Corpus Linguistics in the South 9

Oxford Brooks University, 18 April 2015

Corpus Linguistics and CDA Revisiting the notion of *synergy*

Costas Gabrielatos & Alison Duguid

(Edge Hill) (Siena)

This is a slightly revised and abridged version of

Gabrielatos & Duguid (2014):

https://www.academia.edu/8358693

Revisiting *synergy* (1)

The notion of *synergy* between CL and CDA seems fairly straightforward – but it isn't.

Simply seeing the synergy as using CL techniques to do CDA underestimates the complexity of the interaction.

Some indications of the complexity can be gleaned from examining core aspects of their nature … ⇨⇨⇨

Revisiting *synergy* (2)

- Both are relatively new (late 70s / early 80s)

  ⇨ Room for development of concepts, constructs, techniques.

- Both are pluralistic

  ⇨ Theoretical and methodological tensions.

- Both can be said to have ‘identity issues’

  ⇨⇨⇨⇨⇨

Revisiting *synergy* (3): Identity

How would you describe your research identity?

- *I'm a corpus linguist doing (C)DA.*
- *I'm a (critical) discourse analyst using CL.*
- *Other:* _____

Neither CDA nor CL are a theory, but …

Both have particular, and strong, theoretical affiliations:

CDA

- Focus, hypotheses and analysis are informed by particular socio-political and (to a lesser extent) linguistic theories (mainly Functional Grammar)

CL (as used in/for CDA)

- Approach: exploratory, rather than hypothesis-driven. However, …
  - …the choice of topic is indicative of covert hypotheses/expectations.
  - …the analysis predominantly depends on a particular linguistic theory (Lexical Grammar) – although this is rarely acknowledged (McEnery & Gabrielatos, 2006).

Are they methodologies?

CDA is not a methodology …

- … but it is traditionally oriented towards ‘qualitative’ approaches

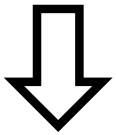
CL is mainly a methodology, with a strong ‘quantitative’ element …

- … but it *does* employ qualitative analyses (annotation) before quantitative analysis -- though this is mostly unacknowledged.

⇨ CL blurs the distinction between ‘qualitative’ and ‘quantitative’ analysis.

Misconceptions about CL
(shared by CDA and CL researchers)

- CL is only useful in identifying rough, large-scale patterns – lacking in fine-grained analysis.
- CL does not examine sufficient amount of text – at best, it examines short concordance lines.
- CL does not take account of the relevant context.
- CL is a ‘quantitative; approach.



- The sole contribution of CL in CDA is in the quantification of patterns.
- CL only provides an ‘entry point’ for CDA.

Currently, criticisms of CL focus on …

What (some) corpus linguistic studies *tend to do*

A more constructive approach to the use of CL for CDA

~~What (some) corpus linguistic studies *tend to do*~~

What corpus linguistics *can do*

(Gabrielatos, 2009)

Using relevant context

- CDA researchers don't have privileged contextual information; CL researchers have no less access to sources of relevant contextual aspects (Gabrielatos, 2009).

- The examination of numerous (expanded) concordance lines enables the researcher to infer useful contextual information (Baker et al., 2008: 279, after Brown & Yule, 1982: 47, 59).

Qualitative / Quantitative

- Many CA studies use "various kinds of quantifying expressions, such as 'routinely', 'regularly', 'frequently'." (ten Have, 2007: 157)
- "Analysis of aggregates requires a preliminary analysis of single instances." (ten Have, 2007: 157)
- "In corpus analysis, the quantitative approach forces the linguist to make constant decisions to achieve a stringent categorisation. […] Thus the qualitative approach, which establishes the basic concepts […], must precede any quantitative investigation." (Schmied, 1993: 85)
- "The terms 'quantitative' and 'qualitative' may be more helpfully regarded as notional methodological extremes." (Baker et al., 2008: 276)

CL can…

- Provide techniques for building topic-specific corpora (e.g. Gabrielatos, 2007)
- Reveal salient contextual elements ("trigger events" – Gabrielatos et al., 2012)
- Reveal differences as well as similarities (e.g. Taylor, 2013)
  ⇨ intertextuality / interdiscursivity
- Pinpoint absence (e.g. Partington, 2014)

CL can…

- Pinpoint systematic patterns of presentation
  - within a text (e.g. speech)
  - across a large number of texts → intertextuality
  - across time

- Provide evidence of representation strategies already identified in CDA, …
- … but also uncover other (aspects of) such strategies,
  ⇨ and contribute to the theoretical frameworks informing CDA

CL can…

Help with downsampling

- More ‘traditional’ CDA techniques can be applied to the analysis of a small number of texts which have been objectively selected.
- **RARELY USED!!**

How?

- Pinpointing periods of increased reporting on a topic
  - Examining the diachronic development of article frequency (Baker et al., 2008; Gabrielatos et al. , 2012)
- Selecting texts using a combination of collocation analysis and manual semantic annotation:
  - Texts containing the highest frequency of strong collocates which, additionally, express a range of concepts (Forchtner & Kølvraa, 2012).

A critical view on *objectivity* in CL

CL studies aim to discover/uncover patterns.

The vast majority of corpus-based/assisted CDA studies use:

- Keyness analysis: frequency comparisons
- Collocation analysis
- Concordance analysis

These techniques involve potential pitfalls that need to be recognised and avoided.

Patterns: Are they really there?

- Apophenia: The tendency to perceive a connection or meaningful pattern between unrelated or random things (such as objects or ideas) (Merriam-Webster online dictionary).

- Technological innovations […] permit us to plough through vast quantities of text in a short time and to reduce it or 'boil it down' to lists and concordance lines […] the pattern-perceiving predisposition of the brain comes into play when it examines such lists (Scott & Tribble, 2006: 5).

Patterns: Confirmation bias

- *"It is the peculiar and perpetual error of the human understanding to be more moved and excited by affirmatives than by negatives."* (Francis Bacon)
- Confirmation bias: tendency to notice what confirms one's beliefs, theory, hypothesis, expectations -- and to ignore, not look for, or undervalue the relevance of what contradicts them.

Patterns: Influence of researcher/theory

- When the approach is exploratory, different researchers may see different (though overlapping) patterns/categories in the same corpus (Marchi & Taylor, 2009).

- When the approach is theoretical, recognised patterns may be restricted to categories predicted by the theory, or even made to fit those categories (*Procrustean* approach).

Frequency

Frequency per N words may not always be appropriate. It's best to calculate on the basis of the number of opportunities.

(Ball, 1994)

Words vs. Opportunities

		Passives	Words	*P/W*
(1)	Three demonstrators were killed and 400 people were injured after thousands of protestors swarmed Pakistani Prime Minister Nawaz Sharif's residence calling for his resignation.	2	24	*8.3%*
(2)	At least three people were killed and 200 were wounded overnight as clashes erupted between authorities and opponents of Prime Minister Nawaz Sharif.	2	23	*8.7%*

Words vs. Opportunities

		Passives	Words	*P/W*	Clauses	*P/C*
(1)	Three demonstrators were killed and 400 people were injured after thousands of protestors swarmed Pakistani Prime Minister Nawaz Sharif's residence calling for his resignation.	2	24	*8.3%*	**4**	***50%***
(2)	At least three people were killed and 200 were wounded overnight as clashes erupted between authorities and opponents of Prime Minister Nawaz Sharif.	2	23	*8.7%*	**3**	***67%***

Collocation

Collocations can be calculated using different metrics:

⇨ The same corpus will return different sets of collocates.

When using effect-size metrics for collocates (usually MI), statistical significance is not taken into account.

⇨ Rare collocates are included -- frequent collocates may be excluded

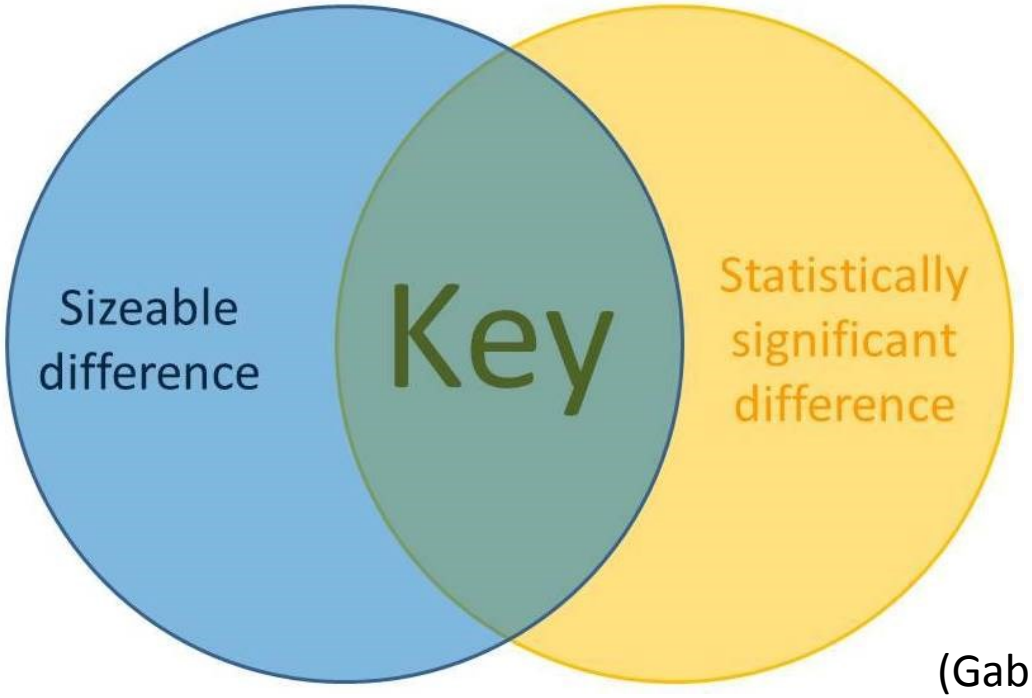
⇨ Results are skewed.

(Gabrielatos & Baker, 2008)

Keyness

Predominantly calculated using statistical significance metrics (usually log likelihood) rather than effect-size metrics.

⇨ Results are unreliable



(Gabrielatos, 2014)

Cherry picking in CL

- Excluding function keywords/collocates
  - They have been shown to index discourses (e.g. McEnery, 2006; Duguid, 2008).
- Only examining top N keywords/collocates
  - Inclusion of less frequent/strong types may, collectively, reveal a different picture (e.g. Baker, 2004).
- Only examining 'interesting' or 'relevant' keywords/collocations.
- Ignoring patterns not recognised by hypothesis/theory /algorithm.

A warning

Corpus linguistics is very easy to do badly.

(Gabrielatos, 2013)

Towards Synergy

- Moving away from CL/CDA caricatures.
- Realisation that CDA has to do with research goals.
  ⇨ Methodology and informing theory can vary.
- Discarding misleading polar distinctions:
  – Objective vs. subjective
  – Qualitative vs. quantitative
- The synergy within:
  – CLs becoming better aware of other CDA approaches.
  – CDAs becoming better aware of CL approaches.
- Closer collaboration ⇨ Dialogue

References (1)

- Baker, P. (2004). Querying keywords: Questions of difference, frequency and sense in keywords analysis. *Journal of English Linguistics,* 32(4), 346-359.
- Baker, P., Gabrielatos C., KhosraviNik, M., Krzyzanowski, M., McEnery, T. & Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society,* 19(3), 273-305. [e-copy]
- Ball, C.N. (1994). Automated text analysis: Cautionary tales. *Literary and Linguistic Computing*, 9(4), 265-302.
- Duguid, A. (2007). Soundbiters bit: Contracted dialogistic space and the No. 10 team. In Fairclough, N., Cortese, G. and Ardizzone, P. (eds.) *Discourse and Contemporary Social Change* (73-94). Bern: Peter Lang.
- Duguid, A. (2008). Men at work: How those at Number 10 construct their working identity. In Garzone,G & Sarangi, S. (eds.) *Discourse, Ideology and Specialized Communication* (453-484). Bern: Peter Lang.

References (2)

- Forchtner, B. & Kølvraa, C. (2012). Narrating a 'new Europe': From 'bitter past' to self-righteousness? *Discourse & Society,* 23(4), 377-400 .
- Gabrielatos, C. (2007). Selecting query terms to build a specialised corpus from a restricted-access database. *ICAME Journal,* 31, 5-43. [e-copy]
- Gabrielatos, C. (2009). Corpus-based methodology and critical discourse studies: Context, content, computation. Siena *English Language and Linguistics Seminars*, University of Siena, 9 November 2009. [e-copy]
- Gabrielatos, C. [@congabonga]. (2013, September 13). The allure of #corpuslinguistics: it's quite easy to do badly [Tweet]. Retrieved from https://twitter.com/congabonga/status/377331048448671744
- Gabrielatos, C. & Baker, P. (2008). Fleeing, sneaking, flooding: A corpus analysis of discursive constructions of refugees and asylum seekers in the UK Press 1996-2005. *Journal of English Linguistics,* 36(1), 5-38.
- Gabrielatos, C., McEnery, T., Diggle, P. & Baker, P. (2012). The peaks and troughs of corpus-based contextual analysis. *International Journal of Corpus Linguistics*, 37(2), 151-175.

References (3)

- Marchi, A. & Taylor, C. (2009). If on a winter's night two researchers …: A challenge to assumptions of soundness of interpretation. *CADAAD Journal* 3(1), 1-20. [e-copy]
- McEnery, A.M. (2006). *Swearing in English: Bad language, purity and power from 1586 to the present.* London: Routledge.
- McEnery, T. & Gabrielatos, C. (2006). English corpus linguistics. In Aarts, B. & McMahon, A. (eds.) *The Handbook of English Linguistics* (33-71). Oxford: Blackwell.
- Partington, A. (2014). Mind the gaps. *International Journal of Corpus Linguistics,* 19(1), 118–146.
- Partington, A., Duguid, A. & Taylor, C. (2013). *Patterns and Meanings in Discourse: Theory and practice in corpus-assisted discourse studies (CADS).* John Benjamins.
- Scott, M. & Tribble, C. (2006). *Textual Patterns: Keyword and corpus analysis in language education*. Amsterdam: Benjamins.
- Taylor, C. (2013). Searching for similarity using corpus-assisted discourse studies. *Corpora*, 8(1), 81-113.

More on counting and keyness

- Gabrielatos, C. (2013). Counting: Word frequency and beyond. *Dubrovnik Fall School in Linguistic Methods*. Centre for Advanced Academic Studies, Dubrovnik, Croatia, 20–26 October 2013. [e-copy]
- Gabrielatos, C. (2013). Keyword analysis. *Dubrovnik Fall School in Linguistic Methods*. Centre for Advanced Academic Studies, Dubrovnik, Croatia, 20–26 October 2013. [e-copy]
- Gabrielatos, C. & Marchi, A. (2011). Keyness: Matching metrics to definitions. *Corpus Linguistics in the South: Theoretical-methodological challenges in corpus approaches to discourse studies - and some ways of addressing them*. University of Portsmouth, 5 November 2011. [e-copy]
- Gabrielatos, C. & Marchi, A. (2012). Keyness: Appropriate metrics and practical issues. *CADS International Conference*, Bologna, Italy, 13-15 September 2012. [e-copy]

Bibliography of discourse studies using corpus techniques

www.gabrielatos.com/CLDA-Biblio.htm