International Conference

Filling Gaps, Building Bridges:
Qualitative and Quantitative Approaches to the Study of Literature

Dept. of Linguistic and Literary Studies
University of Padua, Italy

7-9 June 2018

Book of Abstracts

GUEST LECTURES

Nigel FABB (University of Strathclyde, UK; N.Fabb@strath.ac.uk)

"What do we know when we know a literary form? Corpus studies and weighted expectations"

Literary form can be represented as statistical generalizations over literary texts, and these generalizations may be psychologically modelled as weighted expectations. The expectations are partly derived from experience of literary texts, and partly from learning explicit rules of form. Thus, for example, when listening to the first line of an English sonnet, we formulate an expectation based on counting syllables, such that the tenth syllable must be word-final, and that stresses in polysyllabic words will be on even-numbered syllables (cf. Fabb, Halle 2008); e.g.:

Shall I compare thee to a summer's day?

The expectation is constructed partly from explicit knowledge of sonnets and iambic pentameter lines, and partly from exposure to a large number of sonnets and iambic pentameter lines from which weighted expectations (relating to how rhythmically periodic the line should be, how long it should be, etc.) are derived. Expectations can be met (as here) or violated. The general possibility of violating expectations relates to the aesthetic principle of 'uniformity amidst variety' and presumably contributes to the emotional experience of aesthetic form (Fabb 2016; Turpin, Fabb 2017). Further, since we are continually revising expectations based on prior experience, what counts as uniform (expected) and what as varied (unexpected) will also be changing dynamically. (The notion of literary form as weighted expectations is a reinterpretation of the proposal in Fabb (2002) that literary form holds by virtue of stronger or weaker inferences about a text.)

Corpus studies – for example studies of rhythm-metre relations across a corpus – give us statistical generalizations. Hence the results of corpus studies for a set of texts may have some relation to what we know when we know the literary form of those texts, though there are many possible gaps in the relation between what a corpus study reveals and what a person's knowledge of the form might be. In this paper I consider some of the ways in which we might reinterpret corpus studies of form in terms of the psychological experience of form. One of the issues relates to the procedural nature of expectations, which can change or be different at different stages in the reading of a text; how can this be captured in a corpus study which is often not procedural but declarative? Another issue relates to whether forms vary in their learnability, and that this is independent in principle of their frequency in a corpus; for example, some forms might violate basic schemata, or principles relating linguistic to literary form. A third issue is that certainty is unlikely in aesthetic experience; that is, we expect our expectations sometimes to fail. How is this uncertainty realized statistically across a corpus? Finally, in order to map the statistical characteristics of a corpus onto the experience of poetic form, we might need to understand the specific ways in which characteristics aggregate to determine our experience: how do two kinds of expectation combine?

References

Fabb N. (2002) *Language and literary structure: the linguistic analysis of form in verse and narrative,* Cambridge, Cambridge University Press

Fabb N., Halle M. (2008) *Meter in poetry: a new theory*, Cambridge, Cambridge University Press.

Fabb N. (2016) "Processing effort and poetic closure", in *International Journal of Literary Linguistics* 5 (4), 1-22.

Turpin M., Fabb N. (2017) "Brilliance as cognitive complexity in aboriginal Australia", in *Oceania* 87(2), 209-230.

Michaela MAHLBERG (University of Nottingham, UK; m.a.mahlberg@bham.ac.uk)

"CLiC – studying fiction in the digital age"

Corpus linguistic methods can be seen as early manifestations of digital humanities. However, corpus research has traditionally focused on non-literary texts. Increasingly, the umbrella term *corpus stylistics* (cf., e.g., Semino, Short 2004) is used to describe approaches that employ corpus linguistic methods to study literary qualities and individual literary texts. Still, in order to be able to account as fully as possible for features of literary texts we need to create new tools and develop methodologies that are tailored to the task at hand. In this paper, I will illustrate key functionalities of the web application CLiC (http://clic.bham.ac.uk/) that has been specifically designed for the corpus linguistic study of narrative fiction. The case study that I will present looks at textual patterns that contribute to the creation of fictional characters. The examples will be drawn from the CLiC corpora. The CLiC corpora comprise more than 130 books across four subcorpora: the corpus of Dickens's Novels, the 19th Century Reference Corpus (19C), the Corpus of 19th Century Children's Literature (ChiLit) and the Corpus of Additional Requested Texts (ArTs). For all CLiC texts, direct speech and specific places around speech have been marked up (Mahlberg et al. 2016). Hence, CLiC can run searches across defined textual subsets and support the analysis of features of narrative fiction. The findings that can be generated with CLiC also open up possibilities for the comparison of patterns in literary and non-literary language. In this talk, I will not only illustrate the innovative potential of corpus methods for the study of literature, I will also make links to relevant current debates in the digital humanities.

References

Mahlberg M., Stockwell P., Joode J. de, Smith C., O'Donnell, M. B. (2016) "CLiC Dickens: novel uses of concordances for the integration of corpus stylistics and cognitive poetics", in *Corpora*, 11(3), 433–463 (https://www.euppublishing.com/doi/full/10.3366/cor.2016.0102).

Semino E., Short M. (2004) *Corpus stylistics. Speech, writing and thought presentation in a corpus of English writing*, London, Routledge.

Jan Christoph MEISTER (University of Hamburg, Germany; jan-c-meister@uni-hamburg.de)

"Phenomenology vs. empirics?
From qualitative to quantitative analysis of literature – and back"

Qualitative and quantitative approaches in digital literary studies are often presented as mutually exclusive methods. Franco Moretti's (2013) – consciously polemic – coinage of the term *distant reading* as an approach to literature that is antagonistic to the traditional *close reading* approach is a fairly recent example for such a dichotomy. However, what is often misinterpreted as an almost manichean methodological distinction triggered by the advent of digital humanities is in fact historically rooted in the 19th century discourse which tried to draw a clear distinction between *Geisteswissenschaften* and *Naturwissenschaften*. In the field of literary studies and linguistics this distinction has been played out numerous times ever since – during Russian Formalism, Prague Structuralism, French Structuralism, Poststructuralism, and more generally in the growing divide between the more empirical approaches favoured in linguistics on the one hand, and the more speculative and hermeneutic approaches favoured in traditional literary studies on the other.

In my paper I will argue that this commonplace distinction of apparently opposing methodologies is a misleading simplification. Qualitative and quantitative approaches should rather be conceptualized as a methodological continuum. I will use the concrete example of digital text annotation or *markup* to illustrate this point and demonstrate the mutual benefits to be derived by both "camps", the empirical and the phenomenological.

Reference
Moretti F. (2013) *Distant reading*, London & New York, Verso Books.

Federica PERAZZINI (Sapienza University of Rome, Italy; federica.perazzini@uniroma1.it)

**"The figure in the discourse:
computational criticism and the quality of quantitative literary research"**

As a combination of data-driven research methods and in-depth analyses of literary texts, *computational criticism* has allowed contemporary scholars to pursuit compelling literary and historical analyses on an unprecedented level. Indeed, from Burrows' (1987) book *Computation into Criticism* until the most recent attempts of application of *distant reading* (Moretti 2013), *computational macroanalysis* (Jockers 2013), and *literary pattern recognition* to studies of large-scale cultural dynamics, research in the digital humanities has always had to cope with a certain imbalance between quantitative guarantees and qualitative expectations.

In the perspective of a possible reconciliation, the article presents a series of experiments in computational literary history aiming to identify the keywords that stand as the ideological common ground of the new discursive practices encompassed in the rising genres of Augustan England such as journalism and prose fiction. Reflecting, if not directly regulating, the worldview and tastes of the growing civil society, these genres soon established themselves as the privileged devices for the formation of modern subjectivity. The analysis of high frequency words and lexical specificities is performed on a reference corpus composed of both works of fiction published between 1688 and 1815 and five most representative periodicals of the time. The aim is to determine the stylometric singularity of the novelistic genre as well as the distinctive character of canonical fiction over its non-canonical counterpart.

References

Burrows J. F. (1987) *Computation into criticism: a study of Jane Austen's novels*, Oxford, Oxford University Press.
Jockers M. L. (2013) *Microanalysis: digital methods and literary history*, Champaign, University of Illinois Press.
Moretti F. (2013) *Distant reading*, London & New York, Verso Books.

Arjuna TUZZI (University of Padua, Italy; arjuna.tuzzi@unipd.it)

"Statistical methods for the analysis of literary texts"

In your next research project, are you planning to take into account a large number of novels in order to map their main contents or grasp similarities and differences among them? Are there definitely too many texts for one scholar to read in a lifetime? Why don't you try asking a computer to do this for you? A software package (and a statistical model behind it) cannot *close read* a text. On the contrary, it might be smart enough to *distant read* a text, i.e. collecting data, retrieving relevant information, summarising features, finding patterns, etc.

By means of practical examples this workshop aims to show how quantitative methods can be part of the research instrumentation and the "toolbox" of literary scholars. This perspective is innovative because quantitative methods for text analysis are fostered in a field that is traditionally explored through qualitative approaches. The workshop shows that it is possible to exploit "texts as data" and that quantitative methods can work successfully alongside qualitative methods in many studies. The quantitative analysis adopted is essentially based on words and word counts (i.e. the *bag of words* approach), and, in particular, on the presence, absence and occurrence of words over texts, authors, groups of authors.

FULL PAPERS

Lucia ASSENZI (University of Padua, Italy; lucia.assenzi@phd.unipd.it)

"Quantifying auxiliary *tun* to study 17th century German linguistic reflection?"

This paper will show how the quantitative analysis of the syntactic phenomenon called *tun* periphrasis can contribute to better defining the characteristics of the common literary German language proposed by the 17th century German linguistic society, the *Fruchtbringende Gesellschaft*. The quantification of auxiliary *tun* in the *Erzehlungen* (1624), one of the many translations produced in connection to the linguistic reflection of the *Fruchtbringende Gesellschaft* is expected to reveal the syntactic regional markedness of the ideal standard German the society was trying to propose as a common literary language.

Pursuing this kind of analysis is particularly important because the theoretical linguistic works and grammars produced by members of the *Fruchtbringende Gesellschaft* are quite specific as to the correct phonology, morphology and orthography of the proposed form of standard German, while they however explicitly mention syntax. The syntactic features of the ideal standard German proposed by this 17th century linguistic society can therefore only be explored thanks to an accurate quantitative study of syntactical phenomena in the literary texts written in connection to the linguistic reflection of the *Fruchtbringende Gesellschaft*.

Anna BONIFAZI (University of Stuttgart, Germany; anna.bonifazi@ts.uni-stuttgart.de)

"Research on anaphoric markers in literary texts: how to combine qualitative and quantitative analyses?"

Linguistic accounts of anaphora processing could very much profit from literary evidence. In fact, through textual units that exceed the sentence level, literary discourse widens and deepens the range of pragmatic and cognitive usages of anaphora, especially if we consider discourse structure and viewpoint. The paper offers a qualitative analysis of various anaphoric markers from texts belonging to different genres and ages. It takes into account not only the criteria of accessibility and associativity, but also narrative projection, narrative discontinuities, genre requirements, and the marking of viewpoint. From the analysis carried out, the paper derives questions and challenges concerning the electronic recognition of anaphoric strategies, and most of all the classification/categorization of the various lexical markers being used in a large corpus of data.

Scott BORGESON (Stanford University, USA; borgeson@stanford.edu)
Arto ANTTILA (Stanford University, USA; anttila@stanford.edu)
Ryan HEUSER (Stanford University, USA; heuser@stanford.edu)
Paul KIPARSKY (Stanford University, USA; kiparsky@stanford.edu)

"Prose Rhythm and Antimetricality"

With the recent availability of metrically annotated texts and computational tools for detecting meter, the study of prose rhythm is ripe for a revival. The automatic scansion tool *Prosodic* (Heuser, Falk, Anttila 2017) can distinguish not only between verse and prose (Anttila, Heuser 2016), but also degrees of metricality in prose. Our study tests the hypothesis that some prose writers avoid trespassing on the formal space staked out by metrical verse, while others flout this precept or ignore it.

We examined Shakespeare's *Sonnets* and three prose texts: Walter Pater's *Two Early French Stories*, Thomas Browne's *Religio Medici* and Dickens' *David Copperfield*. After converting them to lineated versions – where each line consists of five consecutive words (Anttila and Heuser 2016) – we produced a RANDOMIZED text that replaces words with words of the same part of speech, keeping the syntax relatively intact, and a SCRAMBLED text that disregards part of speech, destroying the syntax. We then annotated the texts automatically using *Prosodic*, based on five metrical constraints (e.g. *S/UNSTRESSED 'No unstressed syllables in strong positions'; W-RESOLUTION: 'In disyllabic positions within a word, the first position is light and stressed'; Hanson, Kiparsky 1996). Finally, we quantified metricality in two ways: (i) the presence vs. absence of constraint violations in a line (i.e. perfect vs imperfect lines); (ii) the number of possible scansions of a line (ranging from 1 to 64), where the more scansions, the more metrically ambiguous the line.

The analysis shows: (a) that Shakespeare's original text is highly metrical, while Pater's and Browne's original texts are antimetrical compared to the baseline (randomized Dickens) and (b) that the number of scansions/violations is higher in the randomized text than the original texts, and the highest in the scrambled text, indicating that preserving syntax also preserves metrical structure. The findings suggest that automatic scansion is a viable methodology for the study of prose rhythm.

References

Anttila A., Heuser R. (2016) "Phonological and metrical variation across genres", in *Proceedings of the Annual Meeting on Phonology* 2015, 1-12.

Hanson K., Kiparsky P. (1996) "A parametric theory of poetic meter", in *Language* 72(2), 287-335.

Heuser R., Falk J., Anttila A. (2017). *PROSODIC: a metrical-phonological parser*, written in Python. https://github.com/quadrismegistus/prosodic.

Erik CASTELLO (University of Padua, Italy; erik.castello@unipd.it)

"Dickens's *Pictures from Italy* and Murray's *Handbook to Northern Italy*:
a quantitative/qualitative investigation of the use of adjectives in the two texts"

This study explores the use of adjectives in Dickens's *Pictures from Italy* and Murray's *Handbook to Northern Italy,* with a view to exploring similarities and differences between them. It is based on a 17,645-word corpus made up of sections chosen from the *Handbook* and a 17,600-word corpus comprising chapters 5 to 9 of *Pictures from Italy.* The *Corpus of Late Modern English Texts* (De Smet 2005) is used as the reference corpus. The analysis focuses both on the adjectival keyword lists obtained from the comparison of the two corpora to the reference corpus and on the syntactic patterns the adjectives are used in (i.e. "adverb + adjective" and "adjective + adjective + noun"). The preliminary findings reveal that, compared to Murray, Dickens uses a higher percentage of adjectives and also more infrequent ones, as well as more patterns of hyphenated compound adjectives. This suggests that *Pictures from Italy* is overall a more inventive and sophisticated piece of writing than the *Handbook*, although the former takes inspiration from the latter to some extent, and that Dickens's work mainly addresses independent travellers rather than conventional tourists.

Reference
De Smet, H., 2005. A corpus of Late Modern English texts, in *ICAME Journal,* 29, 69–82.

Fabio CIOTTI (University of Rome "Tor Vergata", Italy; fabio.ciotti@uniroma2.it)

"Italian narrative between Realism and Modernism: a computational analysis of verbal tenses"

The transition from Realism to Modernism marks a literary-historical period that has attracted the attention of a number of critics and historians of literature. Our research program aims to identify the intrinsic textual features that characterize that transition in Italian narrative texts of that period. This literary phenomenon, with its chronotopical dimension, is an optimal candidate for an analysis based on the distant reading approach (Moretti 2013) and on related computational analytical methods, provided we can identify an adequate narrative feature that we can properly operationalize. We have identified a candidate in the usage of verbal tenses in the text, following Harald Weinrich's (1964) suggestion. The syntagmatic and paradigmatic distribution of temporal morphemes is a textual linguistic property that can be relatively easily analyzed with computational tools. The results obtained from the analysis of a corpus of about 600 Italian narrative texts (from 1850 to 1920 ca.) suggest that the literary phase under consideration is indeed marked by a restructuring of the tense system in the texts.

References
Moretti F. (2013) *Distant reading*, London & New York, Verso Books.
Weinrich H. (1964) *Tempus: besprochene und erzahlte* (*Sprache und Literatur*, vol. 16), Stuttgart, W. Kohlhammer.

Valentina COLONNA (University of Genoa and University of Turin, Italy;
valentina.colonna@unito.it)

"Prosodies of *Congedo* – Comparative analysis of twelve readings"

The reading of poetry is one of the most overlooked aspects of Italian studies in Criticism and Linguistics. This experimental research arises from Beccaria's (1964) studies, and explores the musical *apparatus* of a poetic composition through methods of instrumental Phonetics. The study focuses on the prosody of *Congedo del viaggiatore cerimonioso* read by Giorgio Caproni. The approaches are qualitative and quantitative: from the analysis of the author's reading, the observation has been extended to other eleven voices, representing different professional domains across different time periods (contemporary poets, radio speakers and actors). Listening represents the start of the research, combined with the observation of oscillograms and spectrograms and their annotation using *Praat* (Boersma, Weenink 2017). However, the sound data is always related to the textual data, by identifying the main trends in melody and prosodic segmentation. The study, in addition to highlighting considerable speech variation, reveals aspects of Italian poetry that have always been neglected, but which can now be explored, thanks to a quantitative method. The is also meant to be a starting point for "VIP - Voices of Italian Poets", a project investigating original and contemporary readings of Twentieth Century Italian poems, accompanied by linguistic annotations at different levels.

References
Beccaria G. L. (1964) *Ritmo e melodia nella prosa italiana*, Olschki, Firenze.
Boersma P., Weenink D. (2017) *Doing phonetics by computer* (http://www.fon.hum.uva.nl/praat/).

Joseph DAVIS (The City College of New York, USA; jdavis@ccny.cuny.edu)

"Theme and quantification in the service of linguistics"

Quantitative analysis of literary texts can support linguistics. Predictions about the distribution of linguistic forms in texts, grounded in an awareness of theme, lead to data that support grammatical hypotheses. Three counts on Italian texts will illustrate the approach at three levels of discourse. In twentieth-century literature, pronoun *egli* 'he' tends to refer to principal characters, while *lui* 'he' tends to refer to secondary characters. But Lampedusa's (1984 [1958]) *Il gattopardo* 'The Leopard' patterns differently; the reason has to do with the interaction of the forms' structural properties and the peculiar nature of that novel. At a text-internal level, a chapter in Devoto's (1951) *Gli antichi italici*, namely "The Ancient Italic Peoples", which is devoted to "Alphabets and Dialects", patterns differently from one on "Becoming Part of the Roman World" in terms of the distribution of clitic *si* versus the accusatives. And two sections of a pivotal scene in Silone's (1937) *Pane e vino*, namely "Bread and Wine", differ markedly in the distributions of locative clitics *vi* and *ci*, both unrevealingly glossed as 'there.' These results serve to support grammatical hypotheses of non-canonical categories. In addition, they enhance our understanding of the texts, especially their overall themes and characterizations, internally varying emphases, and the dramatic structure of episodes.

References
Devoto G. (1951) *Gli antichi italici*, 2nd edition, Firenze, Vallecchi.
di Lampedusa G. T. (1984 [1958]) *Il gattopardo*, Milano, Feltrinelli.
Silone I. (1937) *Pane e vino*, Lugano, Capolago.

Gabriel EGAN (De Monfort University, UK; gegan@dmu.ac.uk)
Paul BROWN (De Monfort University, UK; paul.brown@dmu.ac.uk)

"N-gram matching as a test for authorship, genre, and date-of-composition of London's early modern plays"

A number of authorship attribution problems regarding early modern drama have recently been tackled by investigators searching for short phrases (n-grams for *n* from 2 to 7) common to the work being attributed and works in the canons of candidate authors. These investigations have involved searching for the n-grams in question – by more or less automated means – within electronic text corpora. Recently, however, a complete listing of all the n-grams common to each of 527 early modern plays has been published online and it offers an opportunity to attempt to independently replicate the findings of these recent attribution investigations. The present paper will attempt such replication and see what other purposes this n-gram data may be put to, including the study of whether n-gram sharing correlates with genre and date of composition.

Daniel GUTIERREZ-TRAPAGA (Universidad Nacional Autónoma de México, Mexico; dgtrapaga@hotmail.com)

"Qualitative and quantitative approaches to primary sources in Spanish literary history"

This paper will address methodological approaches to primary sources in Spanish literary history exemplified with two genres: romances of chivalry and gothic literature. It will focus on the gap between the existing data on the said genres, derived from primary sources, and the depiction and assessment of their relevance in several recent textbooks devoted to the history of Spanish literature. The data derived from primary sources makes it possible to argue that both romances of chivalry and gothic literature had an important role in the Spanish literary landscape of the sixteenth and nineteenth centuries, respectively. However, these two genres are usually deprecated or even completely ignored in Spanish literary history due to the preponderance of qualitative approaches. This methodological divergence raises two questions – the first is how qualitative arguments and methods may have justified the omission of prolific literary genres, and the second is what qualitative criteria may have been employed in Spanish literary histories to interpret and abandon quantitative information obtained from primary sources.

Geneviève HENROT (University of Padua, Italy; genevieve.henrot@unipd.it)

"Le Nom propre modifié [Dét+Np] dans les *Mémoires d'Outre-tombe* de Chateaubriand"

La consistance du personnage dans un texte (littéraire) tient non seulement aux différentes séquences descriptives, dialogales et narratives qui le campent sur la scène de la diégèse, mais aussi à la façon dont ses désignations sont relayées et son nom est modulé: en particulier, la construction [Dét.+Npr] est à même de faire subir au nom, et partant, à son porteur, toute une gamme de transformations qui révèlent la qualité des regards qui en filtrent l'image: anonymat ou notoriété, exemplarité ou emphase, métaphore, métonymie ou antonomase dessinent autant de changements possibles dans le système des places du personnel fictionnel. Les *Mémoires d'Outre-tombe* de Chateaubriand frappent en particulier par l'abondance de constructions emphatiques. Un relevé exhaustif sur corpus catégorisé (Frantext) sera le point de départ d'une analyse linguistique et pragmatico-poétique des effets de sens produits, localement et globalement, par les constructions du syntagme nominal proprial.

Elsi KAISER (University of Southern California, USA; emkaiser@usc.edu)
Catherine WONG (University of Southern California, USA; catherjw@usc.edu)
Gwenyth PORTILLO-WHIGHTMAN (University of Southern California, USA; gportill@usc.edu)

"Subjectivity from an experimental perspective:
effects of sensory modality on readers' interpretation of subjective adjectives"

To understand subjective adjectives (e.g. *amazing, disgusting)*, we need to know whose opinion/perspective is being expressed. Although these adjectives can involve different sensory modalities (e.g. sight, taste, smell), consequences of sensory differences on adjective interpretation are not yet well-understood. Does readers' interpretation of who the attitude-holder/experiencer of an adjective is depend on whether the situation involves seeing, smelling or tasting (e.g. 'it looked/smelled/tasted disgusting')? We report quantitative data suggesting that yes, sensory modality impacts identification of the attitude-holder. There are more first-person/narrator interpretations – as opposed to third-person/character-in-narrative interpretations – in the visual than the gustatory or olfactory domains.

Kobie van KRIEKEN (Radboud University, The Netherlands; k.vankrieken@let.ru.nl)

"Understanding perspective in narrative:
from qualitative analysis to quantitative assessment"

Both fictional and nonfictional narratives are characterized by the representation of perspective. This presentation discusses how a mixed-method approach can advance our understanding of the linguistic manifestation of perspective in narrative discourse as well as readers' interpretation thereof. Specifically, three studies illustrate how the combination of frameworks from literary theory and linguistics may help bridging the gap between qualitative and quantitative research on the language of stories: (1) a qualitative text-linguistic analysis of perspective in two stories; (2) a quantitative text-linguistic analysis of perspective in a historical corpus of 300 stories; and (3) a quantitative assessment of readers' interpretation of perspective in a range of story fragments. The combined results shed light on how linguistic perspective may manifest itself at a very local level (i.e. the word level), while revealing that its impact on readers' interpretation of stories stretches to the discourse level. These findings underscore the value of combining qualitative and quantitative research in understanding the language of stories.

Zequan LIU (Henan University, P. R. China; Zeqliu@163.com)

**"Applying a multivariate analysis in literary style investigation:
a case study of *The Old Man and the Sea*"**

The primary objective of this study consists in the construction of a comprehensive and multivariate quantitative approach for the investigation of literary style. For this purpose, this paper proposes six variables as an inter-subjectively verifiable network for quantitative stylistic query, which includes, (1) the involvement and (2) explicitness dimensions of spoken/written texts, (3) the standardized type-token ratio, (4) the moving average type-token ratio, (5) entropy and (6) the relative repetition rate of vocabulary richness. The applicability of the approach is tested in a case study conducted on the stylistic features of Hemingway's (1952) *The Old Man and the Sea*. Statistics thus derived not only confirms the utility of the approach, but also justifies the spoken nature of the Hemingway text as well as the simplicity of its vocabulary.

Reference
Hemingway E. (1952) *The old man and the sea*, London, J. Cape.

Ines LOZANO PALACIO (University of La Rioja, Spain; ines.lozano@ unirioja.es)

"Irony from a cross-disciplinary perspective: towards an integrated account"

In the study of irony – investigated by such disciplines as rhetoric, literature, philosophy, and linguistics – context-based literary discussions of the phenomenon contrast with analytical linguistic accounts of it. A comparative between the two types of approaches reveals their methodological gaps, which can be filled by enhancing cross-disciplinary dialogue with a view to developing a constructive integrated perspective. In the light of this, this paper puts forward an approach to the study of irony that combines insights from inferential pragmatics (e.g. Pretence Theory (Barnden forthcoming; Clark, Gerrig 1984), Relevance Theory (Wilson, Sperber 2012)), cognitive linguistics (e.g. Ruiz de Mendoza forthcoming) and literary theory (e.g. Muecke 1970, Hutcheon 1994, Booth 1974). It acknowledges the centrality of the relevance-theoretic notion of echo, taken as a cognitive mechanism rather than just as a pragmatic phenomenon. Irony arises from the clash between an echoed and an observed scenario. The construction of the former is constrained by socio-cultural, communicative, and personal factors. This view allows for a classification of ironist and interpreter types and of echoed scenarios from the standpoint of their grounding in an array of personal and presumed interpersonal beliefs, and in socio-cultural stereotypes. It also allows for a correlation between irony types and echoed scenario types. The resulting account achieves comprehensiveness while keeping the explanatory apparatus simple by postulating a reduced set of cognitive mechanisms.

References

Barnden J. (forthcoming) "Irony, pretence and fictively-elaborating hyperbole" in Colston H., Athanasiadou A. (eds.) *Irony in language use and communication,* Amsterdam/Philadelphia, John Benjamins.

Booth W. C. (1974) *A rhetoric of irony*, Chicago & London, The University of Chicago Press.

Clark H., Gerrig R. (1984) "On the Pretence Theory of irony", in *Journal of experimental psychology*, 1, 121-126.

Hutcheon L. (1994) *Irony's edge*, London & New York, Routledge.

Muecke D. C. (1970) *Irony and the ironic: the critical idiom*, London & New York, Methuen.

Ruiz de Mendoza F. (forthcoming) "Cognitive modeling and irony", in Colston H., Athanasiadou A. (eds.) *Irony in Language Use and Communication,* Amsterda T/Philadelphia: John Benjamins.

Wilson D., Sperber D. (2012) "Explaining irony", in Wilson D., Sperber D. (eds.), *Meaning and relevance*, Cambridge, Cambridge University Press, 123-145.

Ersilia RUSSO (La Sapienza University of Rome, Italy; ersilia.russo1@gmail.com)
Francesca MAROCCO (La Sapienza University of Rome, Italy; francesca.marocco@libero.it)
Teresa GARGANO (La Sapienza University of Rome, Italy; teresagargano_91@hotmail.it)

"PhiloEditor: an interactive study platform for analytical genetic edition"

PhiloEditor is a web application that automatically detects the variants of two or multiple draft of a text, providing a Digital Critical Edition of it, which highlights both philological and exegetic aspects of the various versions of the text (Di Iorio, Italia, Vitali 2014). This digital resource is not only a philological-linguistic analytical tool, but also a markup tool, which makes it possible both to study texts synchronically and diachronically and to highlight their qualitative and quantitative features – visually displaying them in pie charts and diagrams – within an interactive infrastructure useful for both research and pedagogical purposes.

The paper will show how PhiloEditor has been tested on two case studies: Manzoni's *Promessi Sposi* – compared in its first and second version – and Collodi's *Le avventure di Pinocchio* – compared in its book and magazine version – so as to reveal their micro-linguistic and stylistic features (e.g. linguistic reduction, tuscanization, graphical and punctuation mark variations) but also narrative patterns and character portrayals.

Reference
Di Iorio A., Italia P., Vitali F., "Variants and Versioning between Textual Bibliography and Computer Science" (paper presented *at AIUCD '14*, Bologna, September 18-19, 2014).

Iride SANTORO (University for Foreigners of Siena, Italy; s.iride88@live.it)
Liana TRONCI (University for Foreigners of Siena; tronci@unistrasi.it)

"Les temps verbaux dans *L'Étranger* de Camus: analyse distributionnelle et valeurs narratives"

Cet exposé propose l'analyse de la distribution des temps verbaux dans le roman *L'Étranger* de Camus (1942). À côté du passé composé, qui est le temps narratif principal du roman, d'autres temps sont aussi employés, à savoir le présent, l'imparfait, le plus-que-parfait, le passé simple, le futur. Ils marquent parfois des oppositions fonctionnelles par rapport au passé composé, voire narration (passé composé) vs description (imparfait) ou énoncé (passé composé) vs énonciation (présent), mais ils peuvent aussi remplacer le passé composé dans certains cas. La base de données est constituée par la totalité des occurrences des temps verbaux à l'indicatif dans le roman. Dans l'analyse nous avons suivi une approche distributionnelle et fonctionnelle, en faisant aussi recours aux deux traductions du roman en italien. Les résultats que l'on s'attend sont les suivants: (1) fournir le recueil complet des occurrences des formes verbales dans le roman de Camus; (2) proposer une analyse de ces occurrences en termes distributionnels et fonctionnels, même par rapport aux traductions italiennes du roman; (3) essayer de comprendre le rapport entre l'usage de certaines formes verbales et la constitution du style de l'auteur.

Référence
Camus A. (1942) *L'étranger: roman*, Paris, Gallimard.

Daniel C. STRACK (University of Kitakyushu, Japan; dcstrack@kitakyu-u.ac.jp)

**"Pinnacles in long-form literary texts:
cross-textual evidence for the pervasiveness of megametaphorical expression"**

Various scholars including Werth (1994) and Stockwell (2002) have recognized how literary authors occasionally use megametaphor (i.e. scattered metaphorical domain references) to unobtrusively shape their narratives. Corpus analysis proves useful not only in identifying the presence of such metaphorical systems in a text but also in assessing their relative prominence across multiple texts. This presentation will examine 50 randomly selected literary texts so as to discern the extent to which the relatively uncommon but metaphorically replete word *pinnacle* contributes surreptitious metaphorical meaning to works in which it is found. Results of the study confirm that, when used, the word *pinnacle* serves to highlight climactic scenes or emphasize key turning points in the development of the protagonist's character. The fact that metaphorical lines of interpretation relating to certain words and phrases may be detected not only by way of critical intuition but also through the electronic searching of multiple text corpora demonstrates the value of cross-textual analysis strategies in certain cases. It also hints that megametaphor, rather than being a rare and idiosyncratic type of literary artifice, may actually be more prevalent than has been previously acknowledged.

References
Stockwell P. (2002) *Cognitive Poetics: An Introduction*, Abingdon, Routledge.
Werth P. (1994) "Extended metaphor—A text-world account", in *Language and Literature* 3(2), 79-103.

Gaia TOMAZZOLI ("Ca' Foscari" University of Venice, Italy; gaia.tomazzoli@unive.it)

**"'Totus poema eius ubique mirabiliter figuratus':
identifying, classifying and describing Dante's metaphors"**

This paper presents how metaphors in Dante's *Commedia* can be identified, classified and organized in a database as to provide thorough and solid data for their understanding. The metaphors in the poem are first identified through a linguistic method called *MIP* (i.e. *Metaphor Identification Process*; Steen et al. 2010), based on the comparison between the contextual and the basic meaning of each word-unit. Next, the metaphors are classified along syntactic, semantic and structural and semantic, taking into account both Medieval explanations of figurative language and the features of Dante's metaphors that have always struck the readers of his poem. The findings from this combined linguistic and stylistic analysis provide evidence in support of tenets put forward by some literary scholars (e.g. that metaphors increase in number and complexity throughout the poem) and show that metaphors converge in sections of political and religious invective, where the main purpose of the author is to raise his style. The study suggests that a systematic multi-dimensional analysis of metaphor can shed light on its conceptual import in a text and raise awareness of all its stylistic traits.

References
Steen G. et al. (2010) *A method for Linguistic Metaphor Identification: from MIP to MIPVU*, Amsterdam & Philadelphia, John Benjamins.

Federica VEZZANI (University of Padua, Italy; Federica.vezzani@phd.unipd.it)
Giorgio Maria DI NUNZIO (University of Padua, Italy; giorgiomaria.dinunzio@unipd.it)
Geneviève HENROT (University of Padua, Italy; genevieve.henrot@unipd.it)

"(Not so) Elementary, my dear Watson! A different perspective on medical terminology"

This contribution aims to provide an overview of the syntactic and semantic behavior of medical terms in the literary works of Conan Doyle. It analyses the scientific terms in the stories of Sherlock Holmes through the model for terminological records set out in a multilingual terminological database (TriMED; Vezzani, Di Nunzio, Henrot forthcoming) implemented for the linguistic analysis of technical medical terms. The terms are semi-automatically extracted and then tabulated for such aspects as: their semantic variation due to temporal and historical factors, different contexts of use, different reference corpora, association with different registers, and different lexico-syntactic collocations/colligations. This type of analysis offers a thorough, in-depth view of the semantic, syntactic and stylistic aspects of the medical terms considered, and suggests avenues for further terminological research in a bilingual comparative-contrastive perspective.

Reference

Vezzani F., Di Nunzio G.M., Henrot G. (forthcoming) "TriMED: A Multilingual Terminological Database", paper to be presented at *the 11th Language Resources and Evaluation Conference (LREC)*, 7-12 May 2018, Miyazaky, Japan.

WORK-IN-PROGRESS REPORTS

Francesca BIANCHI (University of Salento, Italy; francesca.bianchi@unisalento.it)

"The adaptation of a novel to filmic needs:
thematic differences in the subtitles of two filmic versions of *Pride and Prejudice*"

The present work compares two filmic adaptations of Jane Austen's *Pride and Prejudice* by extracting and analyzing key-domains compared to the BNC. The analyses showed that the dialogues in the two filmic versions reflect all the core themes of the novel and share a set of common themes that are relevant to the plot. However, both quantitative and qualitative differences were observed. The observed differences suggest that different weight is given to shared themes in each version. The results also suggest that the automatic extraction of key domains and their manual analysis could be a viable method to provide evidence in support of insightful, but impressionistic, interpretations of films, thus making replication studies possible. The method also allows researchers to describe alternative views on the "same" story, both quantitatively and qualitatively. Finally, compared to subjective reading, this method apparently succeeds in bringing to the fore a wider range of themes and better highlighting the richness of the topics discussed.

Francesca COCCETTA ("Ca Foscari" University of Venice, Italy; francesca.coccetta@unive.it)
Daniela CESIRI ("Ca' Foscari" University of Venice, Italy; daniela.cesiri@unive.it)

"Narrative descriptions in Meyer's *Twilight Saga*: a corpus stylistics perspective"

The present study draws on corpus stylistics (cf. Semino, Short 2004; Hoover 2007; Mahlberg 2014; Hoover, Culpeper, O'Halloran 2014, to name a few) to investigate the narrative style in Stephenie Meyer's *The Twilight Saga* (2005-2008). In particular, it focuses on keywords generated using *Wordsmith Tools version 7* (Scott 2017), and the BNC as reference corpus. Qualitative and quantitative analyses show that the most frequent lexical words (nouns, adjectives, and verbs) reflect the writer's focus on carefully selected physical elements of the characters. More specifically, the nouns preferentially denote specific physical attributes of the characters, the adjectives provide a general description of the protagonists and the setting, while the verbs used present the actions of the characters, their interaction with the other protagonists, and their position in the narrative situation.

References

Hoover D. L. (2007) "Corpus stylistics, stylometry, and the styles of Henry James", in *Style* 41(2), 174-203.

Hoover D. L., Culpeper J., O'Halloran K. (2014) *Digital literary studies: corpus approaches to poetry, prose, and drama*, London & New York, Routledge.

Mahlberg M. (2014) "Corpus Stylistics", in Burke M. (ed.) *The Routledge handbook of stylistics*, London/ New York, Routledge, 378-392.

Meyer S. (2005) *Twilight*, London, Atom.

Meyer S. (2006) *New Moon*, London, Atom.

Meyer S. (2007) *Eclipse*, London, Atom.

Meyer S. (2008) *Breaking Dawn*, London, Atom.

Scott M. (2017) *Wordsmith Tools Version 7. Stroud: Lexical Analysis Software.*

Semino E., Short M. (2004) *Corpus stylistics. Speech, writing and thought presentation in a corpus of English writing*. London, Routledge.

Marina GORLACH (Metropolitan State University of Denver, USA; gorlach@msudenver.edu)

**"The qualitative-quantitative analysis of *Fahrenheit 451*:
mapping the linguistic make-up of literary texts"**

This paper discusses the implications and applications of a mixed qualitative-quantitative analysis of a literary text through the consideration of the role of word systems in conveying the message of the text. The analysis is based on a semiotic theoretical and methodological approach, which treats text as a multilayered and multidimensional structure, where all semantic, lexical, morphological, syntactic, and phonological systems are interrelated and act in conjunction. The linguistic and extralinguistic signs used in the novel create a robust effect due to the synergetic intensification and amplification of the message through the recurrence of interconnected words forming quantifiable, close-knit systems. The text in focus is Ray Bradbury's (2012) *Fahrenheit 451°* – a lyrical anti-utopia portraying the massive attack of "consumer civilization" standards on the traditional cultural values of society. The central message of the text can be interpreted as a critical view of state-sponsored censorship, or, according to Bradbury's 2007 interview (Weller et al. 2010), as the effects of television and mass media on the reading of literature and alienation of people. This idea is conveyed via an array of word systems based on the marked distribution of phonological forms, grammatical forms, semantic conceptual fields, metaphoric systems, and communicative strategies. The profound ideological effect of Bradbury's work is achieved through the author's sophisticated use of multiple word systems at all levels of language structure.

References
Bradbury R. (2012 [1964 [1953]]) *Fahrenheit 451°*, New York, Simon & Schuster.
Weller S., Bradbury R., Francis B. (2010) *Listen to the echoes: the Ray Bradbury interviews,*
    Chicago, IL, Stop Smiling Books.

Hongxia JIA (Beijing Information Science and Technology University, P. R. China;
hongxiajia@126.com)
Xitian YANG (Beijing Institute of Technology, P. R. China; 452854057@qq.com)

**"Quality & quantity:
a mixed-method approach to legal concepts in Charles Dickens' *Bleak House*"**

This paper adopts a mixed-method approach to the study of Charles Dickens' *Bleak House*, examining qualitative and quantitative aspects of the novel. Its goals are to describe the verbal representation of legal institutions (e.g. court and prison) and the characters involved in them (e.g. the judge, the lawyer and the criminal), as well as to test qualitative interpretations of and hypotheses about legal concepts contained in the novel, as put forward by previous scholars on the basis of subjective impressions. Through a comparison of *Bleak House* against other novels by Dickens as well as other literary works of the same period, it is possible to offer evidence in support of Holdsworth's (1928) view of Charles Dickens as a Legal Historian.

References
Dickens C. (2003) *Bleak House*, London, Penguin Books.
Holdsworth W. S. (1928) *Charles Dickens as a legal historian*, New Haven, Yale University Press.

Christos STAVROU (Technological Educational Institute of Epirus, Greece; cstavrou@teiep.gr)
Anna CHITA (Technological Educational Institute of Epirus; Greece; chita@teiep.gr)

"The metaphor in literature and the effect on translation"

Metaphor has been widely discussed within the field of translation, mainly with respect to transfer methods, including such renderings as substitution, paraphrase and deletion. This paper focuses on the translation of metaphor as a cultural concept. Based on the theory of Newmark (1982), who distinguished six types of metaphors – dead, cliché, stock, adapted, recent and original – the present paper focuses on the translation of metaphor as a cultural concept by analysing Oscar Wilde's *The portrait of Dorian Gray*. Through a contrastive analysis of its Greek and German translations, and following Dagut (1976: 32), we explore the particular cultural experiences and semantic associations exploited by translation and the extent to which these can, or cannot, be produced non-anomalously into the target language, depending on the degree of overlap in each particular case, so as to determine the translatability of a metaphor. In line with Snell-Hornby (1995: 41), the findings suggest the extent to which a text is translatable varies with the degree to which it is embedded in its own specific culture.

References
Dagut M.B. (1976) "Can 'metaphor be translated?", in *Babel* XII, 21-33.
Newmark P. (1982) Approaches to translation, Oxford & New York, Pergamon Press
Snell-Hornby M. (1995) Translation studies: an Integrated approach (revised edition), Amsterdam & Philadelphia, John Benjamins.

Rosemary WESTON (Eastern Michigan University, USA; rweston3@emich.edu)
Francesco D'INTRONO (University of Massachusetts & Michigan State University, USA; franco@spanport.umass.edu)

"Quantitative and qualitative analyses in literature: a case study"

A novel can be seen as a microcosmic representation of a society (we call it the *literary society*), in which characters interact by establishing relationships that define their allegiances and conflicts, their feelings towards each other, such as alienation, love, etc. In this paper we argue that some literary works, in particular novels, can be studied from a quantitative point of view, by considering specific socio-linguistic variables (Labov 1972) of the text, which unveil the author's interpretation of that society, and to link it to the *intertextual society* present in the writings of other authors. We examine a novel by Argentinian writer Julio Cortázar (1980), *Los Premios*, from a socio-linguistic point of view, by considering certain linguistic variables, such as personal pronouns, and the communicative interactions among the characters of the novel. The analysis shows that that Cortázar identifies himself with hero of the novel because of similarities, physical and intellectual, between him and the hero. The findings also suggest that the minimal social structure at the base of the *literary society* can be found in an *intertextual society*, that is, that it is present in the work of other authors.

References
Cortázar J. (1980) *Los premios*, Barcelona, Bruguera.
Labov W. (1972) *Sociolinguistic Patterns*, Philadelphia, University of Pennsylvania Press.

Tobia ZANON (University of Padua, Italy; tobia.zanon@unipd.it)
Sara GIOVINE (University of Padua, Italy; sara.giovine@unipd.it)

**"The TRALYT research project:
translation and lyrical tradition between Italy and France (19th-21st Century)"**

This work-in-progress report presents the research project: *Translation and Lyrical Tradition between Italy and France (19th-21st Century)* (SIR 2014 RBSI14URLE). This project focuses on digital mapping and critical discussion of the relationships between French (and Francophone) modern lyrical poetry and contemporary Italian (and Swiss-Italian) lyrics. The goals of the project are: 1) to enlarge the body of translations currently available and their bibliographies; 2) to analyse the forms and functions of translation in the two languages and the two literatures; 3) to better understand the practice of translation and its topological effects (i.e. its chrono-topological function and aesthetic function); 4) to draw meta-literary implications on socio-literary and socio-cultural aspects of literary transfer; 5) and to evaluate the reception of an author in a foreign country.

Dandan ZHANG (Qiqihar University, P. R. China; zhangdan1223@163.com)

"A corpus-based comparison of six Chinese translations of the Japanese novel *The Tale of Genji*"

The Tale of Genji (or *Genji Monogatari*), a Japanese novel written in the early 1000s, has been likened to *Hong Lou Meng* (or *Dream of the Red Chamber*) of the Tang Dynsty (618-907) of China, and, owing to its supreme achievements in the monogatari genre, has been deemed as the world's first modern novel. The tale comes in 54 chapters with nearly one million words in *kana* and _tanka_ (archaic Japanese phonetic script and poetry). Ever since Taiwanese translator Lin (1978) and Mainland Chinese translator Feng (1980) published their respective complete Chinese translations, nearly a dozen new versions or "retranslations" have sprung up in the past three decades. But some studies come up with claims or even open outcries over the poor quality of the later translations, and even the "over-reference" to or "plagiarizing" of their predecessors, especially Yin (1996), Liang (2003), and Yao (2006). The current study sets out to make a corpus-based comparison of six Chinese translations of *The Tale*. The parameters to be examined are threefold: lexical, including Type-to-Token-Ratio, density, as well as the occurrences of some remarkable types of lexis (e.g. reporting verbs, conjunctions, modal particles, and "love-hatred" expressions); average sentence length; and techniques whereby the _tanka_ verses are dealt with. The study intends to address two issues, namely, what linguistic and textual features characterise the style of the six translations, and to what extent they are similar to one other.

References
Liang (2003) PLEASE COMPLETE!
Lin (1978) PLEASE COMPLETE!
Yao (2006) PLEASE COMPLETE!
Yin (1996) PLEASE COMPLETE!