Life Beyond Dictionaries

Life Beyond Dictionaries

Edited by

Olga M. Karpova
and Faina I. Kartashkova

Cambridge
Scholars
Publishing

Life Beyond Dictionaries

Edited by Olga M. Karpova and Faina I. Kartashkova

This book first published 2015

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-4438-7794-8
ISBN (13): 978-1-4438-7794-7

TABLE OF CONTENTS

Part II: Tourism and Heritage Dictionaries with Special Reference to Culture

Part III: Projects of New Dictionaries

INTRODUCTION[1]

Ivanovo State University has always paid special attention to the development of friendly relations with universities and non-governmental institutions from other countries which contributes to general research, teaching and educational activities carried at the establishment.

The fruitful contacts with the Fondazione Romualdo Del Bianco®—Life Beyond Tourism® (Florence, Italy) started in 2004 with the signing of an agreement of cooperation in the field of cultural and scientific communication. Over the years, this resulted in more than fifteen conferences, seminars, workshops and meetings in both Ivanovo and Florence.

Our scholars had the chance to become not only partners but also the first-hand developers and prime movers of several projects carried out by the Foundation. With the foundation's support we involved over two hundred and fifty teachers and students from a dozen universities in Russia and twelve UN countries (e.g. Great Britain, Germany, Spain), Asia and the USA. The cooperation with the foundation allowed the University to find new partners at the University of Florence and Accademia della Crusca in Florence.

Professors from Ivanovo State University worked as international experts of the foundation, invited by the President of the Foundation, Mr. Paolo Del Bianco, to hold round table discussions in 2008, 2010 and 2011. This led to the development of further cooperation and the signing of the Memorandum of Understanding between the foundation and the city of Ivanovo. During these visits, alongside fruitful meetings, we participated in numerous cultural events and workshops together with members of the foundation's governing board.

The University Academic Council awarded Mr. Paolo Del Bianco the title of Professor Emeritus for the great contribution of the foundation in the integration of our university into the European educational environment. In turn, the foundation board awarded Ivanovo State University with a Medal of Honour for the effective cooperation.

---

[1] First published in *Life Beyond Dictionaries*. Proceedings of X Anniversary International School on Lexicography. Ivanovo, Florence, September 12–14, 2013.

Six international student workshops organized by the Vice-Rector for PR, and Head of the English Department, Professor Olga Karpova in 2008–2015 deserve special mention. The project suggested by Professor Karpova is devoted to the compilation of a new encyclopaedic dictionary entitled *A Dictionary without Boundaries: Florence in the Works of World Famous People. Project of a Dictionary for Guides and Tourists*. This long-term project is devoted to the Italian, particularly Florentine, influence on different national cultures. All the participants benefit from visiting Florence and exchanging opinions and ideas in international students groups, thus spreading the general principle of intercultural dialogue. They learn to work, communicate and live together in the international community

The workshop on September 8–15, 2013 was of a special kind because it was part of the X Anniversary International School on Lexicography *Life Beyond Dictionaries.* Introduced in 1995, the Ivanovo School on Lexicography is an outstanding international event that biennially attracts more than three hundred participants from all the continents (except Antarctica). This autumn, the Ivanovo School is moving closer to the world academic community by extending its geographic reach and joining researchers from 17 countries in Ivanovo and Florence.

I wish fruitful and interesting work for all the participants during the conference and I am looking forward to further cooperation with Fondazione Romualdo Del Bianco®—Life Beyond Tourism® in many respects.

Vladimir Egorov,
Rector of Ivanovo State University,
Doctor, Professor

It is common knowledge that the Fondazione's activity is designed to foster intercultural dialogue. In this connection, it is crucial for individual countries and individual regions to draw closer to their tangible and intangible cultural heritages, and above all in their impact on the daily lives of their residents. Travel can introduce people to knowledge only if it is conceived as an opportunity for meeting and trading experience and knowledge, rather than as the mere use and enjoyment of consumer-driven services. Fostering mutual knowledge among people from different countries through aspects of their daily lives is the basic aim of the Life Beyond Tourism Cultural Portal, a virtual forum where cultures meet to promote dialogue.

The work that Ivanovo State University is doing with its X International School on Lexicography *Life Beyond Dictionaries* fully reflects the spirit of the Fondazione and the Life Beyond Tourism Portal-Heritage Community, and it is a major contribution to furthering better intercultural knowledge and understanding.

The Fondazione Romualdo Del Bianco®—Life Beyond Tourism® is particularly grateful to Ivanovo State University and the Organizing Committee as a whole for this ongoing and increasingly valuable commitment, which encourages the development of a sensibility in young people helping to make them feel increasingly involved and responsible. This is crucial if we are to ensure harmonious development based on a determined will to understand our fellow human on a planet whose population is drawing close to the 10 billion mark.

Paolo Del Bianco,
President of Fondazione Romualdo Del Bianco®—Life Beyond Tourism®, Florence, Italy

PART I:

LEXICOGRAPHY WORLDWIDE— HISTORICAL AND MODERN PERSPECTIVE

CHAPTER ONE

MANY SHADES OF EUROPE: APPLICATIONS OF COLLOCATIONAL NETWORKS AND RESONANCE

GEOFFREY WILLIAMS
LICORN, EUROPEAN UNIVERSITY OF BRITTANY, LORIENT, FRANCE

J. R. Firth's adage that "you shall know a word by the company it keeps" is much cited, but the full implications of what he says are not always understood. Collocation is too often seen as something static, just a relationship between two words, but J.R. Firth points to something much more than that as studies from M. Phillips (1985) and G. Williams (1998; 2002) have amply shown. The big problem with collocation is that it has always been with us, but we simply did not know it. It has taken big corpora and the insights of J. McH. Sinclair (1996) to show how important the idiom principle is. What is more, there is often a strange confusion between collocation studies and the phenomenon itself. The former has been with us for some 80 years since the publication of H. E. Palmer's second report (Palmer 1933), while the latter is part of language and has thus existed as long as the human speech faculty, and maybe longer.

This paper does not seek to give a history of collocation, or even to discuss functional collocation—that is to say the phraseologically restricted forms generally entered in dictionaries. What this text does seek to do is to bring together two threads of research in contextualist collocation, collocational networks and collocational resonance, as well as with the technique of lexicographical prototypes to show their applications in solving some of the problems facing lexicography in general and e-lexicography in particular.

The text will briefly deal with some problems facing lexicography and then show how the techniques of collocational networks, collocational resonance and lexicographical prototypes have evolved. It then applies this to a case study of political representations of Europe.

The Lexicographer's Conundrum

Lexicography is about finding appropriate solutions to age-old problems of mediating between words found in context and words taken out of context. This point is often overlooked and needs underlining. Medieval glosses were simply written into manuscripts, thus demonstrating what that orthographic word could mean in that context. From the time these were collected into glossaries to the current world of dictionaries we have decontextualized meanings and constantly seek means to better recontextualize them through definitions, citations and examples. The move from gloss to glossary also brings about another problem—word selection, and its corollary word deselection. Medieval glossaries were anything but systematic. They collected words that had proven ambiguous, difficult or controversial and grouped them into a single volume. It is thus no wonder that the first European monolingual dictionaries such as R. Cawdrey's 1604 *Table Alphabetical* were dictionaries of so-called hard words. It is only with works such as Furetière's monumental *Dictionnaire Universel* of 1690 that we get dictionaries that attempt to cover all the words of a language. Even when we get more complete works, the question of what goes in remains an issue. Mentioning R. Cawdrey also highlights another issue—that of the alphabet.

Many definitions of the word dictionary declare these works to be alphabetical by nature. If we accept that the first dictionaries appeared in China and India, this is an amazing piece of Eurocentrism. These languages did and do not benefit from the Latin alphabet, nor do the great majority of the world's languages. It is true that, from R. Cawdrey on, European dictionaries did adopt the Latin alphabet and this has become synonymous for many with the word dictionary, but all lexicographers are aware that whilst this may be convenient in some ways, alphabetical ordering has many drawbacks, the first of which is the dispersion of conceptually related words.

The move from glossary to catholicons and more global collections of wordings also led to other problems. Some words have multiple meanings, thus leading to a necessity to differentiate homonyms and polysemes. Whilst the former may be relatively straightforward in synchronic works, a diachronic analysis can lead to more complex relationships. Polysemy is more complex as the lexicography must decide on the degree of granularity in describing senses, or better meaning potentials as P. Hanks (1987) would define dictionary contents. Monolingual polysemy is difficult enough, but if we consider the radical differences of meaning between languages, above all connotative meaning, the problem deepens.

A lexicographer must face many more problems, but those introduced above provide sufficient matter to demonstrate the organising power of collocation.

Collocational Networks

Collocational networks were first posited in 1996 with the methodology published in a major article in the *International Journal of Corpus Linguistics* in 1998 and with a more advanced application in 2002 (Williams 1998, 2002). Since this date the methodology has been applied in a number of projects by the present author and many others, notably C. Magnusson and H. Vanharanta (2003), and in at least two very innovative theses in terminology extraction (Alonso 2009) and data mining (Magnusson 2010). Other formats have also taken up the idea, such as so-called clouds and constellations that are collocational networks in essence, although they rarely acknowledge the fact. So where do networks come from?

The inspiration is obviously J.R. Firth's notion of "the company words keep" (Firth 1957, 11). Firth opened the door for collocation to move from simple phraseological co-occurrence to dynamic collocation. If J.R. Firth opened the theoretical door, then J.McH. Sinclair took us into the real world of real language by showing the intimate connection between corpora, concordances and collocation through the idiom principle (Sinclair 1991). J.McH. Sinclair's work in Edinburgh, and later Birmingham, opened the way to corpus studies that showed how collocation is at the centre of all language.

Computers were at this time slow and difficult to access as priorities were elsewhere. Nevertheless, J.McH. Sinclair managed to create corpus linguistics and the entire terminology that is now taken for granted (Sinclair et al. 2004). Numerous studies followed, particularly after the move to Birmingham and the beginning of the COBUILD project.

Amongst the studies developed in Birmingham, that of M. Phillips (1985) tackled meaning in text structure by the construction of local networks. The present author took up this reasoning, realising the value of mimicking Kohenen neural networks on texts and hence collocational networks, described as being: "… a web of interlocking conceptual clusters realised in the form of words linked through the process of collocation" (Williams 1998, 156), in which: "… a web of interlocking conceptual clusters [is] realised in the form of words linked through the process of collocation" (Williams 1998, 156). In practice, what this means is that links are established through statistical co-occurrences which lead

to interdependent chains of collocates moving out from a central node. The relationships between individual items lead to an inevitable clustering of related wordings which thereby conceptualize a given thematic trend. There is not, and cannot be, one set statistical measure as the calculation depends on what is being sought. This is similar in essence to the refusal of any notion of a definitive definition of collocation itself. Collocation is a phenomenon of language, and any definition must take into account the application with dictionary usage requiring a more formalised definition, and one for natural language processing that is even further restricted. These restrictions are only for convenience, however, and not a reality as such. The two initial measures were mutual information and Z-score, the former being mostly applicable to scientific fields where terminological groupings are sought, as opposed to more general language where Z-score is adapted more. The methodology remains the same—what counts is the application and the purposes to be achieved. In all cases, the starting point is a single lexical unit or multiword unit. The network is then built outwards from this, as in Figs 1.1 and 1.2 below (Williams 1998) showing the networks for DNA.

What Figs 1 and 2 demonstrate is the natural thematic groupings within a specialized corpus, in this case of the biology of parasitic plants. In a dictionary, this analysis allows for the selections of headwords and explorations of the relations between individual lexical items forming both complex terms as well as the semantic environment of these terms. At another level, the network shows the natural relations that can be exploited when navigating between items in an electronic dictionary.

Fig. 1.1. Immediate collocates for “DNA”

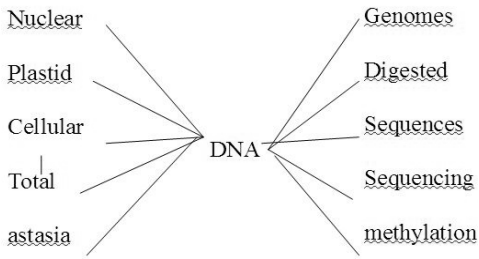
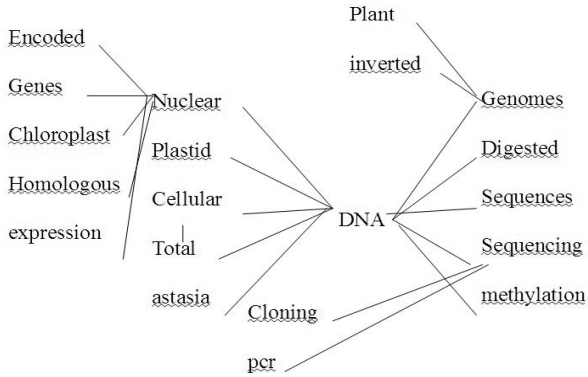


Fig. 1.2. Widening Network for "DNA"



This particular network was built using mutual information as the interest was in rarer and more terminological words. These networks were built on non-lemmatised corpora, and many still are, so as to avoid the influence of the artificial part of speech categories. However, again depending on the situation, POS tagging can be extremely useful and this is what the Dictionary of Science Verbs project, that uses the specialized *BioMed Corpus*, has done. In this experimental dictionary, networks are used both for the extraction of headwords and internal navigation within the dictionary. Lemmatised networks are then reduced to handle different features in the dictionary, as can be seen with the network for the verb *show* (see Fig. 1.3 below).

Fig. 1.3. Noun-verb groupings for *Show*

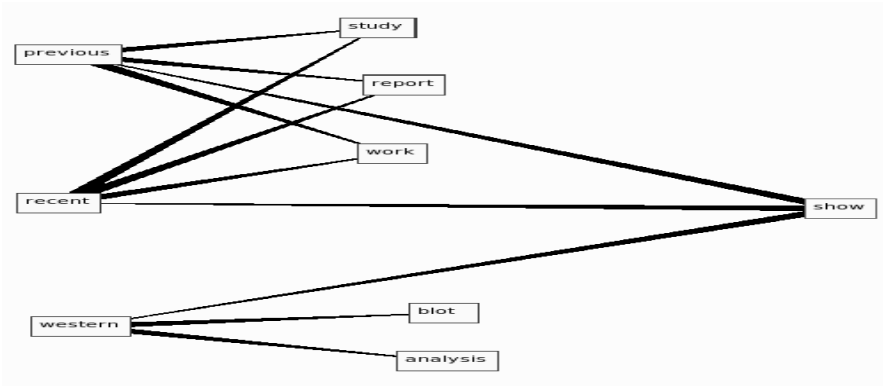
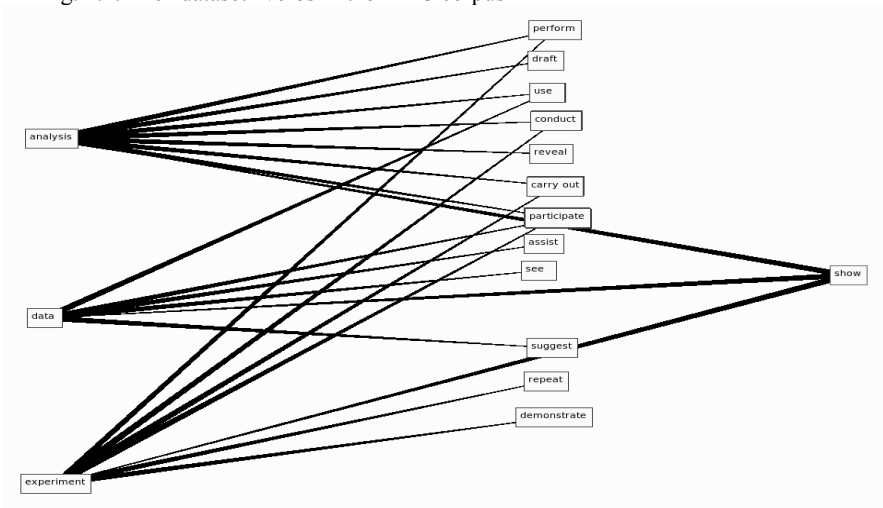


Fig. 1.3 above concerns part of a wider network for the lemma *show* explored in the BioMed Corpus (Williams and Millon 2010). In this case, lemmatisation has allowed us to isolate two different types of groupings: one citing other research and another showing a process to illustrate outcomes. Phraseological patterns can be extracted from both that can show a non-native speaker of English how to use this verb when writing research papers. A far more complex network for the same verb (see Fig. 1.4 below) shows how collocates can naturally group other verbs.

Fig. 1.4. The "dataset" verbs in the BMC corpus



Here, three nouns acting as subjects for the verb show are linked to other verbs that they share. The grouping shows potential synonyms for show as well as other verbal collocates of three words concerning the analytical part of scientific results.

In all cases, the methodology is iterative with a move from the building of a network from a given lexical unit followed by the identification of potential groupings and the extraction and typifying of new networks arising from these groupings. This process is then repeated as the semantic relationships are explored by the linguist/lexicographer. In the cases here, the aim is to build networks that can help select headwords, allow natural conceptual groupings and make easier navigation possible between words and concepts within an electronic dictionary. The alphabet is still there, but a new search method allows for jumping between related word-forms.

The full automation of the process is a possibility, but also a danger.

A network on its own shows only links that require interpretation and it is often better to do this during the building process. As Sinclair so aptly pointed out, we must "trust the text" (2004), which means never losing sight of the contexts that create the meanings we seek to illustrate.

Collocational networks are thus the first of the collocation-related tools at our disposal. The second, which leads on from the first, is that of collocational resonance.

Collocational Resonance

Resonance starts from observing the obvious. This means that whilst synchronic and diachronic analyses of word usage have their uses, the division is purely a convenience and the past is always in the present. In reality the so-called metaphor must be live to be the real metaphor, and once the exploitation (Hanks 2013) has become a norm it is the meaning of the word and no longer figurative. In this way, language changes over time. However, it is equally obvious that past uses of a word influence current uses. As exploitations become norms through association with new wordings, the half-lives of their atomic past remain partially active. Etymological meanings may continue to survive, but generally they do not and are only artificially revived by pedants eager to show what the word *really* means, when the reality is that it means what it means in context and nowhere else.

Thus, rather than falling into the speculative trap of so-called cognitive metaphors with their suppositions about some underlying strata of meaning, resonance seeks to use the power of collocational networks to demonstrate what meaning patterns are displayed in the corpus of earlier usage and how these map onto the present so that a collocational remanence can unconsciously colour meanings. Thus, resonance posits that language users carry aspects of meaning from previously encountered usage, consciously and subconsciously, colouring the meanings and prosodies in use. In such cases, etymological aspects are carried over only if the user is aware of past usage or if this is still active. No distinction is made between the so-called literal and figurative meanings and metaphor is considered to occur only if the user is conscious of the fact.

Collocational resonance first emerged at the Phraseology 2005 conference at Louvain-la-Neuve (Granger and Meunier 2008). The strange thing is that two papers had similar titles—one by the author (Williams 2008) and another by P. Hanks (unpublished). The former took an intertextual view while the latter was concerned more with metaphor, but both were clearly drawing lessons from the teachings of J. Sinclair and

dealing with the same phenomenon.

The first publication of G. Williams (2008) concerned the use of religious terms found in an electronic version of the King James Bible and the works of Shakespeare compared to usage found in the British National Corpus. A gradual loss of the sacred aspect of many wordings as they simply become idiomatic expressions of annoyance or surprise become clear, such as "Jesus wept" or "My God." This is part of a gradual secularization of religious language in which users are generally unaware of the origins of expressions. This subconscious use of wordings is highly relevant to resonance studies and can be illustrated by a citation from Boswell's *Life of Johnson*, in which we read that:

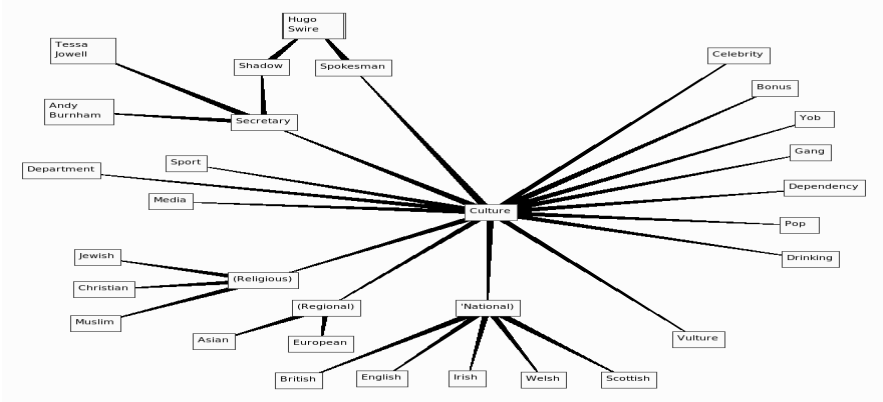
> Mr Allen, the printer, brought a book on agriculture, which was printed, and was soon to be published. It was a very strange performance, the author having mixed in it his own thoughts upon various topics, along with his remarks on ploughing, sowing, and other farming operations … he had this very curious reflection: "I was born in the wilds of Christianity, and the briars and thorns still hang about me."

Mr Allen was obviously suffering from collocational remanence with his denied cultural background showing through involuntarily in his language. He is far from alone in this.

Other studies look at the metaphorical exploitations of words such as *probe* (Williams and Millon 2009), *culture* (Williams 2011a) and *art* (Williams 2011b). In the case of *probe*, we have a noun that refers to a basic medical exploratory instrument to a whole range of investigative instruments and features. In this case the active exploratory features of the instrument and its use are still clearly active (see later for the lexical prototype), whereas in other cases these are far less clear.

Such is the case with *culture* where the usage has moved from cultivation of soil (tillage) to cultivation of the mind. Thus, *culture* can be mapped as a move from a purely agricultural usage to a more metaphorical approach to matters of the cultivated intellect to the current situation. With the English, farming has almost entirely disappeared leaving reference to intellectual arts, societal religious and regional expressions of identity through traditions, as well as an evolving tendency to collocate *culture* with antisocial behaviour (see Fig 1.5 below).

Fig. 1.5. Metaphorical exploitation of the word *culture*



Lexicographical Prototypes

The tool used to demonstrate changing meaning patterns is P. Hanks's (1994; 2000) notion of lexicographical prototypes, as these can allow us to clearly see variations in use. In his celebrated analysis of the noun *bank*, P. Hanks showed how polysemy can be handled through a series of short propositions that illustrate aspects of a word's use and are seen to be activated with polysemy as a continuum rather than discrete items. A good illustration is that as for *probe*, developed by G. Williams and C. Millon (2009). The initial network clearly possesses four key elements: the words "instrument," "long," "blunt" and "investigate," all adequately describing something used to explore and clean wounds. Later uses see a bird's beak described as a probe, hence the verbal form of "to probe" meaning to investigate with a long thin instrument. This does not refer to any bird, but specifically the waders which have long thin beaks with which they seek food in water. From being an exploratory instrument in the real sense to a figurative one takes us to *probe* as a judicial investigation. The prototypes below with only certain features being activated in any one instance can handle these variations:

A precision instrument
Long and thin
Not cutting (blunt)
Very pointed (sharp)
Used to explore and examine

Can penetrate an organism or cell
Does not damage the organism
Means of investigation (ext. by judicial authorities)
Supplies Information

Prototypes are based on L. Wittgenstein's idea of family resemblances and are therefore open-ended classifications. This is important because the prototype can be altered or tweaked as meanings and wordings evolve (Williams 2006).

Initially created to handle polysemy in the dictionary, Williams has linked them to collocational resonance as the prime tool to handle meaning variation. This has meant that prototypes can handle diachronic variation as well as changes across different thematic areas of usage and, most importantly, differences in meaning across languages (Williams et al. 2012). If the propositions are simple then they are translatable and can illustrate changes in meaning that simple translation equivalence will just ignore. In all cases, the methodology is the same. Dictionaries are used to extract the different senses and sub-senses found by lexicographers, which are mapped across languages to find clear correspondences. This slows for an initial list which evolves as corpus analysis is carried out, as will be shown in the case of "Europe."

Europe as a Variable Entity and Variable Space

We now have the three analytical tools we need to explore word usage across languages and across time: collocational networks, collocational resonance and lexicographical prototypes. These are precious aids in the studies carried out within the IntUne-funded project into European identity and provide interesting insights into the usage of the proper noun "Europe."[1]

IntUne was a major European funding initiative looking at the development of a potential European citizenship and identity. The project brought together political scientists, sociologists and a group specialized in media analysis, which is what we are concerned with here.

[1] This research was funded by a grant from the INTUNE project (Integrated and United: A Quest for Citizenship in an Ever Closer Europe) financed by the Sixth Framework Programme of the European Union, Priority 7, Citizens and Governance in a Knowledge Based Society (CIT3-CT-2005-513421).

To study changing attitudes towards Europe, two large press corpora were build, each taking in national and regional press sources and transcribed television news in four languages: English, French, Italian and Polish. These were designed as comparable corpora using a series of parameters to ensure a maximum of comparability (Bayley and Williams 2012). A different team looked at different aspects of Europe as reported in the press, and also at the concept of Europe itself.

"Europe" is an interesting concept as it is used so easily by so many without any real thought as to the variations of meaning that it holds. For instance, it can refer to an ancient Greek Goddess, a broad geographical space and a political unit. It can carry very positive connotations of hope, as is the case for many countries wishing to join as members or privileged partners, or strongly negative ones for those who see the bureaucracy of a centralized institution in Brussels as a threat. It is thus an ideal or an existing institution, a descriptive unit or a metonym. For a citizen of the United Kingdom, Europe is often "the continent," the mainland as opposed to the island, albeit geologically the island is linked to the mainland. Thus, a United Kingdom citizen can claim to have been born both in Europe and outside Europe, depending on the definition adopted. A country can decide to join Europe or leave it, but as you cannot join a continent only one meaning is being selected. And yet, despite this broad range of meanings, Europe is a proper noun and therefore excluded from most dictionaries.

Some dictionaries do give definitions for Europe—Larousse does, but only in its encyclopaedic section, and the *Oxford Dictionary of English* does because it is a good "universal"[2] dictionary. It is thus possible to build an initial prototype from dictionaries that define Europe as:

> a continent of the northern hemisphere, separated from Africa to the south by the Mediterranean Sea and from Asia to the east roughly by the Bosporus, the Caucasus Mountains and the Ural Mountains (ODE),
>
> or:
>
> Used ellipt, and allusively to refer to membership of the European Economic Community (the Common Market) OED.

[2] I use "universal" in the sense applied by the great French seventeenth century lexicographer Antoine Furetière. He set out to go beyond the dictionaries of so-called hard words, and beyond the so-called "polite" words of the dictionary of the French Academy. Universal meant that French as a language could cover all known concepts, and this is the sense I use here. Furetière did not supply a definition of "Europe," however.

Unfortunately, the ODE gives only the standard schoolbook definition, whereas the OED was outdated by 2009, the year of the DVD's publication, the actual edition being from 1989 (supplemented in 1993). The EEC became the EC in 1993, becoming part of the EU in 2009. The interesting thing is that although the political unit has been regularly rebranded, the metonym "Europe" has at least remained stable. The definition is a perfectly good dictionary definition as it allows for how the word is used in general parlance and accepts the vagueness of language. The problem is that the borders to the east and southeast are vague, and vagueness poses major problems in politics as much of Turkey and the ex-Soviet Union should be considered as in Europe when politically they are not. Much of the East is assigned to Asia which terminologically is where the problem lies, as for geographers the landmass is part of a broad Eurasia, if not an even broader Eurafrasia. This could be written off as anecdotal for the common person, but geology includes undersea landmasses and, when petrol is at stake, geology can be very political. Thus, any prototype should take into account a geological Europe, as does that published in Williams et al. (2012). The latter, however, was designed to handle a modern political Europe and fails to mention the goddess, although she is very present in the eighteenth-century literature and must therefore be added. Before looking at how the modern day "Europe" is described, it is interesting to see how it was represented in the past.

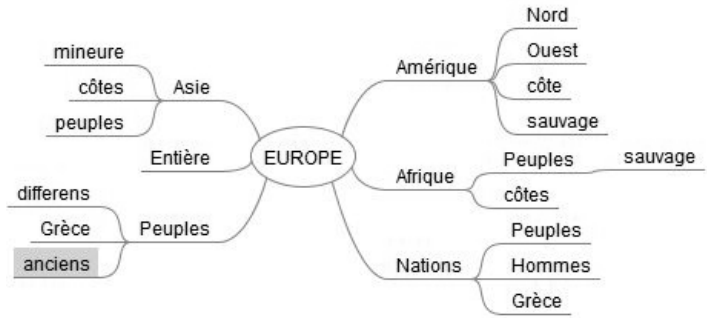
Collocational resonance can be demonstrated using dictionaries and corpora. Ideally both are used, but it is very much a question of being pragmatic about what is available. For English, the OED is a wonderful source and citations can form basic prototypes. However, citations are not corpora and give a limited, albeit useful, picture. Decent diachronic corpora of English are rare or expensive. For French there is no equivalent to the OED, but there is a large archive of texts that can be used to build some form of corpora [3] and an excellent library of freely available texts at the French national library—the BNF—which includes most early French dictionaries. These are mostly in PDF format but the lengthy and expensive task of producing fully machine-readable documents is underway.[4]

[3] *Frantext* (www.frantext.fr) is an archive of texts originally assembled as a source of citations for a major reference dictionary. It should not be considered a corpus as it is by no means a balanced sample. It is, however, an excellent window into usage during periods in the past. Access is by a restricted license, but larger sections of the archive are available at a sister website, www.cnrtl.fr.

[4] Available at http://gallica.bnf.fr.

Using a selection of texts from the second half of the eighteenth century, a network concerning "Europe" can be built (see Fig. 1.6 below).

Fig. 1.6. Eighteenth-century French Europe

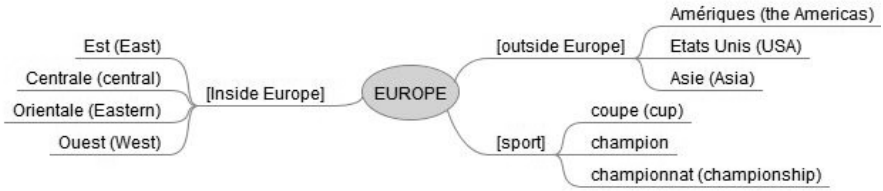


This provides an undefined Europe as consisting of unnamed nations. The geographical spread is probably that given in the great encyclopaedia of Diderot and Alembert[5] that corresponds to the vague definition given earlier (see ODE) bounded by coasts and Asia to the east. As with here, the encyclopaedia divides the world into four, with Europe as the civilized part and the rest mostly peopled by savages [*sauvage*]. This is the revolutionary period in France so it mostly justifies itself by reference to ancient Greece [*Grèce*] and proposes spreading their ideals to the people [*peoples*] of the world, and notably Europe. "Nations" is another interesting word as the nation is seen as the incarnation of the people and thus with no room for the individual, which explains the mass executions and purges that were such a feature of the French revolution. A similar English corpus of the time would probably be less sanguine.

If we now turn to contemporary Europe, as shown in the FrWac corpus on Sketch Engine,[6] a very different picture emerges (see Fig. 1.7 below).

[5] Available at http://http://portail.atilf.fr/encyclopedie.
[6] www.sketchengine.co.uk.

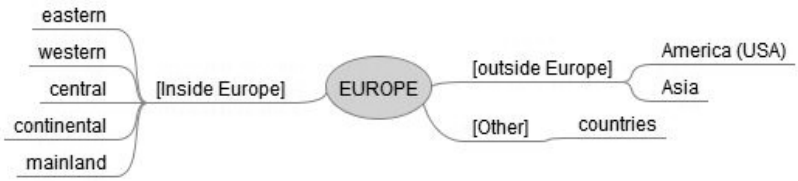
Fig. 1.7. “Europe” in the French IntUne corpus



We still have the division of the world into large groups on a more or less continental basis, but the Americas are no longer savage and are largely reduced to the Unites States as a major trading partner. Similarly, Asia is also mostly present in academic terms. The other new element is the major place taken by sport in which the emphasis is on competition between countries. It is mostly commerce, not ideals, that dominates with Europe remaining a broad geographical space.

The important change is that this Europe is subdivided into country, i.e. political units with defined borders rather than peoples or nations. The major division is however into two blocks seen as having different values: the East and the West. This same division is found in texts in the English corpus (see Fig. 1.8 below), so we shall turn to this first.

Fig. 1.8. “Europe” in the English IntUne corpus



The most obvious feature in the two corpora is that the only resonance is a vaguely shared political space. The notions of nations and peoples have gone and are replaced with countries. Countries are deemed as having stable fixed borders, which, within most of political Europe, are believed to be permanent fixtures, even if the situation is unclear in the Balkans and an underlying instability exists in some other countries. Countries, as units, are easily countable, which is essential in a political unit that counts its members. From 1945 onwards, it was accepted that Europe could be divided into two large units: the West with its capitalist values and the

Communist-controlled East under Soviet domination. Such a division ignored a pre-war situation where a highly cultured Mitteleuropa was emerging. Once the USSR had collapsed, the countries of Mitteleuropa that had been outside political Europe and joined under the title of new entrant countries again found themselves designated as a form of Mitteleuropa, albeit one enlarged to include the Baltic states and part of the Balkans as central Europe. This implies that there is a centre and thus a new Eastern Europe, notably the Ukraine and other countries east of central Europe. This explains the use of "central" and "eastern," but also a certain confusion as to whether east is "east" or "central." In French, the word "orientale" refers to central and eastern countries, with the former being EU members. Thus, corpus data shows multiple political Europes without even talking of the other Europes revealed during the IntUne analyses (Williams et al. 2012). These political Europes can be shown diagrammatically (see Fig. 1.9 below).

Fig.1.9. Europe with a centre

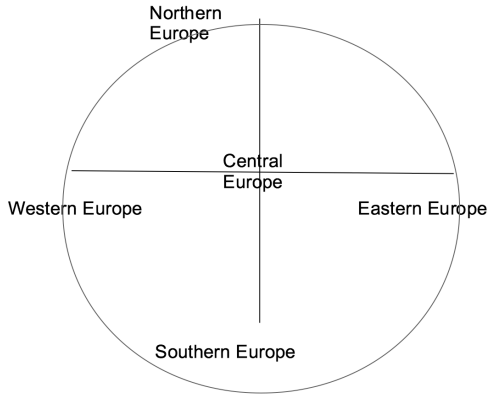


Fig. 1.9 above shows a Europe that does not exist and has never existed, politically. It is, however, one that collocational resonance points to. If Europe is the geographical (but not geological) space referred to in dictionaries then it should have a centre. It points to another fiction that arises from the resonance image, and that is of Europa the goddess. She is not mentioned, but the underlying fiction is that all of "Europe" is based on a Greco-Latin cultural root. This is partially true from school education, but ignores Northern and Eastern cultural traditions that have been undervalued in formal education. What corpus data shows is several realities, one of which is shown in Fig. 1.10 below.

Fig. 1.10. A divided political space

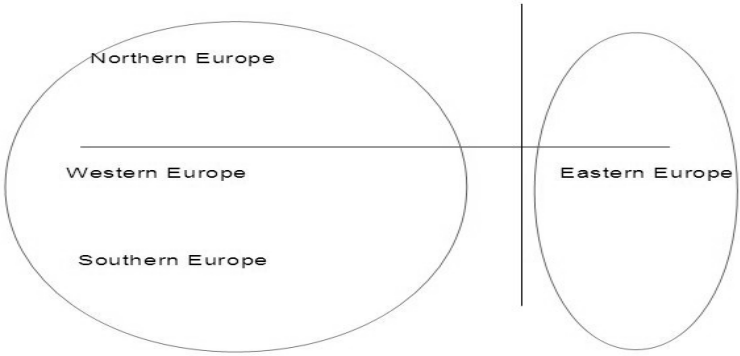
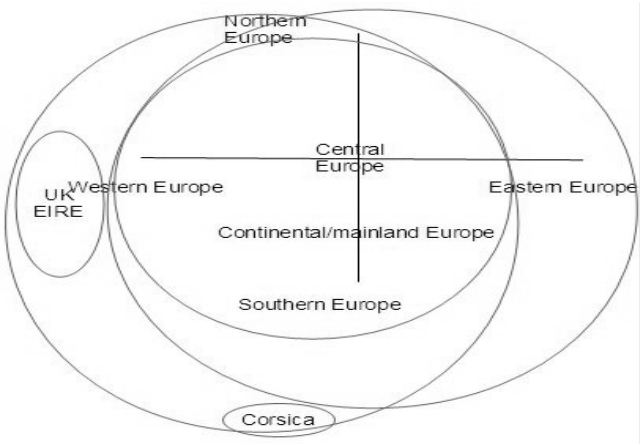


Fig. 1.10 shows a post-1945 and up to the entry of the so-called new entrant countries description of Europe with no centre. This has a very strong resonance as new entrant countries are often still referred to as "east" rather than "central." This is the archetypal Western Europe, which supposedly has a set of shared values and is rich and industrial. This Europe is even part of a wider fictional "West" that includes the USA. The resonance is strong and easily seen in the analysis of press corpora. If we move to current political Europe, and that shown in the English IntUne corpus, then an even more complex picture emerges (see Fig. 1.11 below).

Fig. 1.11. Complex Europe



In Fig. 1.11 above, the wider circle to the left covers current EU member countries, as well as islands such as the United Kingdom and French Corsica that still tend to talk of "the continent" and the mainland," which puts them some way outside of Europe. It is true that this diagram cannot account for Switzerland—very much in but out, and Norway—again, in and out. Thus, a confusion remains as to what Europe is, and yet we are told that some European countries join Europe and others want to leave it. Resonance is clearly a fiction, but these fictions are also quite real. Corpus linguistics shows us the folly of glibly using the word "Europe" as a metonym, and yet so much hangs around this terminological inexactitude.

Conclusion

To sum up, this paper has drawn together three threads in corpus-driven research and shown how they can be applied to a highly used and yet highly ambiguous concept—Europe. From its initial use in showing thematic relations in scientific corpora, collocational networks are now shown to demonstrate how all language has thematic groupings that can show current and past meaning patterns. Collocational resonance is an application of networks that demonstrates meaning and thematic changes over time. Lexicographical prototypes are the tools that mediate between past and current uses with collocational remanence being the degree to which earlier usage remains reflected in the present.

As such, these computer-assisted analyses can be taken as just another intellectual discussion of meaning. If they are applied, however, they are powerful tools in dictionary making that can lead us to genuinely electronic dictionaries that go beyond showing fictions of meaning to the continually evolving contextual meaning that language use is all about.

References

Alonso, A. 2009. *Características del léxico del medio ambiente en español y pautas de representación en el diccionario general*. PhD diss., Universitat Pompeu Fabra.

Alonso, A., H. Blancafort, C. de Groc, Chr. Million and G.. Williams. 2012. "METRICC: Harnessing Comparable Corpora for Multilingual Lexicon Development". In *Proceedings of the 15th EURALEX International Congress*, 389–403. Oslo: University of Oslo.

Bayley, P. and G. Williams (eds). 2012. *European Identity: What the Media Say*. Oxford: Oxford University Press.