Lexicogrammatical patterns as the building blocks of text:

Identifying and interpreting patterns in Financial Press Releases

Sophie-Hélène Cimon
Christopher Gledhill

Université Paris Diderot

Symposium on Corpus Approaches to Lexicogrammar
10 June 2017
Edgehill University

Introduction:
Doctoral research project

Financial Press Releases (Business English *BUSE* vs Business English as Lingua Franca *BELF*)

Motivation:

Interest in **automated polarity detection** in financial text, by providing a practical and theoretical approach based on Appraisal Theory (Martin and White) and SFL.

Interest in identifying the **Pollyanna effect** in BuseE/BELF (NB in psychology 'Sentiment Analysis')

Genre: Financial Press Release (FPR)

« *News that is sent out or released by the company making the news*. If it's an earnings press release, the release will discuss the company's financial results for the recently completed quarter and may provide comments from management. *Press releases often list valuable contact information that can assist investors in their research…* »

« *Remember that press releases are written by the companies, not by a reporter, so they can be full of company bias. … »*

(Press Release,www.investopedia.com)

Genre: Financial Press Release (FPR)

- Generally one or two pages long, available online
- Link on a generic/discursive chain
    - Regulatory filing
    - ➔FPR
        - ➔Webconference
            - ➔ Transcript
        - ➔Financial press
            - ➔Online investor discussions

Linguistic features of FPRs

- 2 main 'embedded' discourse types:

(1) Accounting Narrative:

Textual representation of numerical information, regulated discourse

(2) Management Narrative:

Evaluation of financial performance, usually mediated by quotes from a company executive

Accounting narrative

"[COMPANY] today reported earnings per diluted share of $1.09 for the second quarter ended August 1, 2015. This included an increase in earnings per diluted share of $0.16, which primarily represented the reclassification of receivables as "held for sale" associated with the pending credit card transaction.

For Fiscal 2012, income before income taxes is anticipated to be between $494 million and $501 million, up 64% to 67% (year ended January 29, 2011 ("Fiscal 2011"): $300.4 million). Diluted earnings per share are expected to be between $3.67 and $3.72, an increase of 58% to 60% (Fiscal 2011: $2.32).

Gross margin continued to reflect the impact of higher product costs, declining to 45.2% from 46.6% in the 2010 period. Operating income was $358 million on an adjusted basis in the XTER. This included earnings from the Timberland acquisition of $50 million, excluding acquisition-related expenses of $6.7 million. On a GAAP basis, XTER operating income was $351 million."

Management narrative

"We had a great first quarter driven by our unrelenting commitment to delivering innovative products and services to athletes around the world," said Mark Parker , President and CEO of NIKE, Inc. "Our powerful portfolio of businesses combined with unmatched leadership and resources allows us to capitalize on opportunities that drive long-term value for our shareholders. I am more excited than ever about our potential to continue to innovate with purpose, and fuel NIKE's growth."

"We are pleased with the third quarter result, which confirms that the business has re-established a base platform to build from. Third quarter was also the last quarter with the easy comparative figures created by last years' stock balancing programme. This obviously means that we are now entering a phase with more modest headline growth, but still underpinned by strong operating performance."

Initial observations

Accounting narrative impersonalizes financial results (typically realized by Relational processes, often expressing evaluative discursive functions (attribution, comparison, contextualization), and with the general aim of framing specific information on financial amounts, time scales, and directions of financial travel (*increase / decrease*)).

Management narrative personalizes financial results (typically expressed by Material processes, or modulated by e.g. positive engagement (Behavioural, Mental processes) or metaphors of 'conation' (*capitalize / drive / perform*))

Next step: look for n-grams, tag-grams and lexicogrammatical patterns (LxP)

(a) *[financial concept]* **totaled** *[sum] in [period]* **compared to** *[sum] in [period]*.

(b) *The [sum] [increase/decrease] in [financial concept]* **was [driven/offset]** *by an [increase/decrease] in [financial concept] of [%]*.

(c) *During [period], [financial concept]* **increased [%], or [sum], as compared** *to [period]*.

(d) **As a percentage of** *[financial concept], [financial concept]* **increased** *[# points]* **to** *[%] from [%]*

Starting point for LxPs: discontinuous sequences of grammatical items

1) Strip a (sample) sequence of words of its lexical items, eg

a) *< * * * is a * * of * *s and some *s * that *ly * of *s are *ed. >*

b) *<*s * a * *er to *ing an * * * and * * of * of the most * and * *s.>*

2) Look for the longest sequence possible in the corpus (or Google)

3) Given two sequences, such as a) and b) above, it should be possible to predict whether they belong more or less to a typical discourse (eg a research article (RA) or a journalistic article (JA)), and within these sequences it should be possible to identify those patterns that are typical of the genre and those which are 'merely' local innovations

Re-definition of Lexicogrammatical Patterns

1. **Statistical-Textual unit**: a cascade of co-selected lexical items that may be extended / discontinuous

2. **Syntactic-Semantic unit**: a complex sign (the pattern has a particular frame of reference, not always syntactic constituent, can be composed of lexical and / or grammatical items); the pattern includes at least one predictable item (pivot) and a productive structure (paradigm) allowing for reformulation and integration into longer stretches of text

3. **Discourse-Rhetorical unit**: a rhetorical move, with a distinctive discourse function (tied to specific discourse or LSP)

A Prototypical LxP
(<failure to...> 'warning in technical manuals')

*(1a) **Failure to** comply with safety precautions described in the manuals ... **may result in** death, serious injury or property damage.*

*(1b) WARNING! Read all instructions. **Failure to** follow all instructions listed on pages 4 to 7 **may result in** fire, serious injury and/or DEATH.*

*(1c) **Failure to** heed these instructions **may result in** personal injury and/or property damage.*

*(1d) WARNING RISK OF BURNS. **Failure to** observe these instructions **could lead** to serious injury or death.*

*(1e) **Failure to** understand the proper operation of this product **can result in** serious injury and/or property damage.*

<failure to do (this, so)...>
Sub-pattern (or co-pattern?)

(2a) **Failure to do this could result in** *the cable parting from the drum under load.*

(2b) *CAUTION: ...* **Failure to do so will result in** *the outer wraps pressing against the inner wraps resulting in the damage of the cable.*

(2c) **Failure to do so could result in** *pressure buildup which can cause the gearbox to leak or damage the equipment.*

(2d) *All electrical work must be performed by a licensed electrician.* **Failure to do so could result in** *electric shock or poor winch operation.*

(2e) *Remove port caps from exhaust ports for operation.* **Failure to do so may result in** *equipment damage and limit performance.*

Research questions about lexicogrammatical patterns

1. Hypothesis: it's possible to find at least one really good (=prototypical) LxP in any LSP corpus (well-defined corpus representing a specific genre), therefore:

2. What is the **distribution** of LxPs? What is the proportion of LxPs in a given LSP corpus? 100%? 10%?

3. Is there a **hierarchy** or **ontology** of LxPs?

4. What's the relationship between LxPs and similar patterns in the **general language** (LGP)? (NB I assume that there is no such thing as LGP)

Back to Financial Press Releases

- What LxPs are found in Financial Press Releases (FPR)?

- Are LxPs relevant to the study of evaluation in
    - Specialized genres?
    - Second-language writing in specialized genres?

Two corpora: BusE and BELF

- **BUSE (Business English)**: Companies operating or publicly listed in countries where **English is an official language** (United States, Canada, United Kingdom, Australia)

- **BELF (Business English as Lingua Franca)**: Companies operating or publicly listed in countries where **English is not an official language** (France, Italy, Brazil, Switzerland, Austria, Denmark)

Method – Corpus

- Two comparable corpora of *Financial Press Releases* published between 2012 and 2016
- Contrasting lexicometrics (the 'bag-of -words' approach) with textometrics (intensive comparative analysis)

	BusE	BELF
Tokens	756,557	728,619
Words	575,136	557,360
Zip files (companies)	25	22

Starting points for identifying candidate LxPs using SketchEngine

1. n-grams

    A. Generic salient units (BusE and BELF)

    B. Specific salient units (BusE or BELF)

2. Tag-grams

    A. Generic salient units (BusE and BELF)

    B. Specific salient units (BusE or BELF)

RB+JJ in the BusE corpus

Salient compared to CoCA (fr.: 1.560), Relatively neutral polarity

(NB *Sainsbury's has posted **flat like-for-like** sales during the golden quarter as its non-food offer delivered a "strong" performance.*)

primarily due (488)	broadly unchanged (64)	very strong (40)
primarily attributable (62)	relatively flat (33)	newly transitioned (23)
principally due (44)	relatively unchanged (25)	highly promotional (20)
largely due (24)	approximately equal (9)	particularly strong (17)
principally attributable (15)	relatively consistent (7)	mandatorily redeemable (16)
partially due (11)	roughly flat (5)	up low
also due (13)	largely unchanged (5)	exceptionally strong (14)
mainly due (7)	essentially flat (5)	unseasonably warm (12)
mainly attributable (7)	relatively stable (4)	flat like-for-like (12)
partly due (6)	broadly consistent (4)	…
also attributable (5)	broadly flat (3)	
primarily owing (2)	approximately flat (3)	
largely attributable (2)	relatively constant (2)	
directly attributable (2)		

688 (44%)   169 : 12%

[directional result] + "primarily due (to)" + [commercial activity] (BusE)

(3a) *The [$] <u>increase</u> in net sales is **<u>primarily due to</u>** a [%] <u>increase</u> in comparable store sales […]*

(3b) *This <u>decrease</u> was **<u>primarily due to</u>** <u>reduced</u> earnings.*

(3c) *Average merchandise transaction value <u>declined</u> in Jared, driven primarily by sales mix **<u>primarily due to</u>** <u>higher</u> bead sales.*

(3d) *The <u>improvement</u> in wholesale operating margin was **<u>primarily due to</u>** <u>higher</u> gross margin […]*

[evaluated result] + "partially offset" + [commercial activity] (BusE)

(4a) *These impacts were **partially offset** by higher representative and sales leader investment […].*

(4b) *The decrease in inventories was **partially offset** by an increase in kate spade inventory to support growth initiatives, including retail expansion.*

(4c) *Within the region, strong growth in France was **partially offset** by a decline in distributor markets.*

(4d) *This was **partially offset** by favorable channel and product mix and improved margins in global full price retail stores.*

RB+JJ in the BELF corpus

- Salient compared to CoCA and more frequent (2,053) than BusE. Greater variety in positively evaluated combinations (virtually unchanged: 16, broadly stable: 12)

mainly due (286)
primarily due (150)
also due (37)
mainly attributable (34)
largely attributable (8)
entirely attributable (8)
primarily attributable (7)
entirely due (6)
partly due (4)

537: 26%

very good (76)
very strong (55)
very positive (55)
particularly strong (40)
very high (27)
very solid (23)
very interesting (14)
very important (14)
fully meaningful (14)
highly positive (9)
very successful (10)
particularly dynamic (9)
highly dynamic …

slightly negative (20)
severely unfavourable (12)
very challenging (9)

[directional result] + "primarily due (to)" + [financial activity] (BELF)

(5a) *The* <u>decrease</u> *is* ***primarily due to*** <u>the reduction</u> *of the Danish corporate tax rate [...]*

(5b) *The sales* <u>contraction</u> *in this area was* ***mainly due to*** *the* <u>weak performance</u> *of Hong Kong and Macau.*

(5c) *The* <u>decrease</u> *was* ***primarily due*** *to an extraordinary* <u>repayment</u> *from the German tax authorities [...].*

(5d) *The* <u>increase</u> *is* ***mainly due to*** *the* <u>measurement</u> *at fair value of outstanding hedging instruments.*

Discussion

- The study of LxPs provides insights on phraseological variation within very specific genres

  (Stability of LxPs enhance cohesion, and inter-textual generic cohesion plays a part in defining a unified, authoritative and reliable authorial voice)

- Apart from identifying regularities, it it is now relatively easy to identify sub-patterns and non-prototypical pattens in a systematic way

  (Non-prototypical uses may be linked to marked polarity)

References

Biber, Douglas. 2006. *University language: A corpus-based study of spoken and written registers*. Amsterdam: Benjamin

Biber, Douglas, Susan Conrad, & Viviana Cortes. 2004. If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25/3, 371-405.

Evert, Stefan. 2004. *The statistics of word co-occurrences: Word-pairs and collocations.* Doctoral Thesis, University of Stuttgart.

Gledhill, Christopher. 1995. Collocation and genre analysis. The Phraseology of grammatical items in cancer research abstracts and articles. *Zeitschrift für Anglistik und Amerikanistik* 43 (1/1). 11-36.

Gledhill, Christopher. 2015. On the Phraseology of Grammatical Items in Lexico-grammatical Patterns and Science Writing. In Paul Thompson & Giuliana Diani (eds.) *English for Academic Purposes: Approaches and Implications.* Newcastle-upon-Tyne: Cambridge Scholars Publishing, pp 11-42.

Hoey, Michael. 2005. *Lexical priming: A new theory of words and language.* London: Routledge.

References

Hunston, Susan & Gill Francis. 1998. Verbs observed: A corpus-driven pedagogic grammar, *Applied Linguistics* 19/1, 45-72.

Jones, Martha & Sandra Haywood. 2004. Facilitating the acquisition of formulaic sequences: an exploratory study in an EAP context. In Schmitt, Norbert (ed.). *Formulaic sequences*. Amsterdam/Philadelphia: John Benjamins:

Legallois, Dominique & Jacques François (eds.). 2006. Autour des grammaires de constructions et de patterns. *Cahiers du CRISCO* 21. Caen: Presses de l'Université de Caen.

Renouf, Antoinette & John Sinclair. 1991. "Collocational frameworks in English." In Aijmer, Karin & Bengt Altenberg (eds.) *English Corpus Linguistics.* London: Longman. 128-143.

Siepman, Dirk. 2005. Collocation, colligation and encoding dictionaries, Part I: Lexicological aspects. *International Journal of Lexicography* 18, 409-443.

Sinclair, John McH & Anna Mauranen. 2006. *Linear Unit Grammar: Integrating Speech and Writing*. John Benjamins.

References

Stefanowitsch, Anatol & Stefan Gries. 2003. Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus linguistics* 8/2: 209-243.

Stubbs, Michael. 1995. Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of language* 2/1, 23-55

Tucker, Gordon. 1998. *The Lexicogrammar of Adjectives: A Systemic functional approach to lexis.* London: Cassell Academic