1999)
- Lexicogrammatical information not normally / consistently provided
 - lexis and grammar treated in compartmentalised fashion
- Information / classifications may not
 - be informed by corpus-based studies
 - have internal consistency
 - account for all corpus instances

RefMats and informing sources: Types

Corpus-based

- Based on analysis of appropriate representative corpus (sample).
- Adhere to the principle of total accountability.
- Provide quantitative and distributional information.

Data-informed / Corpus-informed

- Use attested/corpus examples.
- Examples selected ad hoc → no claim to representativeness.
- Cannot provide (reliable) quantitative and distributional information.

Introspective

- Examples derived from (and very probably reflecting) the analyst's introspections and/or preferred theory.

(Gabrielatos, 2010: 11-12)

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)

- Posited the *idiom principle* (exemplified by *collocation*), which operates alongside the *open-choice principle* (words fill in particular syntactic positions).
- 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:

- He 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.

Components of the *lexical item* (Sinclair, 1996: 75; Stubbs, 2009: 123-126):

- the core (i.e. a word or phrase)
- its collocates
- its semantic preference
- its semantic prosody
- its colligations

(Grey = optional component)

However ...

Colligation was redefined in a manner consistent with LG tenets.

Original definition

- “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.” (Firth, 1968: 181)

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).

So ...

Main features of Lexical Grammar

⇒ Primacy of lexis

⇒ Lexis and grammar not treated “on an equal basis”

However ...

- 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.

(Gabrielatos, 2018: 244)

Findings from corpus studies

if-conditionals:

classification and modal load

(BE) interested:

lexicogrammatical patterns and modality

Classification of conditionals

- None of the classifications in RefMats accounted for all BNC instances.
- Big Grammars still present classifications characterised by all/most of the following:
 - restricted to degree of likelihood expressed -- ignoring other modal notions: ability, volition, obligation/permission
 - ignoring embedded modalities (modality modifying modality)
 - ignoring the type of syntactic link between protasis and apodosis
 - ignoring complex or atypical conditionals

(Gabrielatos, 2010)

Apodosis marked for different modality types

- If physicists had tried to discover a way to release nuclear energy before 1939, they **would** have worked on anything else rather than the field which finally led to the discovery of fission, namely radiochemistry. [B78 1973]
- If I can live with them, so **can** everyone else.[FS9 2538]
- 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]
- 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]

Multiple modal marking

If you should decide to concentrate on one particular nursing specialty then you **will probably want** to undertake a clinical nursing studies course. [CHT 248]

Companies **potentially need to** acquire information about all the environmental factors shown in the diagram, if they are to survive and prosper. [B15 636]

Different 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

- If antibiotics are likely to clear up the infection, why are we having this long discussion? [CH1 5292]
- He's not a bad sort for a brother if you know what I mean [AN7 3257]

Embedded conditionals

[I]f a producer controls the production of a given commodity he is a monopolist -- if he is such -- not by virtue of any entrepreneurial role, but as a result of a resource monopoly. [HH2 743]

The embedded *if*-clause (*if he is such*) modalises the matrix clause (*he is a monopolist*).

They are people whom we rarely consider in this House, but when there is a suicide or accident on the railway, the driver, and his mate if appropriate, may be mentally scarred for life by the experience. [HHX 119]

Impossible to 'reconstruct' without tampering with the data.

Untypical conditionals: non-contiguous

Abdomen: When to seek advice

Urgently, Right now!

.....

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

[B1R 681]

Title
+ First part of
Apodosis

Second part of
Apodosis

Protasis
(part of list)

– "You intend to reside there ...
wherever ... for some time?"

– "If I like it."

[FPD 1593]

Apodosis

Protasis

Limitations of collocation as ‘purely lexical’

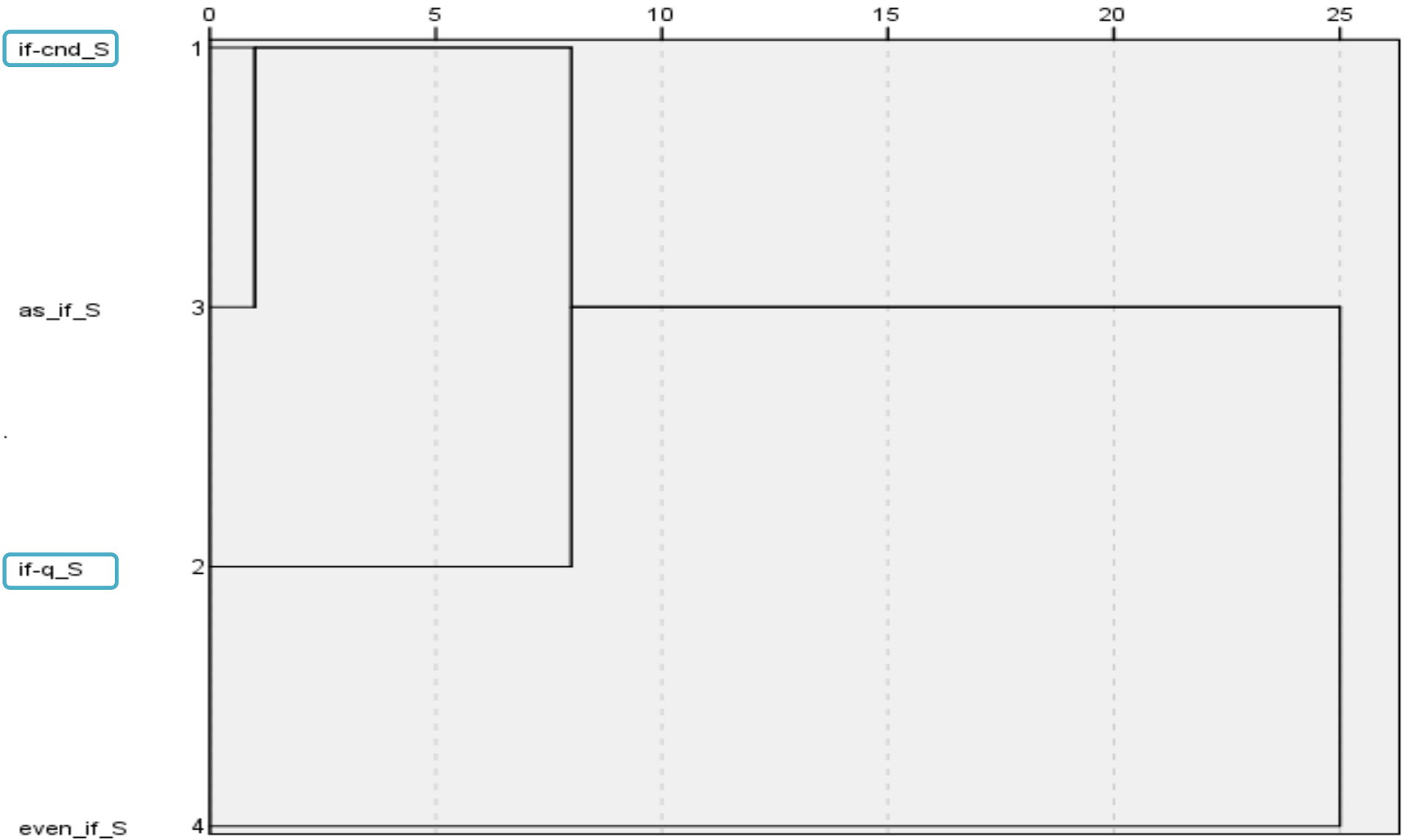
- The word *if* is not a ‘free agent’; it is part of a very small number of structures.
 - On its own: conditional, indirect interrogative.
 - As part of a MWU: conditional-concessive (*even if*), comparison (*as if*).
 - In the written BNC, about 85% of *if* tokens are subordinators of conditionals.
- ⇒ A grammar-independent (bag of words) collocation analysis would essentially mirror the collocations of *if* in conditionals.
- ⇒ Examination of modal load (ML) in subordinate clauses (as a proxy for collocation analysis → semantic preference)

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.

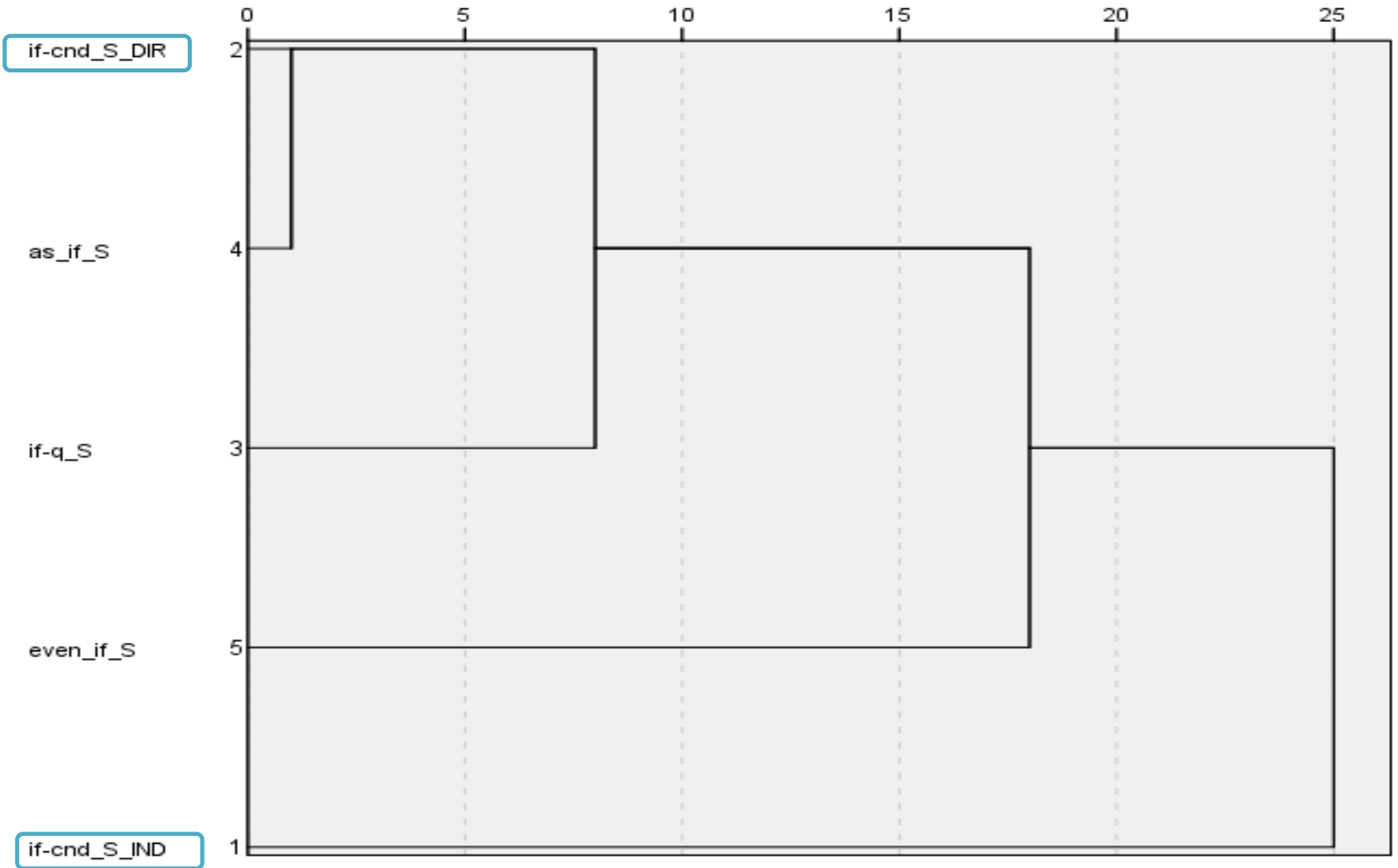
(Gabrielatos, 2010)

Modal Load of subordinate part



What about separating
if-conditionals
into Direct and Indirect ?

Modal Load of subordinate part



If only it was that simple [A0L 3233]

- 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)

(BE) interested

The seemingly lexical starting point (i.e. the word-form *interested*) cannot be adequately defined without recourse to grammar.

- adjective, rather than as the past tense of the verb INTEREST.

BE in *BE interested*

- copular verb ...
- ... in all its tense-aspect permutations

Verb Collocates → Semantic Preference

BE interested in + -ing Clause vs BE interested + to-inf

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

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

BE interested in + -ing Clause vs BE interested + to-inf

	BNCw	BNCs
<i>BE interested + to-inf</i>	53.7%	57.1%
<i>BE interested in + -ing Clause</i>	7.6%	14.6%

Proportion of modalised BE
in different complementation patterns of
BE interested

	BNCw %	BNCs %
in+NP	12.8	16.7
in+ingC	14.9	33.3
in+whC	14.3	15.4
to-inf	36.4	57.1
∅	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 preferences are **lexicogrammatical** features.
- **Every instance of language use is lexicogrammatical**

Compartmentalisation

- RefMats and PedMats: grammars - dictionaries
- However, pedagogical grammars and dictionaries overlap in their coverage
 - grammars also provide lexis-like information
 - dictionaries also provide grammar-like information.
- There is no single source which provides all the lexicogrammatical information that a learner may need in order to form a comprehensive picture of use.
 - Consulting a single pedagogical source cannot be expected to be sufficient, and learners would be wise to combine sources, in particular grammars and dictionaries.

Compartmentalisation

- *interested*: Learner use seems to correlate with the information in pedagogical grammars rather than dictionaries.
 - Lexicogrammatical patterns tend to be presented to learners as grammatical points and/or
 - dictionaries are treated as sources for the meaning of words rather than their use
- Coursebooks have potential for integrated treatment. However ...
- Not only compartmentalised, but also confusing / misleading.

Grammar		Reading	Speaking and Listening
1 People page 8	Present simple and present continuous Adverbs of frequency	Hungry to learn	Voluntary work Meeting a new person Role models People who influence me
	Grammar?		
Use of English	Writing	Vocabulary and Vocab extension	
Adjectives + prepositions so and such	Adding information	Adjectives: personal qualities	Noun suffixes
	Why is this not 'Grammar'?		Morphology
2 Incredible feats page 18	Comparative and superlative adjectives and adverbs Comparative and superlatives with countable and uncountable nouns Double comparatives <i>although/even though, but and however</i>	Extraordinary people	Daredevils Exciting experiences
Parts of speech: -ing words	Informal letter	Descriptions and personal qualities Adjective suffixes Phrasal verbs with <i>look</i>	

UNIT

GRAMMAR

VOCABULARY

EVERYDAY ENGLISH

7 Living history

p54

Morphology

Present Perfect

Unfinished past with *for* and *since**I've lived here for three years.**We've been married since 2010.* p54

Indefinite past

*She's written several books.**I've been to China.* p56

ever and never

Have you ever been in danger? p56

Word endings

Jobs

*philosopher, historian,
economist*

Nouns and verbs

competition/compete

Word stress

'dangerous invi'tation p5

Agree with me!

Questions tags

*It's a lovely day,
isn't it?**You don't like coffee, do
you?*

Adding a comment

*Yes, it is. Beautiful!**No, I don't. I only drink
tea.* p61

8 Girls and boys

p62

Grammar?

have to

*She has to train hard.**I don't have to work late.* p62

should

You should talk to your parents. p64

must

He must get professional help. p64

Things to wear

*belt cap boots**jumper* p68

What things are made of

leather wool cotton p68

At the doctor's

*a sore throat**flu food poisoning**My body aches.**My glands are swollen.**I'll write you a**prescription.* p69

9 Time for a story

p70

Past Perfect

They had walked twenty miles. p71

Narrative tenses

*They saw a bear.**They were looking for work.* p70

Joining sentences – conjunctions

*although, because, so**when, while, before, as, until ...* p72

Feelings

*delighted proud upset**jealous*

Conversations

*Cheer up!**I've got so much to do!**Calm down! You're**so stressed!* p76

Exclamations with

so and such

*I was so scared!**It was such a shock!**I've got so much work!*

p77

Isn't 'vocabulary' part of 'language focus'?

Language focus

Vocabulary

Skills

01
GETTING ON
page 06

Past and present verb forms
Uses of auxiliary verbs

Relationships
Friendship
Wordspot: get

Listening: Your past and present
Reading: Great sibling rivalries
Listening: Characteristics of a good friend

02
UPS AND DOWNS
page 16

Forming adjectives
Forming nouns and gerunds

Describing how you feel
Things that make you feel good

Reading: Happiness: facts and myths
Listening: Things that make you feel good

03
IT ALL WENT WRONG
page 26

Narrative tenses
Continuous aspect in other tenses

Mishaps
Crime and punishment
Headlines

Reading and listening:
News articles

From compartmentalisation to pedagogical lexicogrammars

- Researchers have the knowledge and skills to combine information, ...
- ... but this seems too much to expect from learners.
- Compartmentalisation (with overlaps in coverage) seems practical/inevitable when thinking of hard-copy publications, due to size/cost limitations ...
- ... but online publishing offers possibilities for more comprehensive learner resources that combine the features of pedagogical grammars and dictionaries.

Pedagogical Lexicogrammars

- Learners could access language information starting at any point of the lexicogrammar continuum, ...
- ... and move back and forth combining the information they access.

Example

Learner looking up *interested* to check possibilities/accuracy of use - i.e. starting at the lexis-end of the continuum – will find...

- dictionary-like information -- but also ...
- links to adjective complementation patterns, and complementation in general (the grammar end),
- information on the types of verbs specific to particular patterns (more towards the lexis end)
- issues of modal marking (around the middle).

Pedagogical Lexicogrammars

- Comprehensive
- Combine content of grammars and dictionaries
- Updatable
- Expandable
- Interlinked content
- Links to online corpora
 - ⇒ examples of use in different contexts
 - ⇒ “serendipity” (Bernardini, 2000)
- For examples of what parts of them might look like see Hunston (2018) and Frankenberg-Garcia et al. (2018)

Conclusions and Suggestions

Overall, TALC tends to

- apply theory
- accept RefMats uncritically
- ‘consume’ findings of primary research in L1

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
- should be treated as primary
- any patterns observed can be explained in terms of lexis/grammar only

Conclusions and Suggestions

- 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**

Lexicogrammar and perspectives: an example

A study may examine the frequency that a semantically-defined group of verbs, is used in the progressive aspect, whereas another study may examine the frequency that the progressive aspect is used with particular verbs.

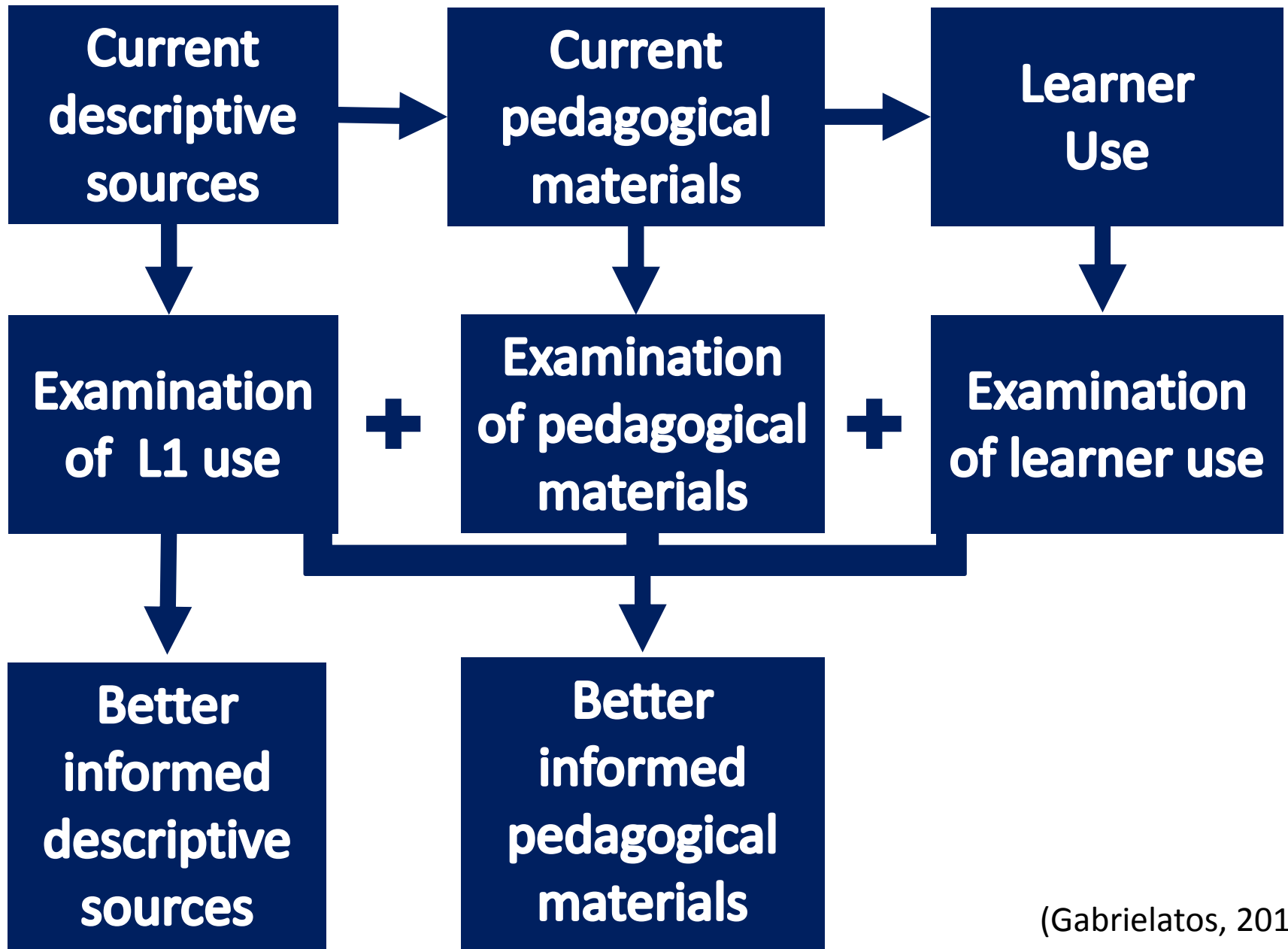
Despite their different starting points (lexis, grammar), both studies would be essentially examining the same lexicogrammatical item – in a complementary way.

Conclusions and Suggestions

- Whatever the focus and starting point of the analysis, all patterns (and their components) 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 both need to be re-integrated.

Conclusions and Suggestions

- TALC can, and should, incorporate primary research in L1
 - ⇒ Critical examination of RefMats
 - ⇒ Improvement and enrichment of RefMats
 - ⇒ Improvement of PedMats
- TALC research can contribute to both theory and pedagogy



(Gabrielatos, 2018)

The 'parting shot'

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

(Gabrielatos, 2011)

References

- Bernardini, S. (2000) Systematising serendipity: Proposals for concordancing large corpora with language learners. In Burnard, L. & McEnery, T. (eds), *Rethinking Language Pedagogy from a Corpus Perspective: Papers from the third international conference on teaching and language corpora*. (Lodz Studies in Language). Hamburg: Peter Lang. 183-190.
- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan E. (1999) *Longman Grammar of Spoken and Written English*. Harlow, Essex: Longman.
- Frankenberg-Garcia, A. Rees, G.P., Lew, R. & Roberts, J. (2018) Developing a tool to help EAP writers with collocations in real time. 13th Teaching and Language corpora Conference (TALC2018), University of Cambridge, 18-21 July 2018. [[e-copy](#)]
- Gabrielatos, C. (2003). Conditional sentences: ELT typology and corpus evidence. *36th Annual BAAL Meeting*, University of Leeds, UK, 4-6 September 2003. [[e-copy](#)]
- Gabrielatos, C. (2006) Corpus-based evaluation of pedagogical materials: *If*-conditionals in ELT coursebooks and the BNC. *7th Teaching and Language Corpora Conference*, University Paris 7 Denis Diderot, Paris, France, 1-4 July 2006. [[e-copy](#)]
- 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](#)]

References

- 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](#)]
- 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. (2013) *If*-conditionals in ICLE and the BNC: A success story for teaching or learning? In Granger, S., Gilquin, G. & Meunier, F. (eds.) *Twenty Years of Learner Corpus Research: Looking back, moving ahead*. Louvain-la-Neuve, Belgium: Presses Universitaires de Louvain. 155-166.
- Gabrielatos, C. & Kormos, J. (2014). Corpus-based examination of *if*-conditionals in an oral proficiency exam. 7th Biennial IVACS Conference, Newcastle University, 19-21 June 2014. [[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.

References

- 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.
- 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](#)]
- 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. (2009) The search for units of meaning: Sinclair on empirical Semantics. *Applied Linguistics*, 30(1), 115–137.