

Discovering clines of variation

The location and analysis of non-canonical forms in general reference corpora

Gill Philip

University of Bologna – Italy

Keywords: categorization, fixedness, units of meaning, variation, word play

1. Introduction

As the growth of large corpora and the resulting move away from the serendipitous collection of language examples makes available more and more data from which language norms, and deviations from them, can be identified, studies into phraseology continually reaffirm the fact that “so-called ‘fixed phrases’ are not in fact fixed” (Sinclair 1996:83). Yet, in spite of this recognition, the discovery of non-canonical forms in corpora is still generally considered to be a matter of good fortune (Moon 1998:51). Further to this, the fact that the data which can be extracted is heavily restricted to that which is entered into the search query is cited all too often as one of the drawbacks of corpus-related research into the phenomenon (Moon 1996:252; Deignan 1999:197). After all, how can one search for something without knowing what that something is? This paper presents procedures for the location of non-canonical ‘fixed’ phrases, and puts forward some arguments in favour of their inclusion in phraseological analysis.

2. Searching for variant forms

Corpus query syntax varies from programme to programme, but a factor shared by the majority of applications is that search procedures are rarely exploited to the full. Although tagging allows fairly abstract searches to be carried out in the search for grammar patterns (see, for example, Hunston & Francis 1999), the retrieval of lexical variation is all too often limited to word forms and lemmas rather than phraseological structures (Moon 1996; Cioppa & Coffey 2000:550). But this need not be the case: serendipity can be replaced by the systematic, chance by trial and error.

This paper presents a method for the retrieval of variant forms from computer corpora (in this case, the Bank of English), demonstrating how key-words, wild cards and phrasal structure all contribute to the identification of underlying schemas which can in turn be translated into search queries. It is worth noting that the same principles are relevant to the use of advanced queries in Internet search engines, with the result that the linguist is able to extend and verify his/her findings on a much larger, if less homogeneous, data set – a factor of considerable importance when dealing with multi-word strings, where the location of sufficient examples for analysis can often pose a problem.

3. The linguistic value of phraseological variation

If finding the data is in itself a barrier to the study of variation, so too is the fairly low esteem in which it is held in language description. Perhaps surprisingly, canonical forms of idioms and other figurative phrases are actually quite uncommon in language corpora and are, as a general rule, outnumbered by their corresponding non-canonical forms. Despite this observation, non-canonical forms are still considered to be exceptions to the norm. They are peripheral to the interests of lexicography and foreign language teaching because they can ultimately be reduced to a canonical form, and it is this underlying canonical form, not its variants, which requires documentation and learning.

3.1. *The semantic effects of variation*

At the far extreme of the scale of variability lie puns and other types of word plays. Unlike their less showy counterparts, these attract considerable interest on the part of linguists. The many layers of meaning that they involve are held in place by an underlying canonical structure which appears to generate an infinite variety of novel utterances – see Partington (1998:121-143) on the exploitation of fixed phrases in newspaper headlines; see also Moon 1998; Philip 2003. In the study of word play, novel utterances are typically contrasted with their canonical relatives, all but ignoring the relationship that they have with the common-or-garden variation created on the fly by language users. A large swathe of language is overlooked as a result of it being neither normal nor exceptional enough to merit attention. Yet, it is precisely to this little-studied middle-ground that phraseology scholars can turn in order to discover more about language and the use that its speakers make of it.

The average language user can and does manipulate conventional structures to the particular communicative situation, but contrary to what our intuition might suggest, this does not necessarily imply that double meanings, humour, irony or other deliberate textual effects are created. What the study of variant forms shows is the extent to which learned forms can be adapted to make them contextually appropriate, whether this be in order to disambiguate, specify or reinforce meanings.

3.2. *Factors affecting phraseological variation*

In the view taken in this paper, canonical phrases are seen as a particular sub-set of collocational frameworks (Sinclair & Renouf 1991). This means that they provide a structure which permits the productive variation of some slots while resisting change to others. The most revealing productivity occurs along the paradigmatic axis where terms are substituted not only by members of the same semantic set but also by apparently unrelated terms. This phenomenon can be accounted for by the *Class Inclusion Hypothesis* (Glucksberg & Keysar 1993; Glucksberg & McGlone 1999), which posits that although classes are traditionally considered to be taxonomic, in the case of metaphorical language, they are attributive. If the relationship between the substituted term and the canonical one is based on common attributes, then everyday, unmarked variation would appear to be functioning along different lines from those which are intuitively believed to be in operation. This contributes significantly to the difficulties encountered in trying to predict how variant forms will manifest themselves. It also helps to explain why computational models of language find non-canonical forms difficult to account for, whereas most language users find both their interpretation and production unproblematic.

The study of canonical forms alongside non-canonical forms in all their guises also highlights the relationships which hold between phraseological items and their co-textual environments. In fact, the analysis of corpus data demonstrates that the central element of an extended unit of meaning (Sinclair 1996), typically taken to be a single word, can just as readily take the form of an entire phrase, canonical or otherwise (Philip 2003). The data suggests that although it is possible for a canonical phrase to occur in an unusual co-text, thus triggering contextually-relevant interpretations, it is extremely rare for a non-canonical form of a phrase to occur in a context which differs markedly from the norm (Philip 2003). When it does occur, it tends to occur in transcribed spontaneous speech when two similar structures are fused during on-line processing, resulting in a 'crack' in the phraseological priming (Hoey 2005:11).

4. Conclusions

It is the unpredictability of non-canonical forms that makes them difficult to extract from a corpus, and this certainly hinders attempts to study them comprehensively. Although this paper shows that their location is not governed by mere happenstance, the manual extraction of the type described here still poses a challenge to automation: automatic extraction requires a degree of stability which non-canonical forms do not always comply with. However, the study of variant forms can contribute fruitfully to existing linguistic knowledge bases, providing a semantic dimension that has yet to be explored in corpus studies. This in turn will contribute towards future success in automatic extraction of variant forms from corpora, as well as to shaping and refining our understanding of how natural languages really work.

References

- Cicogni L. & S. Coffey (2000) A Corpus study of Italian Proverbs: implications for lexicographical description. In *Euralex 2000 proceedings*. Stuttgart: Universität Stuttgart, 549-555.
- Deignan A. (1999) Corpus-based research into metaphor. In Cameron L. & G. Low (eds) *Researching and Applying Metaphor*. Cambridge: Cambridge University Press, 177-199.
- Glucksberg S & B. Keysar (1993) How Metaphors Work. In Ortony A. (ed.) *Metaphor and Thought*. Cambridge: Cambridge University Press, 401-424.
- Glucksberg S. & M.S. McGlone (1999) When love is not a journey: What metaphors mean. *Journal of Pragmatics* 31: 1541-1558.
- Hoey M. (2005) *Lexical Priming: A new theory of words and language*. London: Routledge.
- Hunston S. & G. Francis (1999) *Pattern Grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam & Philadelphia: Benjamins.
- Moon R.E. (1996) Data, Description, and Idioms in Corpus Lexicography. In *Euralex '96 Proceedings*. Gothenburg: Göteborg University, 245-256.
- Moon R.E. (1998) *Fixed Expressions and Idioms in English: A Corpus-Based Approach*. Oxford: Clarendon.
- Partington A. (1996) *Patterns and Meanings: Using Corpora for English Language Research and Teaching*. Amsterdam & Philadelphia: Benjamins.
- Philip G. (2003) *Connotation And Collocation: A Corpus-Based Investigation Of Colour Words In English And Italian*. PhD Thesis. Birmingham: The University of Birmingham.
- Sinclair J.M. (1996) The Search for Units of Meaning. *TEXTUS IX* (1): 71-106.
- Sinclair J.M. & A. Renouf (1991) Collocational Frameworks in English. Reprinted in Foley J.A. (ed.) (1996) *J.M. Sinclair on Lexis and Lexicography*. Singapore: Unipress, 55-71.



UCL

Université
catholique
de Louvain

Centre for English Corpus Linguistics
Centre d'étude des lexiques romans

PHRASEOLOGY 2005

The many faces of Phraseology
An interdisciplinary conference



PHRASÉOLOGIE 2005

La phraséologie dans tous ses états
Colloque interdisciplinaire

Louvain-la-Neuve, 13-15 October 2005

EDITORS

C. COSME, C. GOUVERNEUR, F. MEUNIER & M. PAQUOT

2005

TABLE OF CONTENTS

SECTION 1: KEYNOTE PAPERS

Blumenthal Peter <i>Profil combinatoire des mots: analyse contrastive</i>	1
Gross Gaston <i>Réflexions sur la notion de figement</i>	5
Heid Ulrich <i>Computational Phraseology: Approaches to the computational analysis and representation of phraseological units and to their extraction from text corpora</i>	13
Kennedy Graeme <i>Focus on Form and Meaning: Phraseology in Foreign Language Learning and Teaching</i>	17
Sinclair John <i>The phrase, the whole phrase, and nothing but the phrase</i>	19
Wray Alison <i>Looking at the WHY in Phraseology: a psycholinguistic perspective on patterns in text</i>	23

SECTION 2: PAPERS - POSTERS - SOFTWARE DEMOS

Adolphs Svenja <i>Multi-word units and second language speech fluency</i>	27
Alba-Salas Josep <i>'Hacer miedo' and 'dar miedo': a Corpus-based Diachronic Study of Spanish Collocations with 'hacer/dar' + State Noun</i>	31
Arnaud Pierre, Ferragne Emmanuel, Lewis Diana & Maniez François <i>[Adj+N] sequences in attributive or NP-final positions: observations on lexicality</i>	35
Augusto Maria Celeste <i>'Sourd comme une porte' ou 'sourd comme une caille'? Une approche contrastive des expressions phraséologiques du type comparatif en portugais et néerlandais</i>	39
Ballier Nicolas <i>MWUs licensing noun complement clauses in English</i>	43

Barlow Michael	45
<i>Collocate</i>	
Biorci Grazia & Cini Monica	49
<i>Tirare su un bambino'... L'élever ou le soulever? Aspects phraséologiques dans les verbes syntagmatiques</i>	
Boers Frank & Stengers Hélène	53
<i>Adding sound to the picture: an exercise in motivating the lexical composition of idioms in English, Dutch and Spanish</i>	
Bolly Catherine	57
<i>Constructions récurrentes, collocations et séquences figées avec le verbe à haute fréquence 'prendre': Pour une méthode «mixte» d'analyse de corpus (FL1 et FL2)</i>	
Čermák Frantisek & Blatná Renata	61
<i>Grammar idioms</i>	
Coffey Stephen	65
<i>Exploring the phraseological environments of the particle 'a' in modern English</i>	
Colson Jean-Pierre	69
<i>Improving translation practice by the semi-automatic extraction of set phrases</i>	
Cosme Christelle & Gilquin Gaëtanelle	73
<i>Free and bound prepositions in a contrastive perspective: the case of 'with' and 'avec'</i>	
Coxhead Averil	77
<i>"I like to learn verbs because you only have to learn one word"</i>	
Davis Boyd & Lunsford Ronald	81
<i>Metonymy in Alzheimer's speech: when the whole is more than the sum of its parts</i>	
DeCesaris Janet & Battaner Paz	83
<i>Idioms and Fixed Expressions in Learner's Dictionaries: Proposals for Spanish</i>	
De Cock Sylvie	87
<i>The routine aspects of speech and writing. a comparative corpus-driven study of native and nonnative performance data</i>	
Dister Anne, Fairon Cédric & Watrin Patrick	91
<i>Recherche d'expressions figées dans la presse. Méthode et observations</i>	
Diwersy Sascha	97
<i>La combinatoire lexico-syntaxique des noms de qualité agentive en français et en allemand</i>	

Dubroca Galin Danielle	101
<i>La phraséologie au service de la mercatique</i>	
Dziadkiewicz Aleksandra	105
<i>Analyse de phraséologismes pragmatiques français et polonais en vue d'une traduction automatique</i>	
Evert Stefan	111
<i>Empirical Research on Association Measures – the UCS Toolkit</i>	
Evert Stefan & Krenn Brigitte	113
<i>Exploratory Collocation Extraction</i>	
Eyckmans June, Kappel Jenny, Stengers Hélène, Demecheleer Murielle & Boers Frank	117
<i>Formulaic Sequences and Perceived Oral Proficiency: Putting a Lexical Approach to the test</i>	
Farø Ken	121
<i>On Motivation/Iconicity and Arbitrariness of Idioms</i>	
Fedulenkova Tatiana	125
<i>Isomorphism and Allomorphy of English, German and Swedish Phraseological Units Based on Metaphor</i>	
Fedulenkova Tatiana & Anufriyeva Maria	129
<i>Some Tendencies in Lexicalisation of English Phraseological Units</i>	
Fiedler Sabine	133
<i>Phraseology as a Learning Aim and Aid: Language Play in the Classroom</i>	
Fitzpatrick Tess	137
<i>An exploration of factors affecting the acquisition and use of formulaic sequences in the L2</i>	
Forsberg Fanny	141
<i>Prêt-à-parler: le rôle des séquences préfabriquées en français parlé L2 et L1</i>	
Frath Pierre & Gledhill Christopher	145
<i>Qu'est-ce qu'une unité phraséologique?</i>	
Fujimura Itsuko & Nakao Hiroshi	149
<i>Noms composés et choix de l'article 'de' et 'des' en français</i>	
Gilquin Gaëtanelle	155
<i>To take or not to take phraseology into account: the place of multi-word sequences in corpus data and experimental data</i>	

Gouverneur Céline	159
<i>The phraseological patterns of high frequency verbs in advanced English for General Purposes: new perspectives</i>	
Granger Sylviane	165
<i>Pushing back the limits of phraseology: how far can we go?</i>	
Gustawsson Elisabeth	169
<i>'Straw-clutchers', 'kite-flying' and 'trail blazing': on derivations of verbal idioms in the British National Corpus</i>	
Hammer Françoise	173
<i>Le diable et le bon Dieu - rencontres phraséologiques</i>	
Handl Susanne	177
<i>Collocation - a question of direction and weight</i>	
Hanks Patrick	181
<i>Resonance and the Phraseology of Metaphors</i>	
Heyvaert Liesbeth	185
<i>A phraseological approach to nominalization: issues of collocation and colligation</i>	
Hümmer Christiane	189
<i>Meaning and Use: a Corpus-based Case Study on German Idiomatic Multi-Word Expressions</i>	
Ishida Priscilla	193
<i>A method for the contrastive analysis of idioms: Japanese and English idioms of anger</i>	
Jaglińska Anna	199
<i>Idiomaticity in learner language: a study of the incidence and variability of prefabs in advanced EFL learners' writing</i>	
Jautz Sabine	203
<i>The Use of Expressions of Gratitude in Radio Phone-ins and Radio Interviews in British and New Zealand English: Instances of Politeness or of Impoliteness?</i>	
Jones Steven, Murphy Lynne, Paradis Carita & Willners Caroline	207
<i>Googling for Opposites: a phraseological approach to assessing antonym canonicity</i>	
Kerz Elma	211
<i>A construction grammar approach to the analysis of lexico-grammatical patterns in academic texts</i>	

Klein Jean & Lamiroy Béatrice	215
<i>Relations systématiques entre paires d'expressions verbales figées à travers quatre variétés de français</i>	
Kozak Katarzyna	219
<i>Idioms of paralanguage: a comparative study</i>	
Kübler Natalie	223
<i>Phraséologie verbale transdisciplinaire et spécifique: essai de classification en sciences de la terre</i>	
Lecler Aude	227
<i>J'ai la mémoire qui flanche, j'me souviens plus très bien... Le défigement: réinvestissement et réinitialisation dans le cycle phraséologique</i>	
Legallois Dominique	231
<i>Du bon usage des expressions idiomatiques dans l'argumentation de deux modèles anglo-saxons: la grammaire de construction et la grammaire contextualiste</i>	
Levison Michael & Lessard Greg	235
<i>Generating Complex Verb Phrases: an Extension of Burger's French Verbal Morphology</i>	
Lindquist Hans & Levin Magnus	243
<i>FOOT and MOUTH: the phrasal patterns of frequent nouns</i>	
Maclagan Margaret, Davis Boyd & Tillard Gina	247
<i>Fixed phrases in the speech of patients with dementia</i>	
Manerko Larissa & Tuarmenskaya Angela	249
<i>Spatial Conceptualization of Human Body Parts in English Phraseology</i>	
Martí Solano Ramón	253
<i>Variation and Exploitation of Phraseological Units of Body Parts in the British Press: Types of Modification</i>	
Martin Willy	257
<i>A Unified Approach to Semantic Frames and Collocational Patterns</i>	
Mason Oliver	261
<i>Automatic Identification of English Multi-Word Units</i>	
Meunier Fanny	265
<i>Fine-tuning the lexical approach</i>	
Moon Rosamund	269
<i>'White as snow, red as blood, black as ebony': conventionalized English similes and corpus evidence</i>	

Mortchev-Bouveret Myriam	273
<i>Etude phraséologique du verbe CASSER : fonctions lexicales et classes sémantiques pour un encodage lexicographique</i>	
Mudraya Olga, Piao Scott S.L., Löfberg Laura, Rayson Paul & Archer Dawn	277
<i>English-Russian-Finnish Cross-Language Comparison of Phrasal Verb Translation Equivalents</i>	
Murano Michela	283
<i>La phraséologie dans les préfaces des dictionnaires bilingues français-italien, italien-français</i>	
Naciscione Anita	289
<i>Cognitive Aspects of Visual Representation of Phraseological Image</i>	
Neff van Aertselaer JoAnne	293
<i>Contrasting English-Spanish discourse phrases: A corpus study</i>	
Nenonen Marja	297
<i>Prototypical idioms: evidence from Finnish</i>	
Nuccorini Stefania	301
<i>"Absolutely no? Absolutely not!" On some contrastive aspects in the phraseology of English and Italian</i>	
Oakey David	305
<i>Absolute and pragmatic frequencies of word combinations in academic journal articles in English</i>	
Omazić Marija	309
<i>Conceptual Motivation and Processing of Idioms and Idiom Modifications</i>	
Oncins-Martínez José	311
<i>Another plea for 'phraseo-stylistics' - or how literary texts can contribute to the development of phraseology: the case of Shakespeare</i>	
Orliac Brigitte	315
<i>Extraction automatique de collocations en langue de spécialité: une approche combinée</i>	
Osborne John	319
<i>Phraseology effects as a trigger for errors in L2 English: the case of more advanced learners</i>	
Paquot Magali	323
<i>EAP vocabulary in learner corpora: a cross-linguistic perspective</i>	
Parizoska Jelena	327
<i>False pairs in phraseology</i>	

Pazos Breña José Manuel & Pamies Bertrán Antonio	331
<i>Convergence de critères statistiques et grammaticaux pour la récupération d'unités phraséologiques en corpus électronique</i>	
Pecman Mojca	335
<i>Compilation, formalisation and presentation of bilingual phraseology: problems and possible solutions</i>	
Philip Gill	339
<i>Discovering clines of variation: the location and analysis of non-canonical forms in general reference corpora</i>	
Piérard Sophie & Bestgen Yves	343
<i>Identification automatique des marqueurs globaux du discours par l'analyse des expressions récurrentes</i>	
Piirainen Elisabeth	347
<i>Widespread Idioms in Europe and Beyond: Toward a Cross-linguistic and Cross-cultural Research Project</i>	
Pontonx Sophie de	351
<i>Expressions idiomatiques françaises et allemandes: vrais et faux amis. Une aide à l'apprentissage des phrasèmes</i>	
Prat Zagrebelsky Maria Teresa	357
<i>Difficulties in dealing with phraseological patterns on the part of Italian advanced EFL learners: a corpus-based approach</i>	
Reboul-Touré Sandrine & Pagnier Thierry	359
<i>Rouge comme une tomate' ou 'bleu comme une orange': collocations et créations dans le cadre de la didactique du français langue maternelle</i>	
Renouf Antoinette	363
<i>Issues of automatic phrase retrieval in web text</i>	
Römer Ute	369
<i>A phraseology-driven approach to identifying evaluation in a book review corpus</i>	
Romero Fresco Pablo	373
<i>The analysis of phraseological translation in a parallel audiovisual corpus: a proposal for a dictionary of idioms in use (English-Spanish)</i>	
Sabban Annette	377
<i>Some considerations on the culture-boundness of phraseology</i>	

Schmitt Norbert	381
<i>Processing advantages of formulaic sequences vs. non-formulaic sequences in text passages</i>	
Segermann Krista	382
<i>Les unités phraséologiques comme unités d'apprentissage dans l'enseignement du français langue étrangère</i>	
Siepmann Dirk	387
<i>Phraseology and Learners' Dictionaries: Coverage, Generativity, Semasiological vs. Onomasiological Approaches</i>	
Simpson Rita & Ellis Nick	391
<i>An Academic Formulas List (AFL)</i>	
Stengers Hélène, Eyckmans June, Horemans Arnout & Boers Frank	395
<i>Optimising mnemonic strategies through CALL: a tool called Idiom Teacher</i>	
Stubbs Michael	397
<i>The most natural thing in the world: quantitative data on multi-word sequences in English</i>	
Svensson Maria Helena	401
<i>Un critère de figement très nuancé: la non-compositionalité</i>	
Tschichold Cornelia	405
<i>A computational lexicography approach to phraseologisms</i>	
Tutin Agnès	407
<i>Collocations du lexique transdisciplinaire des écrits scientifiques: annotation et extraction des propriétés syntaxiques</i>	
Valetopoulos Freiderikos	413
<i>Pour exprimer son état émotionnel en grec: le cas des locutions verbales</i>	
Van der Wouden Ton	417
<i>On the phraseology of stop words</i>	
Vangehuchten Lieve	421
<i>L'identification de l'unité phraséologique en terminologie: énoncé du problème à partir de l'analyse d'un corpus de discours économique d'entreprise</i>	
Vásquez González Juan Gabriel	425
<i>"The company a word keeps": collocates as a means for categorizing Old English</i>	

Vergely Pascale	429
<i>Mise au jour de structures régulières dans un corpus oral de communications de travail: l'exemple de la navigation aérienne</i>	
Verlinde Serge, Selva Thierry & Binon Jean	433
<i>Les unités polylexicales: de leur détection à leur apprentissage</i>	
Wanzeck Christiane	437
<i>The Etymology of Phraseological Units</i>	
Wikberg Kay	441
<i>Similes in the BNC</i>	
Williams Geoffrey	445
<i>The Good Lord and his works: A corpus-based study of collocational resonance</i>	