



HAL
open science

A cultivated audience: Comparable corpora and cross language collocation

Geoffrey Williams

► **To cite this version:**

Geoffrey Williams. A cultivated audience: Comparable corpora and cross language collocation. Clavier 09, Nov 2009, Modena, Italy. pp.39-64. halshs-00725346

HAL Id: halshs-00725346

<https://shs.hal.science/halshs-00725346>

Submitted on 24 Aug 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A CULTIVATED AUDIENCE: COMPARABLE CORPORA AND CROSS LANGUAGE COLLOCATION

GEOFFREY WILLIAMS
University of Bretagne Sud

Abstract

Comparable corpora can give an insight into the meanings that have become attached to words that on the surface may seem to have simple translation equivalents, but which in reality overflow with meanings displaying the cultural values of the source and source language. This text uses the IntUne set of corpora drawn from the British, French and Italian press to show how the words 'culture' and 'cultivate' have not only changed over time, but display very different meaning potentials when found in the press of the three countries. In analysing the corpora, the study draws on the notions of collocational networks and collocational resonance.

1. Introduction

We have long been trapped into the notion that words have meanings, and that consequently dictionaries contain the true meaning, and that if meanings have changed overtime, then there must be an original sense that can be found by delving into the past. It is so easy to ask what a word means, rather than making the effort of turning the question round and asking what meanings have been attached to this word form.

Once it has been accepted that the word-form gains meaning only in context and that what it carries as more or less permanent traits is only by convention, it is possible to see word forms as containers into which we pour meanings. The meanings that have been poured in are like liquid in that they do not have a rigid easily dividable structure, but that different senses flow from one to another. The metaphor of container is quite frequent, and can be further exploited when looking at word meanings – containers can overflow or leak.

Leaky containers and spilt meanings can be found everywhere. This is because meanings can only ever be realised in texts. Dictionaries contain potential meanings based on past usage, but do not hold any absolute truth. If leaky meanings are part of *parole*, then a more stable version in terms of *langue* can only reasonably be found in a balanced corpus, and even then the results are only valid within the constraints and time frame of that corpus.

In this text we shall look at words which, whilst being fully lexicalised, are so used that the container is overflowing. This is the delexicalisation process described by John Sinclair using the simile of fading blue jeans (Bill Louw, personal communication). The word meanings will be explored in three corpora, each in a different European language. This raises the question of comparative corpora and the degree to which they are fully comparable if cross-cultural variation is to be taken into account. Meaning will be looked at by applying the use of lexical prototypes (Hanks 1994; 2000) and within corpora using collocational networks (Williams 1998) to explore the different textual environments and, through collocational resonance, (Williams 2008) to look at meaning change over time.

2. The IntUne context

There is little point in doing corpus work if the environment for which the corpus has been created is not discussed. This is because all corpora are biased by the application for which they have been assembled. No corpus is neutral, all corpora have been developed to fulfil a need. Exchanging corpora is a necessary exercise as corpus building is a time-consuming task. However, in doing so it is necessary to understand what decisions were taken and why so as to understand what a corpus

was designed to represent.

In this case the corpora were developed by teams in four different countries all of whom were participating in the European Commission funded IntUne project (www.intune.it). The IntUne project set out to look at aspects of citizenship across Europe by bringing together specialists in sociology, political sciences and media studies. The different working groups all had their specific information sources, mostly interviews and opinion polls for the two former groups. The media, in the form of corpora of written and television news, would be looked at by corpus linguists from France, Italy, Poland and the United Kingdom. Although the project was Europe-wide, costs limited the corpora to only four languages. This paper will deal with the three that the writer can handle with relative ease.

Insofar as the aim was to study whether a sense of European citizenship is developing, the role of the media group was to see how the press might be influencing opinion amongst the populations and the political elites. To this end it was necessary to build two series of corpora, each more or less coinciding with the information gathering periods of the other researchers. These corpora would be comparative.

Comparative corpora aim to look at different languages whilst overcoming the bias introduced by translation in parallel corpora. This means having a set of shared corpus compilation criteria in addition to shared management structures. Whilst the latter, in the form of Text Encoding Initiative-conformant XML data, do not pose a major problem, selection criteria introduce a number of difficulties. As with any corpus, it is necessary to start with external criteria (Sinclair 2005) and then move onto internal ones (Williams 2002) in order to evaluate the corpus.

For the IntUne corpus, the aim was to have stylistic and content-level comparability, which inevitably meant making some hard choices such as not using the British tabloids, for which there are no real equivalents. The situation is similar as concerns the economic press. These factors we shall come back to later.

The first level of comparability was to have newspapers representative of both national and regional press. For the former one was to be to the right, of a more conservative tendency, and one to the left, with a more liberal outlook. For the regionals, the aim was to have two from geographically different regions. To these four newspapers, we also added transcription of the evening news from two television channels, one state, one private. It is a fact that radio remains a highly influential news media, maybe more so than television, but this would not only have complexified the comparability levels but also added costs.

Thus, whilst the top-level criteria give a degree of comparability, it is looking inside the corpora that will show the inevitable skewedness. Skewedness is a factor that cannot be avoided, but one of which all corpus linguists who “trust the text” (Sinclair 2004) should be aware. As all corpus linguists know, a corpus is not a mass of data, but a collection of texts. If the linguist stays at the woods level, then the essential information contained in the trees will be lost, and all meaning with it.

Many factors can skew a corpus, the events of the moment being a particular one. For IntUne, the first French corpus was dominated by the French presidential elections, with the other corpora taking more or less interest, but always concentrating on interpretations of French politics. In the second period, it was the beginnings of the current ongoing financial crisis. The wordings are thus linked to events; we have not the meanings, but the meanings being vehicled at a precise time in a precise context.

As mentioned early, some of the skewedness arises from what is not selected. Newspapers do not form limited discourse communities in which interaction between members is active, but rather wider readership communities with newspapers trying to appeal to a core readership, a potential wider readership, and the others, including corpus linguists and discourse analysts who dip in from time to time. This presumably explains the lack of tabloids – the audience simply does not exist elsewhere in Europe, but in leaving them out we miss a large British audience and a major socio-cultural grouping. Even when comparing the comparable, other factors arise. Is *the Guardian* to the left? Is it really an equivalent to the French *Le Monde* which is centre left, and very Parisian elite in

outlook? In a truly comparable corpus we would be doing a close analysis of readerships, but what this would reveal is that the news and its language is not neutral, but cultural. Comparative corpora will not be comparable, but a window into different cultures. This cultural variation is what this text is about, taking a look at the notion of ‘culture’ between three of the IntUne corpora and showing how a corpus cannot provide simple translation equivalents, but complex ones in which the cultural background must be analysed before generalisations about meanings are made.

3. Methodological background and method

As revealed by Firth, collocation is a central organising factor in creating meaning. The concept of collocation has been developed in a number of ways in both lexicographical and contextualist applications, no one having a universal truth but all revealing different aspects of the phenomenon. The first to talk of collocations in a modern lexicographical tradition was Palmer (1933) in his *Second Interim Report on English Collocations*. This was essentially a collection of what might later have been termed *lexical units*, some of which were classified as collocations. It was from this work however that a phraseological and lexicographical tradition arose leading to a picture of collocation as a combination of selected parts of speech forming syntactically valid patterns. Such an approach is enshrined in the BBI dictionary of English Word Combinations. (Benson *et al.* 1986) and more recent corpus-based dictionaries of collocation. This is tamed collocation, what Firth revealed and Sinclair developed was a much more revealing picture of how word combinations form meaning rather than simply claiming that such combinations exist. The former are essential in that they can be captured in a dictionary, the latter reveal what language and meaning are about. Firth’s revelation was totally independent to that of Palmer, whose work remained largely inaccessible until recently. Firth’s revelations were carried forward by Sinclair in the nascent paradigm of corpus linguistics where it was shown that statistics could be used to extract significant patterning (Sinclair *et al.* 1970|2004). In this way, contextual collocation can be seen as statistically significant co-occurrence of lexical items with a fixed textual window. This has been developed in various ways, including those of *collocational networks* and *collocational resonance* (Williams 1998, 2007) that will be developed here. In handling variations of meaning a third element will be introduced, lexicographical prototypes as described by Hanks (1994, 2000).

4. Collocational networks and resonance

Collocational networks build statistically-based chains of collocations. The growing networks build a lexis that is closely linked to a theme, similarly to, but much more valid than, brainstorming a word. The networks demonstrate semantic groupings that intuitive and formal analyses miss in that they are not textually based. In comparing pairs in the developing networks, lexicographic collocations appear and can be recorded, but in addition to these we find data that brings to light the semantic prosodies (Louw 1993, 2000) that are linked to wordings and give then meaning in context. In early networking, no constraints were applied, but in a POS tagged corpus it is possible to refine and limit the networks by part of speech. This can be demonstrated in the French corpus by looking at the noun *identité* [identity] and isolating its modifiers (Fig.1).

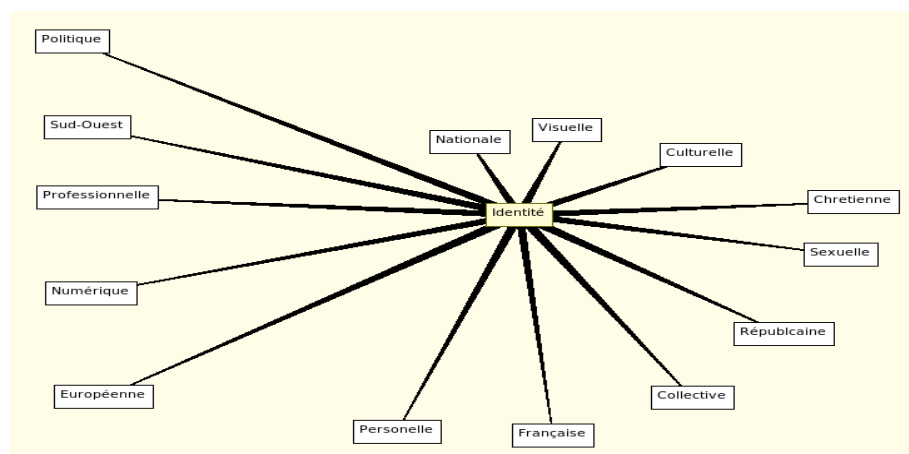


Figure 1. about here

As can be seen, identity can take many forms. It is possible now to classify these into groups, but also to explore in turn the collocates for *nationale*, for example, so as to expand the network and see how common collocates may aid or hinder any initial classification.

Collocational networks give a picture of how a word may be used synchronically within a given corpus. However, as words evolve another means must be found to find how earlier uses of a word may influence current usage, whether that meaning be currently active or not. This is the role of *collocational resonance*.

A collocational network supplies a meaning environment within the context of a corpus. Were it possible to trace networks over time, it would be possible to see what aspects of meaning are stable and those that change over time. Resonance concerns past aspects of meaning that may or not have been carried over in time. This can be seen in two ways: intertextual resonance, whereby formulae are used and reused, or metaphoric resonance wherein a single word evolves in meaning. The first is dealt with in Williams (2008) with reference to the Bible and Shakespeare; in this case the second, first mooted by Hanks (2005), but not published, will be developed. The two are intimately related. Given that diachronic corpora are not always available, a historic dictionary can provide a window into the past; this is what will be done here using the *Oxford Dictionary of English* in its CD-ROM version.

Networks and resonance will bring up the recurrent problem of polysemy. In dictionaries this is often dealt with by giving the impression that word meanings can be broken down into neatly numbered senses. In reality this is never the case. Meanings flow into each other over a continuum with different aspects being activated under different contexts. In addition, as meanings evolve, new characteristics may enter the paradigm with others falling silent. To deal with this, Hanks (1994, 2000) has put forward the notion of lexicographical prototypes, prototypes that work on the basis of a continuum as in Wittgenstein's *family resemblances* (1951). In such a case, a series of propositions that form aspects of a word's potential meanings can be listed. In a given context several of these can be activated, giving rise to a particular sense. This was designed to handle polysemy in monolingual contexts, but it has since been adapted to look at meaning variation across languages (Williams *et al.*, forthcoming) and this is how it will be handled here.

5. The problem of Culture

In an elicited response to the question as to what constitutes 'culture', an immediate answer from an educated westerner might be something along the lines of 'appertaining to the higher arts – painting, opera etc.' This is certainly one of the meanings that will be found in most dictionaries, but, as is often the case, elicited meanings may not coincide with what is actually found in texts. This can be illustrated using a text published in *the Observer* on the 25th October 2009 and entitled "Ending the silence on 'honour killing'". The text concerns the murder of a young Muslim woman by her father for having digressed certain rules within her ethnic group. Quite apart from an analysis of the word *honour* in such a context, it is the word *culture* that appears in a justification of the crime that is refuted by the judge. The first three examples refer to *culture*, the following three to the adjective *cultural*.

- (1) 1 silence is being broken. "It is not a problem of *culture* or religion or education " it is happening in e Honour_killing.txt
- (2) 2 case of Banaz, the judge said that if this is the *culture* then the culture needs to be changed, not the wom Honour_killing.txt
- (3) 5 ant communities where people feel adrift in a new *culture* and try to anchor themselves to

- the past is a key Honour_killing.txt
- (4) 1 attitudes of accepting the crimes in the name of *cultural* sensitivity have also disappeared and the police Honour_killing.txt
- (5) 2 etrators were treated leniently under the name of *cultural* sensitivity. Now there are no reductions in sente Honour_killing.txt
- (6) 3 fi Muslim Council. "Religion becomes infused with *cultural* practices and honour takes on an overinflated imp Honour_killing.txt

What is apparent from the first set is that we are not talking about the arts, but the traditions and customs within a society as being something that may be linked to education or religion, but are not necessarily condoned by either. Culture relates then to behaviour within a particular social grouping and one that should be changed to take into account the norms of any new society in which an individual or group of individuals enters. The behaviour related to a norm can result in cultural practices. Taking these into account by another members of another society entails cultural sensitivity, an awareness and acceptance of other forms of behaviour.

5.1 Dictionary culture

If we are to see how this tribal Muslim culture, as the behaviour described in the article is not that of Muslims in general, relates to the notion of the arts we can turn to dictionaries. Here we have the word as given in three English dictionaries spanning from the first alphabetically organized dictionary of English, that of Cawdrey, via Dr Johnsons' famous 1755 work and then to a current corpus-based dictionary:

- Culture, husbandry, tilling. (
- Cawdrey. *A Table Alphabetical*, 1604)
- Culture. n.s. [cultura, Lat.] The art of cultivation; tillage. Bacon. The art of improvement and melioration. Tatler.
- (Johnson. *Dictionary of the English Language*. 1755)
- Culture. Noun.1. The arts and other manifestations of human intellectual achievement regarded collectively. (Pearsall, Hanks. *Oxford Dictionary of English*. 1998)

What is apparent is that we have a slide from an earlier meaning related to farming to the more recent one where, in English, this aspect has entirely given way to other meanings. This is confirmed in the *Oxford English Dictionary (OED)*, where evidence of the meaning change overtime is mapped. What also becomes clear is the resonance that lies in the preparation of the land to implant seeds, to that of preparing the minds to sow the seeds of knowledge. Gradually it is the metaphorical sense which has evolved to become the literal one in current British usage. Taking all the various definitions in the modern *Oxford Dictionary of English (ODE)* a series of propositions that will form our basic prototype can be isolated. This gives the following list:

- The arts regarded collectively.
- Other manifestations of human intellectual achievement regarded collectively.
- A refined understanding of the arts regarded collectively.
- A refined appreciation of the arts regarded collectively.
- A refined understanding of manifestations of human intellectual achievement regarded collectively.
- A refined appreciation of manifestations of human intellectual achievement regarded collectively.
- The ideas of a particular people.
- The ideas of a particular society.
- The customs of a particular people.

- The customs of a particular society.
- The social behaviour of a particular people.
- The social behaviour of a particular society.
- The attitudes and behaviour of a particular social group)
- The cultivation of bacteria, tissue cells etc. in an artificial medium.
- The cultivation of plants.

If we apply this prototype to our example of ‘honour killing’ then only a certain number of propositions are activated, as can be seen in table one.

Table 1. Culture as group behaviour

The ideas of a particular people	
The ideas of a particular society	X
The customs of a particular people	
The customs of a particular society	X
The social behaviour of a particular people	
The social behaviour of a particular society	X

What is activated relates to a particular society, in this case Kurdish immigrants, rather than to an entire people.

Amongst the uses of *culture* in the English IntUne corpus we find the word used as a modifier with a very different sense. In this case we have an extract from the *Guardian* newspaper in which the politician Tony Benn contacts the minister in charge of the arts, the culture secretary.

- (7) Benn has written to culture secretary Tessa Jowell, asking her to overrule any decision to cut funding. *Guardian*. 14th April 2007.

This activates a different set of propositions to that in the former example as shown in table 2.

Table 2. Culture as higher arts

The arts regarded collectively	X
Other manifestations of human intellectual achievement	X
The ideas of a particular people	X
The ideas of a particular society	
The customs of a particular people	X
The customs of a particular society	
The social behaviour of a particular people	
The social behaviour of a particular society	

Here we are within the context of a national social group within its own country. Reference to funding is to what the government will spend on promoting theatre, films or other socially recognised art forms. The ideas and customs of people is shared, the rest is not, at least within the context of this corpus. However, a word of warning must be added. As Firth (1957) pointed out, saying that “dark” collocates with “night” is insufficient. Neither word retains a simple meaning that is altered by the other; both come together to form a new meaning unit. This is the case with ‘culture secretary’ where the sum is more than the addition of the two words. Thus our prototype looks at ‘culture’ and not ‘culture secretary’. To understand the new meaning units two profiles must be drawn to see how they interact before seeing how they act as a unit. This must be borne in

mind as collocational networks are not simple accumulations of words, but groupings of words that interact to form meanings.

5.2 Cross-language culture

Culture is a highly charged word in English with a variety of meanings both in a diachronic and synchronic aspect. The next stage is to see how it is displayed in bilingual dictionaries before moving onto the IntUne corpora.

Generally we find fairly simplistic translation equivalents as in table 3.

Table 3. Translation equivalents of ‘Culture’ in *Le Robert & Collins* English - French and the *Larousse* Français – Italien dictionaries

English	French	Italian
Culture	Culture	Cultura (civilisation, knowledge)
		Coltura (agriculture)
Cultural	Culturel	Culturale
Cultivate	Cultiver	Coltivare

This is standard bilingual dictionary practice in that one word is deemed as replaceable by another. In fact, *Le Robert & Collins* on CD-ROM does give more detail by showing multi-word units and idioms as can be seen in Figure 2.

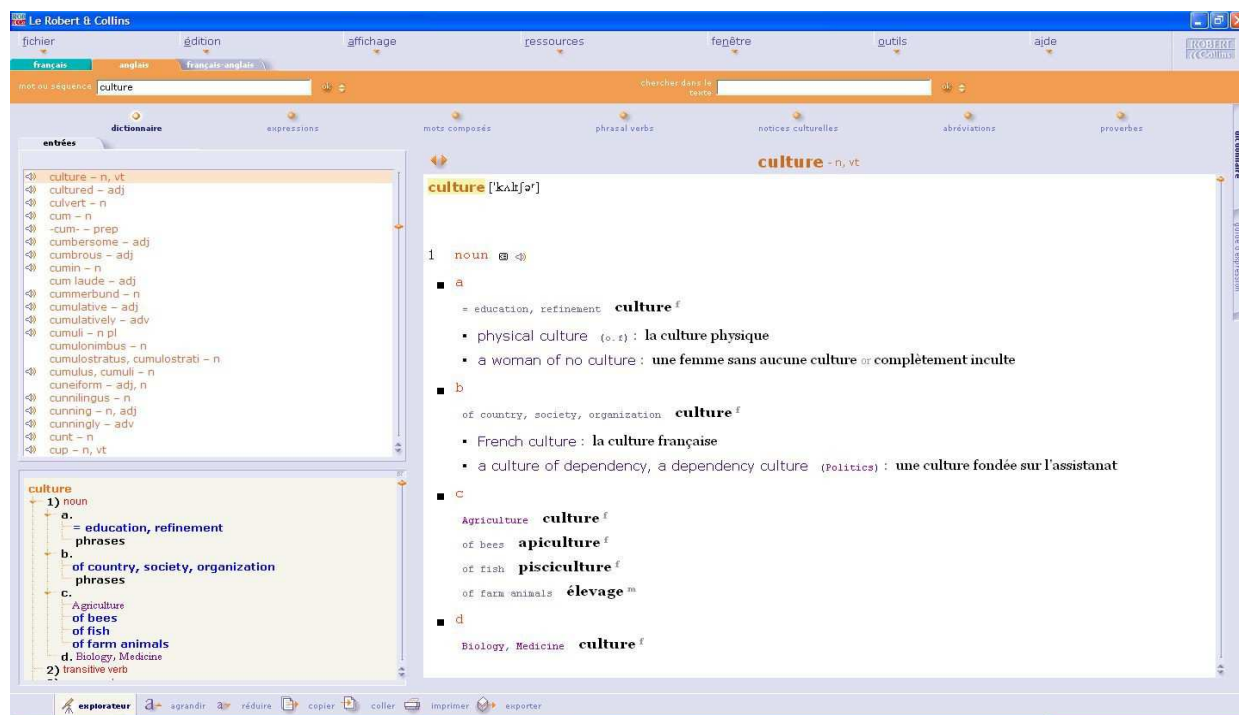


Figure 2. ‘Culture’ in *Le Robert & Collins* English – French Dictionary

As can be seen, ‘higher’ culture is broken down into two groups, both of which have links to the arts, but also other strange elements. In the first group we have education and refinement, both of which would normally appertain to the arts, but the two multi-word units are negations of this. ‘Physical culture’ is a more politically correct form for body-building, a notion generally linked with a connotation of all brawn and no brains. Why we have ‘a woman of no culture’, when there are more than a few men who share this predicament is a mystery, all the more so in that the better translation would be *inculte*, that is to say without learning. Group B gives the culture of a social group, but the examples are those of a higher culture rather than that referred to in the ‘honour

killing' extract earlier. 'French culture' clearly refers to literature and the arts, whereas the second example is of a smaller social group described with an implicit negative connotation carried by the French *assistanat*. Group C deals with what is in English a redundant farming sense that is only active in compounds and in a biomedical sense found in group D. The dictionary is obviously addressing an English user as the contextual clues are in English. Reversing the translation (Figure 3) we find many more multi-word units, but no specific mention of social groupings, although this is the case with *corporate culture*, and a general notion that culture relates to knowledge. The negative collocation of *mass culture* is lost as culture itself has not been delimited. In presenting someone who is not *inculte*, in this direction a 'man of culture', we are also supplied with the derived modifier 'cultured' that we find in Figure 3.

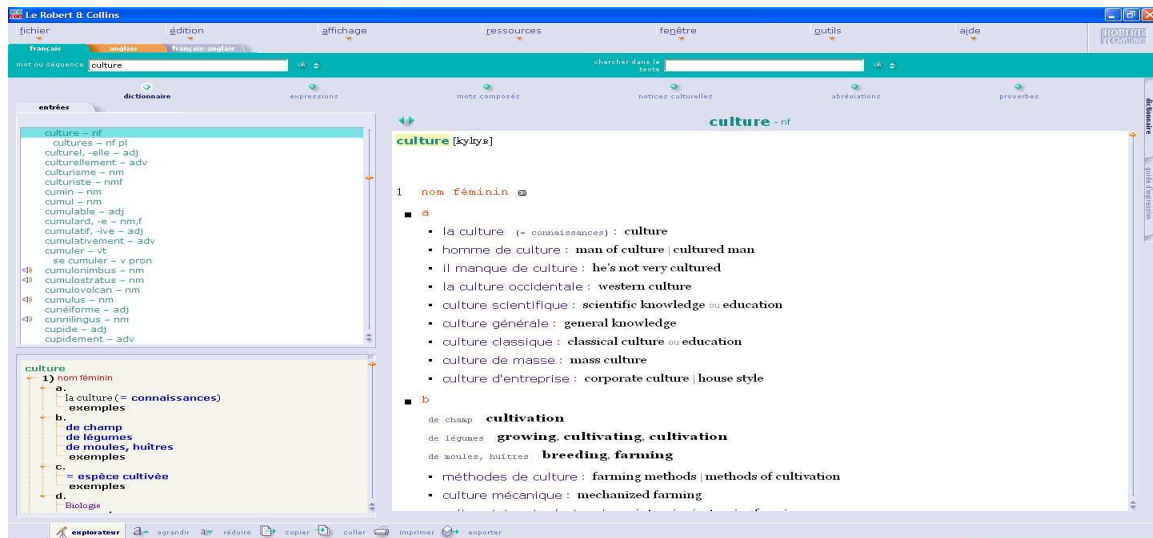


Figure 3. 'Culture' in Le Robert & Collins French - English Dictionary

Group B in this direction needs to clearly present the fact that two large senses remain in French by giving clear references to farming in the examples. In the French-English direction the contextual clues are now in French.

The *Paravia* French-Italian dictionary is much richer in content, with 6 main senses outlined with accompanying examples of potential collocations. The agricultural sense is broken down into the act of growing foodstuffs and the varieties of plants grown. The knowledge meaning has three senses: ‘civilisation’, ‘knowledge’ and ‘the economy’. The sixth sense is the biological one that is also found in English.

- 1 agr. (action de cultiver) coltura, coltivazione; la ~ du blé la coltivazione del grano; ~ d'un champ coltivazione di un campo; mettre en ~ mettere a coltura; petite, moyenne, grande ~ piccola, media, grande coltura; aire, terre de ~ area, terra da coltivazione
- 2 agr. (espèce cultivée) coltura, produzione agricola; ~ d'hiver coltura invernale; ~ d'exportation coltura da esportazione, per l'esportazione; ~ céréalière coltura cerealicola
- 3 biol. coltura; ~ in vitro coltura in vitro
- 4 (civilisation) cultura; la ~ européenne, chinoise la cultura europea, cinese; ~ de masse cultura di massa; ~ d'entreprise, de groupe cultura d'impresa, di gruppo
- 5 (connaissances) cultura; ~ encyclopédique, générale, musicale cultura enciclopedica, generale, musicale; vaste ~ vasta cultura; ~ classique cultura classica; femme, homme de (grande) ~ donna, uomo di grande cultura; avoir de la ~ avere cultura; ne pas avoir de ~ non avere cultura
- 6 econ. cultura; subventionner la ~ sovvenzionare la cultura; budget de la ~ budget per la cultura
- Il cultures f.pl. agr.
- (terres cultivées) colture
- ~ extensive coltura estensiva; ~ intensive coltura intensiva; scol. ~ physique educazione fisica; sport cultura fisica; ~ s vivrières coltura

Figure 4. ‘Cultura’ in the *Paravia* French-Italian Dictionary

Now that we have seen how different language pairs deal with the word, we can move onto the picture painted by the three corpora to see how well the usage is covered and also how comparable the notions are across language cultures.

5.3 Culture in the Corpus

5.3.1 English Culture

For each corpus the approach will be the same: an initial collocational network of nouns, names and modifiers derived from a 5:5 span and using z-score to calculate significance. The network for English can be seen in Figure 4.

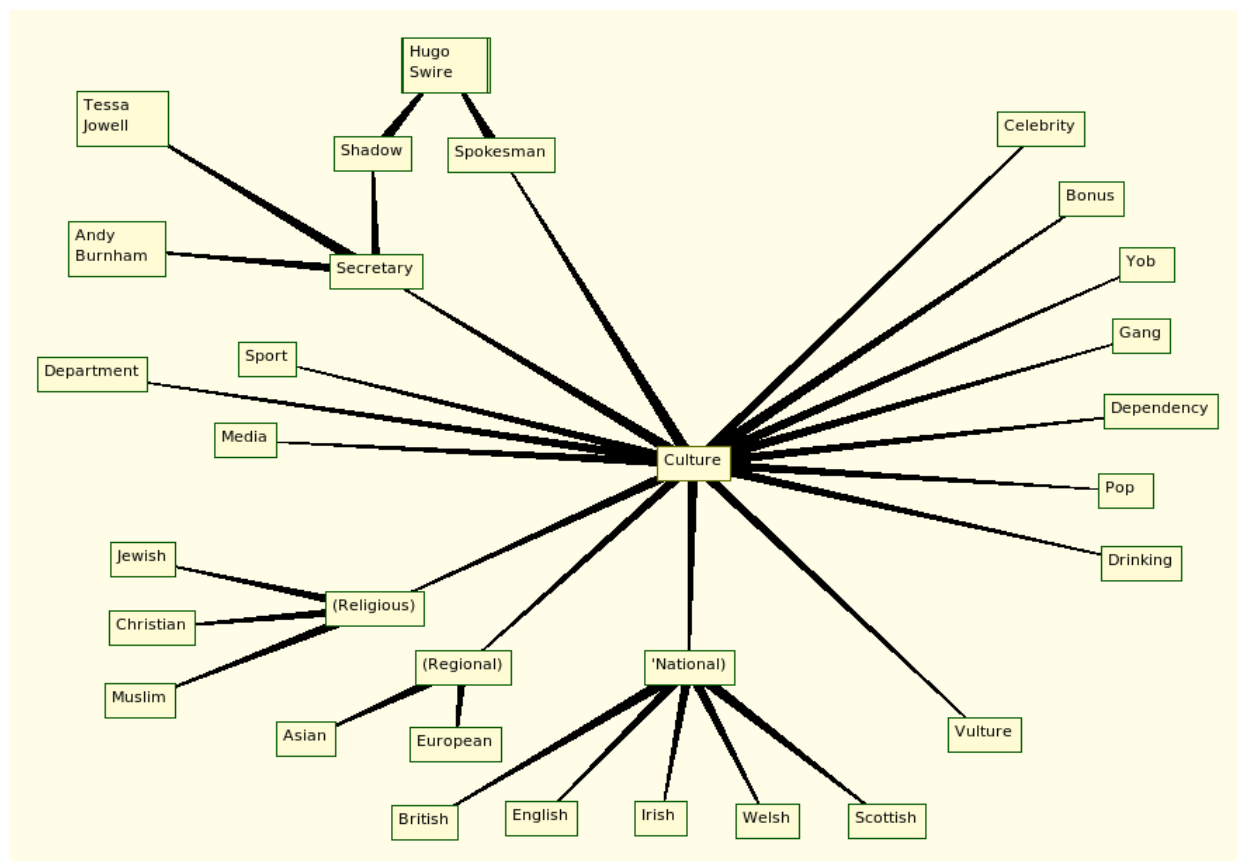


Figure 4. Culture in the British corpus.

What is apparent is that five groups of collocates can be isolated; the government department and its minister, culture linked to religion, regional culture, national cultures and a large group in which, as we shall see, culture is often negatively connoted.

Culture is highly collocated with the behaviour of certain social groups. *Yob* and *gang culture* are obviously frowned upon, with the former associated with football and the latter with inner city youth, generally black. *Dependency culture* was highlighted in the dictionary and concerns those who do not work but require welfare benefits. *Drinking culture* is also related to youth crime, but goes well beyond that in the tendency of Britons of all ages to over-indulge. Adult drinking is reprehensible in others. This leaves *bonus*, *celebrity*, *pop* and *vulture*. The first is recent and criticises people in the financial centres for awarding themselves excessive bonuses. The second and third can be related and may not always be negative. The former is bad when the press is over-interested in the lives of people deemed successful in film, music or sports, while the latter is seen as slightly vulgar and trivial. Culture vultures, on the other hand, are interested in higher culture, but over-indulge without being truly appreciative.

Bilinguals, and English dictionaries in general, fail to take into account this negative grouping of people whose behaviour does not conform to the norms of the community. These can be broken down into two subgroups, what is inherently criminal, and what is over-indulging – taking something that is in itself harmless to an extreme. In fact, five groups can be brought together into two – culture as something that is representative of the intellectual traditions and values of a society that are to be promoted, and the socially reprehensible behaviour of smaller groups within society. Insofar as we are working within a continuum, then different senses will move between categories, with pop and celebrity culture sitting on the outskirts of higher culture, their crime being that they are ephemeral. The others are on a continuum of bad behaviour from the mild to the criminal, depending on the effect on other members of society. Over-indulgence in alcohol can be considered amusing – until a road accident occurs.

If this is how the British press presents ‘culture’, we can now see what happens in other countries.

5.3.2 French Culture

A stereotype of France is that it is obsessed with its own literary and artistic culture. This is clear from the French – English combination in *Le Robert & Collins* where national culture is epitomised by *la culture française*. However, in the IntUne French corpus this does not stand out. As in Britain, there is a Minister, the title and breadth of responsibilities of the ministry may differ from what is found elsewhere, but the function of a government department promoting intellectual traditions remains. Similarly groupings of national, religious and regional culture can be found. The nearest we have to pop culture is hip-hop, but this is the state taking on board imported inner city culture so as to be closer to youth. No negative forms of social behaviour are to be found associated. The big difference with the UK corpus is that agriculture remains central, but that it is here that a certain negativity comes in.

The formula ‘culture de’ can be often associated with negative arguments, in this case it is genetically modified organisms (OGM – *organismes génétiquement modifiés*) that are being grown as are illicit drugs. Obviously, depending on the viewpoint of the newspaper, and on the moment in time, genetically modified organisms can be seen in a more or less positive light. The national press does not take a great interest in the growing of, for example, maize, but highlights only what is newsworthy. Neither drugs nor modified plants can be legally grown for human consumption in France. Given that the European Commission is listening intensely to the agribusiness multinationals, genetic modification is newsworthy.

5.3.3 Italian culture

Unlike French, Italian distinguishes farming, *coltura*, and intellectual activities, *cultura*, so that the network is already more centred. As in both the other corpora, the major groupings are those of the official bodies, and their representatives, along with religious, regional and national cultures. None of the collocates are negative in nature, but the links are often to art forms. To find genetically modified plants, and a great deal more use of metaphor, it would be necessary to turn to *coltura*. Here, however, we return to the *culture* and its adjective form.

5.4 Cultural networks

If the noun *culture* has produced very different collocational networks, the situation with the adjective is very different. In both the UK and French corpora (Figures 5 and 6), *diversity* (*diversité*) is a key collocate as in a *multicultural* society recognising and valuing differences is seen as essential.

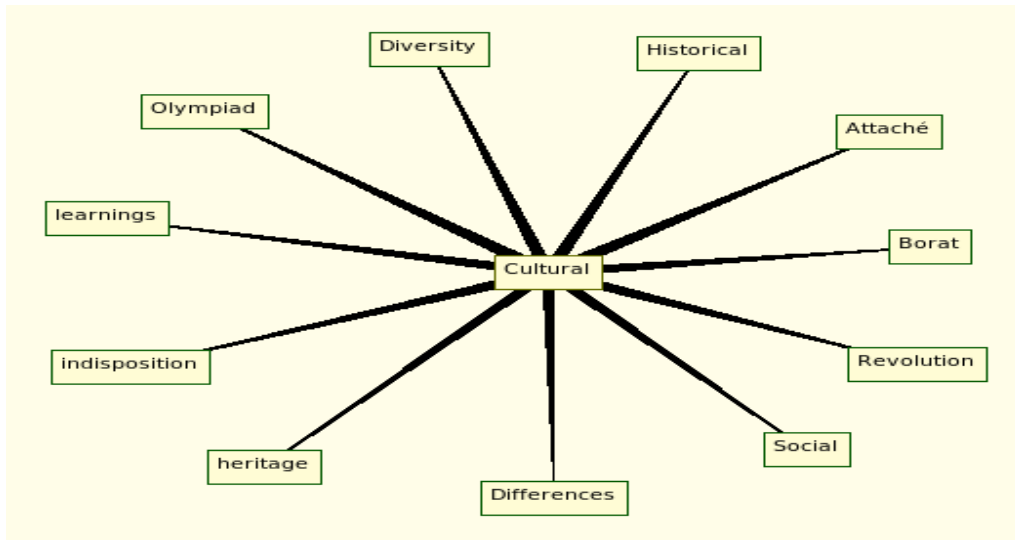


Figure 5. Network for the adjective *cultural* in the UK corpus.

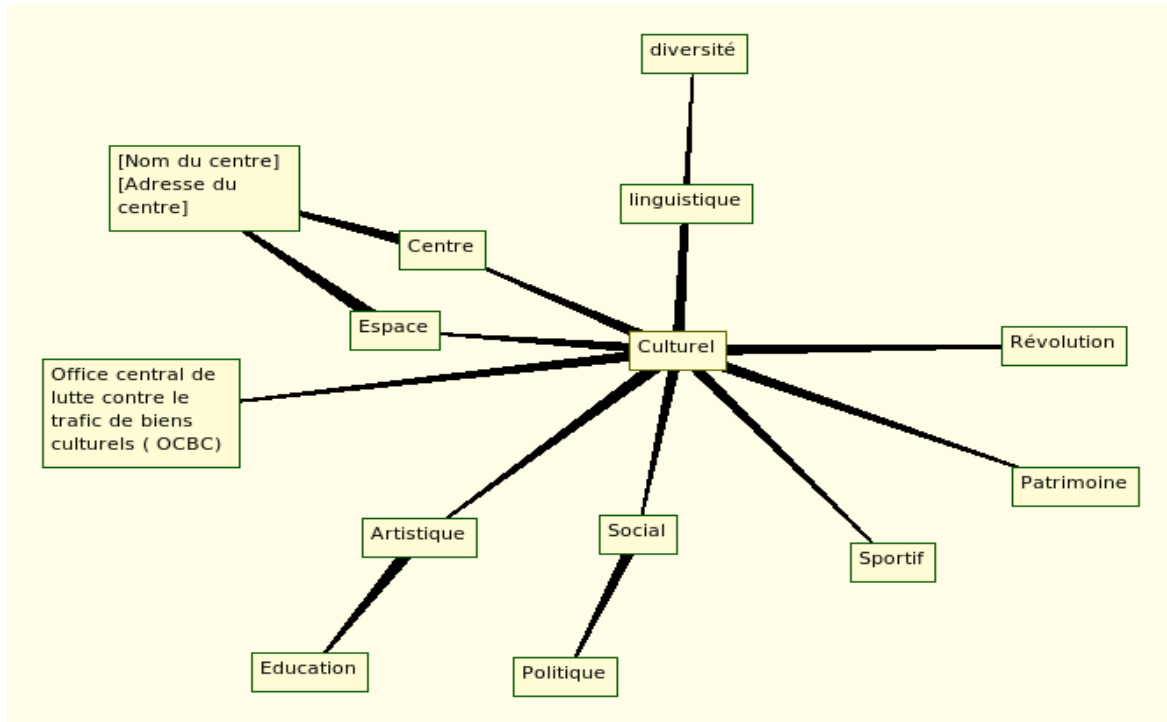


Figure 6. Network for the adjective *culturel* in the French corpus.

However, there are major differences between these two. The UK corpus, in particular, concentrates on social aspects and change in society, with history and heritage playing a central role. *Borat*, the main character in a film is an accidental occurrence. Both corpora use the term cultural revolution for a major social change, but whereas the UK corpus is mainly concerned with the cultural revolution that started under Mao in the 1960s, the term is more widely used in the French press, although there is also coverage of the events in China.

In the French and Italian corpora (Figure 7), heritage is also present as *patrimoine* (French) and *patrimonio* (Italian). The difference with the UK corpus is that he two have cultural “centres”, a phenomenon that is missing in the UK corpus. In all three cases, the adjective refers only to intellectual effort so that there is little danger of ambiguity, which is not the case with *cultivate*.

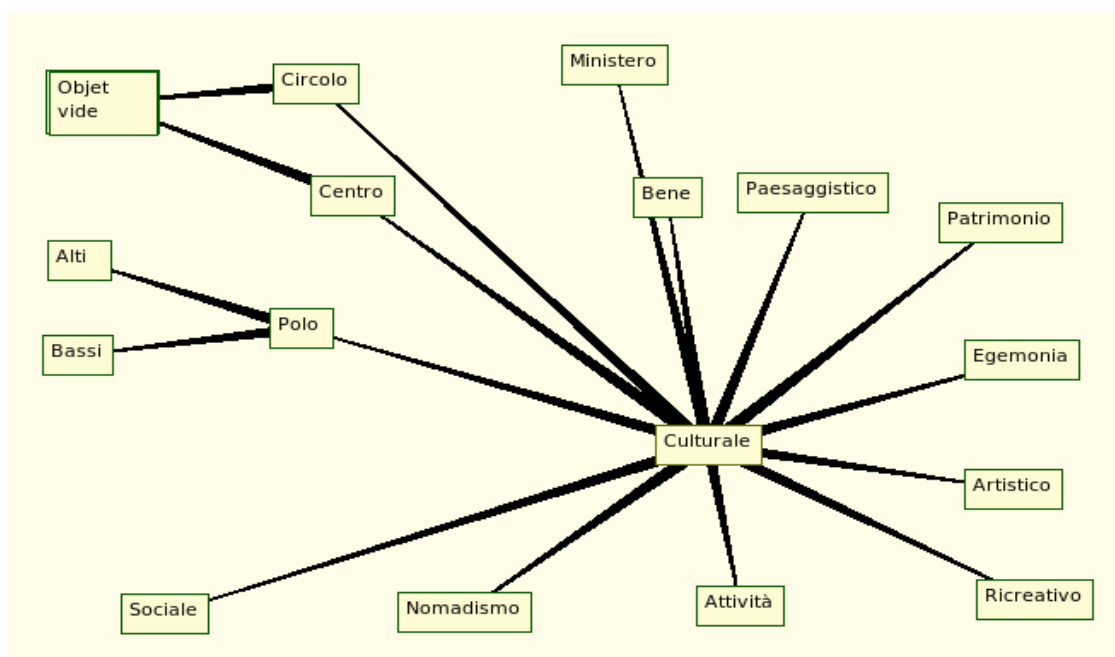


Figure 7. Network for the adjective *culturale* in the Italian corpus.

5.5 Cultivating differences

In English, the verb *cultivate* followed a similar widening of reference to that of the noun *culture*. Initially attested as referring to tillage, the art of cultivating the land, by 1681, the *OED* gives us the figurative use of cultivation of the mind so as to ‘improve and develop by education or training; to refine, to culture’, with the adjectival form *cultivated* referring to persons or minds ‘improved by education or training; refined, cultured’. The link between *culture* and *cultivation* is assured.

The English collocates clearly fall into two groups, with *cultivate* being almost exclusively concerned with gardening and vegetable growing and *cultivated* referring to matters of the mind, interests, sentiments or people may be cultivated, in other words the fact of being to be agreeable to someone who can be of use to one. The nation of gardeners is concerned very much with its vegetables, and its self image.

In French (Table 4), four main senses are apparent: the land to be tilled as a surface (*hectare*) or as a piece of land (*potager* [vegetable garden], *jardin* [garden], *parcelled* [field] ...), the plants to be grown (*tomate* [tomato], *cereals* [cereals] ...), a method of cultivation (*biologique* [organic]) or a person who cultivates (*paysan* [peasant or farmer]). The latter is an interesting case as in English *peasant* is derogatory, but is not necessarily so in France. For some time it was associated with backwardness, with farmers having become *agriculteurs*. Now things are changing with the *paysan* being the protector of the land, whereas the *agriculteurs* are seen as polluters of the soil under the control of the agribusiness multinationals.

Table 4. *Cultivé* and its agricultural collocates in the French corpus.

Headword	Category	Frequency	Z-score	Type	Sub-type
Maïs	NOM	28	77,1	agriculture/horticulture	plant
hectare	NOM	28	56,3	agriculture/horticulture	land
potager	NOM	7	45,8	agriculture/horticulture	land
Jardin	NOM	26	43,8	agriculture/horticulture	land
OGM	????	18	43,1	agriculture/horticulture	plant
aromatique	ADJ	5	39,4	agriculture/horticulture	plant
légume	NOM	14	34,7	agriculture/horticulture	plant
céréale	NOM	7	32,1	agriculture/horticulture	plant
Variété	NOM	9	26,3	agriculture/horticulture	plant
Vigne	NOM	8	25	agriculture/horticulture	plant
Plante	NOM	9	21,5	agriculture/horticulture	plant
parcelle	NOM	6	20,8	agriculture/horticulture	land
agriculteur	NOM	10	18,9	agriculture/horticulture	person
Tomate	NOM	4	17,9	agriculture/horticulture	plant
surface	NOM	9	13,4	agriculture/horticulture	land

Forêt	NOM	7	12,4	agriculture/horticulture	land
Paysan	NOM	5	12,3	agriculture/horticulture	person
biologique	ADJ	7	11,8	agriculture/horticulture	method
Terre	NOM	12	11,7	agriculture/horticulture	land

Self-image also appears in French (Table 5) and is very much a personal matter as one cultivates ones image (*x cultivate son/sa [singularité / différence]*), in addition to one's garden (*x cultivate son jardin*) where the latter can also be figurative with a *jardin secret* and *jardin politique*)

A similar picture appears in the Italian corpus (Table 6), but with different outlooks. As expected, Italians also cultivate their gardens and fields and the plants therein. A personal side also exists, but the collocational tendency attested here is to cultivate personal qualities (*virtù* [virtue], *potenzialità* [potential]) and an inter-relational network (*relazione* [relations], *amicizia* [friends]).

Table 5. *Cultivé* and its self-image collocates in the French corpus.

Headword	Category	Frequency	Z-score	Type	Sub-type
Discrétion	NOM	4	14,7	Personal quality	image
Art	NOM	12	10,4	Personal quality	image
Lien	NOM	9	11,9	Relations with other	interpersonal
Singularité	NOM	4	22,9	Self-image	image
Ambiguïté	NOM	5	19,2	Self-image	image
nostalgie	NOM	5	16,7	Self-image	image
Mystère	NOM	6	16,1	Self-image	image
différence	NOM	10	12,8	Self-image	image
son	DET	131	12,3	Self-image	image
paradoxe	NOM	4	12,1	Self-image	image
image	NOM	13	9,9	Self-image	image

paradoxe		NOM		4		12,1		Self-image	
Headword	cat	Frequency	Z-score	Type	subtype				
ortaggio	NOM	6	62,7	agriculture/hor	plant				
appezzamento	NOM	4	47,2	agriculture/hor	land				
orto	NOM	12	37,2	agriculture/hor	land				
allevare	verb	6	36,8	agriculture/hor	method				
mais	NOM	6	35,9	agriculture/hor	plant				
uva	NOM	4	29,8	agriculture/hor	plant				
verdura	NOM	6	28,8	agriculture/hor	plant				
patata	NOM	4	23,1	agriculture/hor	plant				
vigna	NOM	4	22,9	agriculture/hor	plant				
terreno	NOM	14	22,2	agriculture/hor	land				
giardino	NOM	10	18,5	agriculture/hor	land				
frutta	NOM	4	16,4	agriculture/hor	plant				
erba	NOM	4	14,2	agriculture/hor	plant				
poesia	NOM	8	14,2	art	literature				
continuare	verb	36	24,8	formula					
assecondare	verb	4	19,2	formula					
passione	NOM	44	61,4	Personal Interest					
sogno	NOM	32	45,5	Personal Interest					
interesse	NOM	20	21,7	Personal Interest					
virtù	NOM	6	16,1	Personal qualif	image				
potenzialità	NOM	4	13,9	Personal qualif	image				
relazione	NOM	14	16,4	Relations with	interpersonal				
amicizia	NOM	6	13,3	Relations with	interpersonal				
						image			
Image	NOM	13	9,9	Self-image	image				

Table 6. *Coltivato* and its agricultural collocates in the Italian corpus.

6. Conclusion

Tilling the earth and sowing seeds is obviously a rich source of metaphors, a movement from literal to figurative meaning in which the figurative becomes literal. Whether the usage be literal or figurative, the word has been given many new meanings through developing new collocates in associated contexts. As one sense spills over to the next, the discrete word loses all meaning of its own, if it ever had one.

What is apparent is that the dictionaries divide up senses in very different ways. The lexicographers have carried out an analysis in each of the language combinations and have given potential translation equivalents for the categories provided. The first problem is to know which world-view is being described. If this is implicit in a monolingual dictionary, the situation is not so clear in a bilingual one. The French-English is an odd mixture addressing an English speaking audience but with examples that bear little relevance to English speakers, all the more so in that someone wishing to look up agriculture terms is unlikely to go for the headword *culture*. The French-English version may be of more use to a French speaker, but the multi-word units are far from inspiring. We fare better with the *Paravia* French-Italian combination, apparently addressing a French audience, which is richer in detail. The unanswered question remains as to what ‘culture’ is being described and translated. To approach a cultural environment of use we require a corpus.

Insofar as meanings have words and not the reverse, meanings can only be found in discourse, and for this recourse to a corpus is invaluable. The corpora used in this paper are deemed comparable in that they were built using declared external criteria. The skewedness due to factors of

differing newspaper formats have been discussed and taken into account, so that the three can be considered representative of a certain newspaper coverage at given periods of time. This is the limit of their comparability. The next stage is to take key words that on the surface may seem translatable and explore their environments within the corpora so as to compare how they are used. This gives us a key into how different language cultures see the world, even when they share a strong common cultural background.

It is difficult to avoid the word culture, which is why it has been used here as a basis for this reflection on language and context. Collocational networks are a means of visualising the diversity of environment. Collocational resonance may provide a means to compare what is activated when comparing within and across languages. The interpretative means lie both within the corpus and within the wider discourse of each language.

In this case we start with a common core, that of the practice of arable farming. In an urban Britain, it would seem that the press has lost contact with this agricultural past and links mostly culture to the abstract elements that effect the behaviour of communities in a multi-ethnic country. Culture is both the affair of the government when it affects the arts, but is also the behaviour of smaller social groups, in which case a negative prosody seems to reign. The urban view is also present in the adjective *cultural* and the verb *cultivate* as reference to tillage is more that of the garden and the emphasis lies on personal development and world-view. The French and Italian corpora open yet another vista. France seems to be more centralised with emphasis on its ministry of culture, and its ministry of agriculture. And yet here too the urban dominated press is more concerned by newsworthiness than by agriculture. In language the agricultural aspect is carried by another noun, but shared in the verb where the metaphor element of personal development is present, but different from that found in the other corpora.

This all leaves open a number of questions that none of the dictionaries address. Do we share cultures? If not, how can we give keys to access that other culture rather than through simplistic translation equivalents? Do we cultivate the same values? Can, and should, a dictionary show how these values are reflected in supplying connotative and pragmatic input.

The main question is one of comparability. Is comparing to show what is similar, or to show what is apparently shared whilst mediated through a very different view of society and what differs? Rather than seeking translation equivalents, the first task may be to find what meanings are being activated and to seek the keys to open and appreciate that world-view rather than giving the comfortable impression that language works on parallel lines.

References

Books and articles

- Benson M., Benson E., Ilson R., 1986 (2nd ed.), *The BBI Dictionary of English Word Combinations*, John Benjamins, Amsterdam.
- FIRTH J. R. 1957. *Papers in Linguistics 1934-1951*. Oxford : Oxford University Press.
- Hanks, P., 1994. "Linguistic Norms and Pragmatic Exploitations, or, Why Lexicographers Need Prototype Theory, and Vice Versa", in F. Kiefer, G. Kiss, J. Pajzs (eds.), *Papers in Computational Lexicography: Complex '94*, Hungarian Academy of Sciences, Budapest, pp. 89-113.
- Hanks P., 2000. "Do Word Meanings Exist?", in A. Kilgarriff, M. Palmer (eds), *Senseval: Evaluating Word Sense Disambiguation Programmes*. Special issue of: *Computers and the Humanities*, 34, 1-2, pp. 205-215.
- Hanks 2005????
- Louw B., 1993, "Irony in the Text or Insincerity in the Writer? The Diagnostic Potential of Semantic Prosodies", in M. Baker (ed.), *Text and Technology*. John Benjamins, Amsterdam, pp. 157-76.
- Louw B., 2000,
- Palmer H.E., 1933. *Second Interim Report on English Collocations*, Kaitakusha, Tokyo.
- Sinclair J. McH., Jones S., Daley R., 1970/2004. *English Collocation Studies: The OSTI Report*. London - New York: Continuum.

Williams G., 1998, Williams G. (1998). 'Collocational Networks : Interlocking Patterns of Lexis in a Corpus of Plant Biology Research Articles'. *International Journal of Corpus Linguistics*. Vol. 3/1. pp. 151-171.

Williams G., 2007 Williams G. (2007). 'In search of representativity in specialised corpora: categorisation through collocation'. Teubert W & Krishnamurthy (eds). *Critical Concepts in Corpus Linguistics*. London : Routledge. Vol.3 : 324-346.

Williams G., 2008 Williams G. (2008). 'The Good Lord and his works : A corpus-based study of collocational resonance.' *Phraseology : an interdisciplinary perspective* (Granger, S. & Meunier, F. eds), Amsterdam : Benjamins. Pp 159-173.

Wittgenstein

Dictionaries

Cawdrey . *A Table Alphabetical*, 1604. published in 2007 by the Bodleian Library. Oxford.

Johnson *Dictionary of the English Language* 1755

Le Petit Larousse 2005.

Oxford English Dictionary (OED). CD-ROM Version 4.

Oxford Dictionary of English (ODE). Oxford University Press. Revised second edition. 2005. Soanes C & Stevenson A.

Le Robert & Collins French - English, English - French Dictionary. CD-ROM 2004.

Il Nuovo Hachette Paravia Francese/Italiano – Italiano/Francese. 2nd Edition, 2007. Paravia Bruno Mondadori Edition

Acknowledgements

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).