



# Art for dictionaries sake: Comparing cultural outlooks through dictionaries and corpora

Geoffrey Williams

## ► To cite this version:

Geoffrey Williams. Art for dictionaries sake: Comparing cultural outlooks through dictionaries and corpora. Roberta Facchinetti. English Dictionaries as Cultural Mines, Cambridge Schoiars Publishing, pp.171-200, 2012. halshs-00725345

**HAL Id: halshs-00725345**

**<https://shs.hal.science/halshs-00725345>**

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.

## CHAPTER NINE

# ART FOR DICTIONARIES SAKE: COMPARING CULTURAL OUTLOOKS THROUGH DICTIONARIES AND CORPORA

GEOFFREY CLIVE WILLIAMS

### 1. Introduction

Corpus linguistics is about being surprised. It is about being surprised by the commonplace questions of language that we take for granted, but for which we never take a close enough look at their complexity. If familiarity breeds contempt, then language is a good case. Although contempt may necessarily not be the outcome as we do value our languages, our acquired knowledge as educated individuals often prevents us from recognising the wealth of meanings expressed by even the simplest of words. We tend to judge by what we have learnt, rather than taking a fresh look at language. In decontextualising large amounts of data, corpora can force us into taking this fresh look.

An educated person does not read linearly. We may have done this when we were learning to read, but we now skim, making snap decisions from the key words of the text in context. Our knowledge of language enables us to extract meanings without having to worry about ambiguity, our contextual knowledge helps us decipher texts and handle polysemy with remarkable ease. This is obviously an advantage in fast information extraction, but can be a hindrance if our aim is to formalise and retain lexical information for future use. Generally, when in known waters, we do not need external help when decoding a text, but when in doubt we do turn to that repository of lexical knowledge that is the dictionary. Although this is what our education has taught us to do, this reaction is not without danger as, although most users do not realise it, dictionaries only contain meaning potentials, and not meanings themselves. They do not contain the full truth

and only the full truth as the senses, and not meanings. The senses that they do hold are invariably institutionalised ones. The consequence of all this is that the educated reader, and dictionary user, tends to have two separate knowledges of word meanings, a conscious taught one that enables us to give a 'correct' meaning when challenged out of context, and our acquired knowledge of language in context that may give a very different picture.

Traditional dictionaries are based on accumulated knowledge of language and the use of written sources mediated by the trained intuition of the lexicographer. New editions may be revised, but no publishing house can afford to rewrite all the entries and reorder the content. That non-corpus-based dictionaries do not portray current usage of words has been shown time and time again in corpus linguistics. Intuition is not the best judge of reality, and yet, surprisingly, we still very often accept that the dictionary is right, and real usage is, in some way, wrong.

For those who are willing to stand back and take a fresh look, and who are willing to give the time necessary to learn to read concordances, corpus linguistics has turned our view of meaning on its head. Concordancers, like educated humans, do not present text in a linear fashion, instead they present a key word with sufficient context for its particular textual meaning to become apparent. By sorting and grouping the contexts, recurrent contextual meanings stand out, and these are often at odds with taught knowledge. If we have eyes to see, and are curious about language, we are inevitably surprised with the results.

This text is thus about surprises, the surprises found by looking at one very familiar word in the context of several different corpora. This word is *art*.

*Art* has been chosen as an example, the same problems are to be found with all the words in the lexicon. It has been chosen as it is closely related to *culture*, another word that is widely used without taking a fresh look at its meanings in varying textual environments. Culture then shall be our starting point as this text will set out to show how dictionaries are essentially cultural artefacts that are the fruit of a social, educational and language culture.

First, we shall look at culture as a means of setting the scene for a particular analytical method and for a study of lexis within paper dictionaries. After discussing the nature of paper dictionaries with reference to the lexeme *art*, we shall move on to how this word is represented in English and French corpora. The conclusion calls for a considered rethink of dictionaries wherein the corpus provides more than just source data, but the whole means of structuring the dictionary itself.

## 1. The meaning of culture

Culture is such a simple word. Culture is one of the defining elements of advanced civilised societies. In educated societies we tend to see less developed societies as having traditions, whereas we have culture, meaning something that encompasses all the higher arts. Expressions such as 'primitive culture' or 'popular culture' simply show that lower forms of 'culture' exist, but that true culture is neither primitive nor for the masses, it is something better than that. This may sound presumptuous, but it is very much what we find in published works. This sense of a higher culture is an institutionalised meaning, one we have learnt at school and university, and from dictionaries. This can be illustrated by the following extract from the Oxford Dictionary of English (ODE):)

(1) The arts and other manifestations of human intellectual achievement regarded collectively. (ODE).

The highly popular French Larousse dictionary also gives this meaning, but in fifth place, when it describes, the phenomenon as being:

(5) Ensemble des usages, des coutumes des manifestations artistiques, religieuses, intellectuelles qui définissent et distinguent un group, une société (Larousse)

A collection of usages, customs, artistic, religious and intellectual manifestations that define and distinguish a group or a society.

ODE is a corpus-based dictionary, Larousse is not, but both highlight the intellectual aspect.

The interesting point is that we can easily accept that this is the central meaning of culture, when in reality earlier uses of the word had nothing whatsoever to do with the mind and intellect, but all to do with tilling the soil. The fact that meaning can evolve over time is not new, what we tend not to do is to map these changes and take into account what aspects of earlier meanings may be carried over into new contexts. We also tend to accept that meanings change, but fail to accept that the dictionaries upon which we rely may not be giving the current meanings of the word. Dictionaries contain senses, the meanings are in contexts.

## 2. Cultural resonance

Language changes over time. This is not decay, but change. Language change is brought about by changing circumstances. Words do not have meanings, we give them meanings in contexts. As contexts change, so will meaning. One of the major outcomes of contextualist linguistics in the Firthian tradition and of corpus linguistics is the acceptance that

dictionaries do not contain the meaning of words, but only describe how words have been used, and list meaning potentials (Hanks 2000). From these meaning potentials, we can extrapolate new meanings from new contexts.

Metaphor is a major factor in language change in that we exploit one sense to elaborate another. This is why Hanks (1994) claims that meanings should be described not in definitions, with senses and sub-senses, but through prototypes that allow for continuity between aspects of meaning. The value of using prototypes to compare meanings across time and across languages has already been demonstrated (Williams 2009, Williams et al forthcoming). Prototypes allow us to see that as meanings change, certain features are carried over, thereby colouring futures sense, even if an earlier sense disappears from usage. This process has been termed collocational resonance as aspects of meaning that colour the semantic prosody of their contextual usage are often attributed unconsciously (Williams 2008a).

The word *culture* has already been discussed in a cross-language perspective (Williams forthcoming). This study showed how *culture* has developed a variety of meanings in three contemporary languages – English, French and Italian as revealed in the IntUne comparable corpora. Just looking at English, we can see how from an earlier Latin usage, *culture* has moved from meaning tillage and husbandry, the two synonyms provided in Cawdrey's 1604 *Table Alphabeticall* in lieu of a definition, to agricultural in general, and then via the notion of human intellectual traditions to the current usage, which, in English, has retained only human intellectual behaviour and the arts. The fact of growing things has not disappeared entirely, but *culture* in this sense is limited to biological studies, the earlier sense lives on only in *agriculture*. These changes are well mapped for English in the Oxford English dictionary (OED). When we compare to French and Italian usage, the agricultural aspects are still active. In English however, the word seems to be taking on a new connotative meaning wherein *culture* is frequently linked to negative behaviours. This new aspect does not appear either in the OED or ODE, as such change is insidious and can only be seen through the use of corpora, which only goes to show the difficulty of tracking active language change. What is interesting about Cawdrey's usage is that he does not define, but simply gives two synonyms. In so doing, he is actually giving two senses in that tilling is what we would now call arable farming, whereas husbandry includes animals. In both cases the current word is farming. Neither of Cawdrey's synonyms are frequent in current usage, whereas on the other hand, farming did not have this meaning in Cawdrey's time. What is clear from corpus data is that connotative meanings change over

## Art for dictionaries sake: Comparing cultural outlooks 5 through dictionaries and corpora

time and with context. In addition to this, a much more subtle change is that whereby groups of contexts gradually create what Louw (1994, 2000| 2008) has termed semantic prosody, a situation where an aura of connotation can subtly giving new underlying meanings, often positive or negative, to words. Semantic prosody shows how *culture* has come to take on new meanings, including the notion of as excess in current English.

Thus culture, the word and its meanings, is cultural, that is to say fixed in the behaviour, traditions and expectations of different societies. Dictionaries can describe senses attributed to *culture*, but are themselves cultural in that they reflect the culture of the society for which they have been compiled.

Although the word *culture* has evolved differently in different languages , the relationship to human intellectual experience, and notably what we term the *arts* is a constant across western languages.

Language changes, and consequently so do the senses recorded in a dictionary. However, in a non-corpus-based dictionary, which means all the traditional reference works, the senses are entered and ordered by the intuition and accumulated knowledge of the lexicographer. Accumulated knowledge and a tendency to prescription may run against real usage with the user believing that the dictionary must be right, whereas they are wrong. This fallacy is remarkably current amongst teachers who often seek norms, without questioning the validity of those norms.

Thus, the senses found in a dictionary are not necessarily those that predominate in current usage. The problem lies partly in the nature of dictionaries, and the way we perceive them. Thus, before moving on to real usage as seen through corpora, we shall look at how meanings are encapsulated in a condensed form in dictionaries. To do this, we must first look at how dictionaries view their users and the extent to which dictionaries are driven by inherited notions as to what constitutes such a work and to what extent they are actually user-driven.

In the following section, we shall look at how *art* and related words are dealt with in English dictionaries so as to illustrate the problems with current paper dictionaries before moving on to corpus evidence and how that should change more than just content in a move towards interactive electronic dictionaries..

### 3. Dictionaries and their users

One of the problems with dictionaries is that they are too familiar. We are made to buy one, or more, when we start school, and at home there are

invariably dictionaries, because our parents were also brought up in the cult of dictionary worship. The problem is not necessarily with the dictionary itself, but with the fact that users lack dictionary knowledge necessary to judge quality and the essential question of to whom a given work is addressed. We are generally introduced dictionaries by our school teachers. However, this can be very much a question of the blind leading the blind in that it is highly unlikely that their critical knowledge of lexicography is very developed. This is simply not something that is taught outside of lexicography courses. In addition to this, if they have had a literary training, they may well see dictionaries as a prescriptive tool to uphold a perceived norm. The normative function of dictionaries is important, but whether or not they hold the current norm can be a moot point..

It is obvious that dictionaries do have a prescriptive role when used in the earlier stages of education. It is vital that learners perceive the language norms of the society in which they are being educated. Unfortunately, here is a danger that the dictionary be seen as the source of true meanings, and that is pernicious. Meanings change with the context whereas dictionaries can only contain institutionalised senses, and these change over time. As a consequence, not only do users often have a false idea of the authority of a dictionary, but they are also very conservative about them. If a dictionary appeals, why buy another? To some the perception is that the simpler is the better, polysemy only gets in the way is written off as a 'wrong' meaning. Obviously, in order to better understand and use the tool, and to develop a critical outlook, what users needs to do is to ask themselves what a dictionary really is. So what is it?

The standard view is that a dictionary is a collection of words in alphabetical order accompanied by a definition, an explanation, a commentary, a semi-encyclopaedic entry, one or more 'synonyms' or or, in the case of bilinguals, a translation equivalent. The question as to how these differ is not generally asked.

As we all know, dictionaries are not necessarily alphabetical in order. When Cawdrey entitled his 1604 dictionary 'A Table Alphabeticall', he considered the fact as being a novelty. Then, as now, not all dictionaries are semasiological, and even those that are have to make decisions concerning the order of run-on entries, compound forms and multiword units.

That dictionaries contain definitions is another misnomer in that many still give one or two synonyms instead of a phrasal or semi-phrasal 'definition'. Even when there is a definition, readers do not always necessarily know the difference between a paraphrase, a definition in the

## Art for dictionaries sake: Comparing cultural outlooks 7 through dictionaries and corpora

Aristotelian *definiens-definiendum* format, or the Full Sentence Definitions and explanations introduced in the COBUILD, and now widely used elsewhere (Hanks 1987). In addition, users bilingual dictionaries expect a translation equivalent, but do not necessarily ask how equivalent the equivalent is. Before even endeavouring to give a history of English lexicography, Bejoint (2010) feels the need to give an exhaustive analysis of what a dictionary is and who writes them as most people retain a folk definition of what constitutes a dictionary, and a hazy knowledge as to who writes them. It is therefore important to consider what dictionaries are, and what are the advantages, and drawbacks, of each dictionary type. The next question is who actually writes dictionaries and how they carry out their tasks. The link between 'dictionary' and 'lexicographer' is not known to all, although those who have come across the great Dr Johnson know that the latter is "a writer of dictionaries; a harmless drudge, that busies himself in tracing the original, and detailing the signification of words." (Johnson 1755).

When Johnson was at work, dictionaries were compiled by a single author. To some extent this is still true today, but the contemporary large reference dictionaries are written by teams of lexicographers under the control of an editor. If a lexicographer writes dictionaries alone, or as part of a team, the question as to the degree of competence of a lexicographer in scientific or technical dictionary writing may be raised. Many serious specialist dictionaries are written by scientists or medical practitioners, these are not necessarily lexicographers and may lack the analytical skills of the former, but they do have an in-depth knowledge of the subject areas. Other 'dictionaries' are written for fun or by those with an axe to grind. The lexicographical competence is lacking, they are simply alphabetical lists, but we still call them dictionaries. A so-called wiki dictionary is an on-line collaborative work, which cannot really lay any claim to any rigorous analysis of the lexicon, and yet they are freely available on the internet and much consulted. Even the great dictionary of the French Academy is not written by lexicographers, but by a committee of literati. Thus, however prestigious a dictionary may be, it does not ensure that the content has been written by reliable trained lexicographers.

The question as to what dictionaries are for is another rarely asked question. There is a tendency to see dictionaries from a purely European perspective, when in point of fact dictionaries first enveloped in India and China, well before Europe as an entity even entered human consciousness. In looking at the history of dictionaries, it is interesting to see why they evolved, Chinese dictionaries being linked to the need to standardised

character description for monolingual dictionaries and the integration of Buddhist terminology for bilingual ones (Yong & Peng 2008). Dictionaries have always arisen from a need to name and classify, and a need to teach those classifications.

Thus, dictionaries are a response to language needs and changing language circumstances. In Europe, they evolved from glosses found in medieval manuscripts in Latin with a need to explain and understand words. By the time of the renaissance, the problem was relating latin to the emerging vernacular languages. This is well seen in the development of English language lexicography (Cowie 2009) where we move from a need to understand Latin, then French, and then the English language itself, which led to the emergence of monolingual English language dictionaries, the first of which was that of Cawdrey.

On the continent, academies had been created to set a gold standard for the great nations of Europe and their language. The first of these was the Academia della Crusca which set out to impose the literary output of Tuscany, or more precisely Florence, as a linguistic norm. This was followed by France with its own royal academy. Thus, by the 18<sup>th</sup> century, the need to have a reference dictionary of English was felt, and this led to Johnson's mammoth undertaking. Johnson set out to prescribe, but his pragmatism quickly led him to a requirement to describe. Later, in the 19<sup>th</sup> century, Murray's Oxford English Dictionary responded to the need for a historical and complete dictionary of English as the language rose to the status of a world language. In the mid-20<sup>th</sup> century, Hornby's Oxford Advanced Learner's Dictionary, not at first considered a true dictionary in the noble tradition, addressed itself to the needs of the non-native speaker wishing to understand and produce correct English (Cowie 2010). Changing needs, changing dictionaries.

A similar evolution can be seen in France where the lexicographical tradition started by Pierre Larousse arose from the development of education for all under the *loi Guizot* of 1830, and thus the need for a popular dictionary of French, as opposed to the literary and historical one proposed by Littré (Pruvost 2005). In more recent times, the postwar development of higher education in France brought about a need for a new type of dictionary for educated users, a need fulfilled by the Le Robert dictionary.

Thus, the development of dictionaries is essentially driven by needs, whether societal or cultural as the title of Pruvost's (2006) history of French lexicography shows when he declares them to be "outils d'une langue et une culture" (tools of a language and a culture). The process of dictionary evolution is simple; different cultures bring forth different

## Art for dictionaries sake: Comparing cultural outlooks 9 through dictionaries and corpora

needs, and engender different answers. Nevertheless, the fact that they respond to changes in society, and to perceived associated needs, does not necessarily mean that they are user-driven. Dictionaries are often the reflection of a certain scholarly culture, and that culture may not fully reflect user needs. Dictionaries are eminently cultural tools and may reflect cultural, rather than, user, needs..

Consequently, although dictionaries reflect changes in society, the content may reflect more the position of a controlling elite than the users themselves. Thus, some dictionaries are prescriptive. This is the case of the dictionaries of the national academies which seek to create a prestige language, but are of little use in everyday life where a descriptive dictionary will be of more use. Some dictionaries, like the Oxford English Dictionary are historical. Although they are wonderful reference works, they are not designed for everyday reference usage. Some dictionaries will rely on citations from literary works, fine for prestige, but useless in the production of everyday language. The outcome of this variety of dictionary types, is the fact that to choose the best dictionary for any given purpose we should know when it was written, for whom it was written and why. These questions are rarely asked, and most dictionary publishers are very shy about limiting their audience by declaring to whom their productions are best suited. This is especially the case when we consider bilingual dictionaries, which must handle encoding and decoding in two directions.

To sum up, the ideal dictionary does not exist. If we want to build new dictionaries for new cultural conditions and new needs, we cannot start from a blank page. We need to look at and admire earlier dictionaries as they are not just a source of knowledge, but also provide highly innovative solutions to real problems of their time, and these can be adapted to current needs and technologies. However, corpora bring new sources that are closer to real usage than the lexicographical intuition of the past. Corpora can change our outlook, and may help change the format of the dictionary, but they remain a means and not the end. Thus, to build new dictionaries, we need to understand new technology and its possibilities, whilst keeping a healthy dose of cynicism if we are not to worship the technology more than what it sets out to achieve. In other words, we need to rethink where we are going, and above all, why. This I shall try to illustrate by returning to the notion of culture as being “the arts and other manifestations of human intellectual achievement regarded collectively” and seeing how culture relates to art and the arts. However, in so doing, we shall have to go back over some of the points about dictionaries raised

earlier. As a starting point, we shall start with the treatment of art in contemporary English dictionaries.

### 3.1. *The dictionary and art*

So what is art? We shall start with a corpus based English reference dictionary, the Oxford Dictionary of English.

According to the ODE art is :

- 1.The expression or application of human creative skill and imagination, typically in a visual form such as painting or sculpture, producing works to be appreciated primarily for their beauty or emotional power
- 2.(the arts) the various branches of creative activity, such as painting, music, literature, and dance
- 3.(arts) subjects of study primarily concerned with human creativity and social life, such as languages, literature, and history (as contrasted with scientific or technical subjects)
- 4.a skill at doing a specified thing, typically one acquired through practice

Phrases

- art for art's sake
- art is long, life is short

The above does not cover the entirety of the entry as I have deliberately left out the examples that are given for each sub-sense and phrasal pattern. In looking at the entries for *art* and its related word forms, two things appear clearly. One is the problem of lemmatisation as the above definition gives three different definitions of 'art' with the plural and definitive forms differing from the more general one. The second is that although the notion of human creativity can be found in all four definitions, they do vary greatly in scope. Dividing entries into separate senses is a necessary, but artificial task. This is clearly a case where a prototype would be useful, as this would not only highlight the links between senses, but might also help elucidate the phrasal expressions that follow. However, lemmatisation is not the only problem as other problems arise when listing compound entries and multiword units. Although *arts* may follow *art*, we cannot group all the related forms as we shall see when looking at *art* in three advanced learners dictionaries; the Oxford Advanced Learner's Dictionary (OALD, Collins COBUILD, and the MacMillan English Dictionary for Advanced Learners (MEDAL). Thus, the following takes into account certain 'facts' about dictionaries.

# Art for dictionaries sake: Comparing cultural outlooks <sup>1</sup> through dictionaries and corpora

## ***3.2. Dictionaries are alphabetical***

As can be seen in appendix one, dictionaries are more or less alphabetical, something which gives rise to numerous problems for users. The list shows the headwords found in the ODE compared with the three other dictionaries. Where a word exists in one dictionary, but not in another, the column is marked with an X. In this table, the entries retained are only single headwords and multiword units, the compound forms as artwork and other related forms as artist have not been retained.

That there are variations in headword lists is normal; a reference dictionary such as ODE will be more complete than a learner's one. In addition to questions of scope, there is also the choice of individual editors as to what is to be included in a word list. What is common points to a certain perceived relevance, similar corpus results, or maybe just comparisons with the competitor's lexicons. The COBUILD stands out as having its headword list rigorously based on corpus results, but even here different editions can show different results as the initial corpus has grown from the big COBUILD corpus to the vast bank of English. As in all metalexicographical analyses, similarities are interesting, differences are inevitable. What is most interesting in all cases is the degree of dispersal brought in by alphabetisation which leads to related entries being very far apart. Surely an electronic dictionary should be able to show links between related words?

## ***3.3. Different dictionaries define differently***

Dictionary entries differ in the way words are listed and also on the way they show homonymy and polysemy. For ODE, art has three entries, the first being the noun, the second the archaic verb form and the third the abbreviation for the noun article. Only MEDAL gives the verbal form on the basis that although archaic in usage, it is difficult to read English without coming across either the Bible or Shakespeare, or any work wishing to give a magical or medieval atmosphere. When we look inside the entries we find that they group word meanings, lumping, or divide them, splitting, differently. Thus, OALD and COBUILD each have seven senses, whereas MEDAL has only five. That they agree on the number does not mean that they will agree on the order in which they are to be represented. Again, this results from different editorial decisions. Where differences are most apparent is in defining styles.

OALD

1. The use of the imagination to express ideas or feelings

4. art, music, theatre, literature, etc, when you think of them as a group

#### COBUILD

1. Art consists of paintings, sculpture, and other pictures or objects which are created for people to look at and admire or think deeply about

3. The arts are activities such as music, painting, literature, cinema, and dance, which people take part in for enjoyment, or to create works which express serious meanings or ideas of beauty

#### MEDAL

1. Paintings, drawings and sculptures that are created to be beautiful or to express ideas

3. arts (plural) subjects of study that are not scientific, such as history, literature, and languages

4. the arts (plural) activities such as art, music, film, theatre and dance considered together

These definitions show how the ordering and content can vary between dictionaries. In the full entries for all three we find the difference between art, arts and the art with the key element of the arts being not scientific in outlook. The important point is however that of defining style.

OALD takes the traditional approach which does not repeat the headword but rather relies on the user adding in the formula X is. In many dictionaries a tilde is used to replace the keyword, these dictionaries all prefer a numbering system. The major difference lies in the full sentence definitions used by COBUILD. This is a definite policy decision designed to increase clarity and is based on Hanks call for explanations rather than definitions (1987). This has been the subject of much debate (Rundell 2006/2008) and whilst the definitions given here follow the more traditional format, MEDAL will make use of both depending on the situation. In other words, the format is part of an in-house culture that defines standards in terms of perceived user needs and expectations.

The wealth of dictionaries lies in choice, but do users see the choice? These are all first class dictionaries taking deliberate editorial decisions on a number of essential matters; defining styles, definition v. explanation, content information, highly concise entries v. the more descriptive, ordering of senses follow precise criteria. If the user is not aware of the issues, then the user has not the means to choose.

In addition to the explicit meanings, there are also implicit ones. A rapid prototype built from the learner's dictionaries give us the following facts/

- Art is about objects
- Art is deliberate
- Art is about beauty
- Art is about ideas
- Art is to be looked at

## Art for dictionaries sake: Comparing cultural outlooks<sup>1</sup> through dictionaries and corpora

- Art is to be enjoyed
- Art is to be admired
- Art is an activity
- Art can be studied
- Art has value

Whilst it is clear that art differs from science, which of these characteristics are activated when discussing *art*, *arts* and *the arts*? There is one of the arts, but is not art. And culture? Is culture about all three or just the *arts*? The answer is cultural. Different dictionaries either group senses, the lumpers, or divide them into a myriad of sub-senses, the splitters. This is partly an in-house culture, and partly a national culture.

### **4. Dictionaries as cultural artefacts**

Dictionaries grow out of a society that created them, so inevitably a change of language will bring about a change of style. This is of course simplistic, the English and American dictionary styles are very different (Béjoint 2010). In this article we have only looked at English dictionaries, so in comparing with usage in another language we shall turn to our continental neighbour, France. Obviously, despite a strong prescriptive tradition, the dictionaries found in France and in French-speaking countries as Québec will differ, but that is another story that we shall not attempt to tackle here. Neither shall we look at the academy or historical dictionaries, but concentrate on today's big sellers, *Larousse* and *Le Robert*.

#### ***4.1. Dictionaries of contemporary French: Larousse and Le Robert***

The first thing that becomes clear is that *Larousse* and *Le Robert* are lumpers. Both have very long entries listing what each considers relevant. In terms of multiword units, only *Larousse* has *art deco* and *art nouveau*, presumably because *Larousse* is more open to encyclopaedic information compared to the very literary *Le Robert* dictionary. Both of these are included as run-on entries. One important difference in terms of alphabetisation lies in the nature of the language. In English, art frequently qualifies another noun, as in *art gallery* or *art house*. In French this cannot happen so that *galerie d'art* will be elsewhere in the dictionary unless

included as a run-on. Thus the problem of dispersant of related items is going to be even greater in French than in an English dictionary.

If we turn to the defining style, there is no particular innovation. Both tend to be excessively concise, but are not necessarily clear to the uninitiated. In other words the approach is very traditional and does not seek closeness to the user as in the English dictionaries. The user is expected to know the model. The following gives a shortened list of the sub-senses found in the Larousse along with a translation. This is not a corpus-based dictionary and so ordering has been fixed by the lexicographer in line with in-house style. It is unclear whether this is historical or simply what is deemed as a 'logical' order.

1. Aptitude, habilité à faire qqch

*Aptitude, ability to do something*

2. Ensemble des moyens, des procédés, des règles intéressant une activité, une profession

*Entirety of the means, procedures, rules that concern an activity, a profession*

3. Création d'objets ou de mises en scène spécifiques destinés à produire chez l'homme un état de sensibilité et d'éveil plus ou moins liés au plaisir esthétique

*The creation of objects or specific staging aiming at producing amongst humans a state of sensitivity and awareness that is more or less linked to aesthetic pleasure*

4. Chacun des domaines où s'exerce la création esthétique, artistique

*Each of the fields in which is carried out aesthetic or artistic creation*

5.) Ensemble des usages, des coutumes des manifestations artistiques, religieuses, intellectuelles qui définissent et distinguent un group, une société

*A collection of usages, customs, artistic, religious and intellectual manifestations that define and distinguish a group or a society*

If we turn to Le Robert, we find a much more lengthy and complex entry filled with literary citations. The definitions are differently ordered and much more complex, which is why only a very simplified version is given. The very numerous literary citations have also been left out.

1. Ensemble de moyens, de procédés réglés qui tendent à une certaine fin

*Entirety of means, or regulated procedures which lend themselves to a certain end*

1. Moyen d'obtenir quelque résultat

*1. Means to obtain some result*

2. Ce que l'homme ajoute à la nature; ce qui est artificiel

*1. What man adds to nature, what is artificial*

## Art for dictionaries sake: Comparing cultural outlooks <sup>1</sup> through dictionaries and corpora

3. Ensemble de connaissances et de règles d'action dans un domaine particulier

1. *Entirety of knowledge and rules of action within a given field*

In all Larousse gives five senses, but which have little correspondence with the English definitions. It does however have colour illustrations and encyclopaedic entries. Larousse sets out to be a household dictionary and aims at a much wider audience than its rival.. Le Robert has no illustrations, but plenty of citations. In addition, the dense entries are potentially hard going, even for a French user, due to their concision and philosophical outlook. In other words, if you have not been trained in a certain literary tradition, it is quite impossible to extract information for everyday usage.

Both these dictionaries are aimed at native French users, but, although they do not say so, they essentially aim at very different audiences. It may well be that the Larousse is used more for confirming meanings and for spell checking, whereas the Le Robert aims at a more erudite university-educated literary audience. To an English reader, neither seems particularly clear, but this is normal in that they are not addressed to non-native users. They are, in theory, addressed to users with a particular dictionary culture that enables them to cope with this style of entry writing.

### ***4.1. Summing up dictionary cultures***

The study of the entries for *culture* and *art* do not seek to be an in-depth metalexicographical analysis of these words, but serve simply to show the defining styles, and hence the dictionary cultures, of two publishing houses. They also provide the means of comparison with popular English dictionaries. That they are different within each language and between languages is obvious. The question arises as to whether, in comparing dictionaries across languages, we are simply trying to compare the incomparable.

Audiences differ so that a reference dictionary aimed at a native speaker is obviously different from those aimed at non-native learners of a language. In addition to this there is in-house culture. The OALD and COBUILD have a very different approach to defining, the Larousse and Le Robert differ greatly an audience expectations. In addition to these factors, there is the question of national culture. Although words share a common origin, the meaning patterns will have developed in very different ways. This is clearly the case with *culture*, and apparently that of *art* too.

The difference found between monolingual dictionaries in content and style leads to other problems when we come to bilingual dictionaries. In the latter case, beneath the simple translation equivalents can lie very different culturally-based expectations and meaning patterns. Thus, rather than comparing dictionaries, it might be more profitable to compare language using corpora, and then build new dictionaries that can handle these cultural expectations.

## **5. The way forward for data. Can corpora cope better?**

Having looked at *art* in dictionaries, we can now turn to what corpora offer so as to see whether not only the picture given in published works will need updating, but also whether new data sources will require a new look at user needs and thus new dictionaries. To do this we shall be using corpora in both French and English so as to carry out a brief pilot study rather than a full exhaustive analysis of available data.

Since the advent of the COBUILD project in the nineteen eighties, corpora have revolutionised attitudes towards dictionaries sources, and also in the representation of that data. The COBUILD outlook introduced new ways of handling lexical and grammatical data and new ways of defining (Sinclair et al. 1987) as well as providing analytical methodologies and opening up new lines of research for the entire corpus linguistics community.

Dictionaries that make use of corpora rely on large balanced reference corpora. Standard reference corpora for languages should be built following clear criteria so as to allow representativity (Sinclair 2005). This is the case with balanced corpora such as the British National Corpus. This is fine for studies within a single language, but problems invariably arise when comparing languages. Two problems can be cited; one is that criteria will differ from language to language, which means that corpora in different languages may not be fully comparable (Williams, forthcoming), the other is that not all languages have reference corpora. French, for example, does not.

One way of getting around the representativity issue may be recourse to internet extracted corpora insofar as these can be very large and constantly kept up to date. However, these do suffer from numerous problems as they represent only what is available, which means that vast amounts of written and oral data, and all electronic texts in closed archives, are simply not accessible. Another difficulty is that they no real selection procedures, they are literally what is available at a given time with no means of controlling either the age or source of the data. Nevertheless, despite these

## Art for dictionaries sake: Comparing cultural outlooks<sup>1</sup> through dictionaries and corpora

drawbacks, it has been argued that their sheer size may help in gaining some form of representation by ironing out variables in spelling and non-native speaker grammatical errors. The usefulness of size has been argued by Church (2010) who accepts that these may be the corpora of the future for large-scale generalisations about language. Nevertheless, insofar as the devil is in the detail, we may yet come back to well constructed corpora so as to get a closer knowledge of language, but for the moment those with the power and financial resources to create large corpora are turning towards internet.

Consequently, in order to carry out this study, two web corpora have been used, the Ukwac (UK web as corpus) corpus which has 1 565 274 190 running words and its French counterpart, Frwac at 1 628 667 738 running words. These are thus similar in size and also in collection method, if not selection, procedures. Both are to be found amongst the selection of corpora proposed for use with Sketch Engine<sup>1</sup>, which is that analytical tool adopted here.

What corpus linguistics as expounded by John Sinclair has proven is that collocation lies at the heart of language. Thus, collocation will be the main exploratory tool applied here, notably used the theories of collocational networks and collocational resonance (Williams 1998, 2007, 2008a). These theories draw on a contextualist approach, rather than a traditional phraseological one that tends to concentrate on formulae studies out of context. Given that this study is looking at both dictionaries and corpora, it is importance to consider these two approaches to collocation as both influence lexicographical production, but in different ways. The central question remains collocation, but the starting points are different. However, far from being in opposition, the two approaches can be seen as clearly complementary.

### ***5.1 Collocation: A tale of two traditions***

Historically two main traditions in collocation studies can be found; lexicographical and contextualist collocation. These have their beginnings largely in a tradition going back to Palmer (1933) for the former and a contextualist collocation approach stemming from Firth (1957) for the latter.

If, in corpus linguistics, Firth is generally attributed the fatherhood of collocation studies, it is because his approach was taken up and developed by numerous linguistics working in the Sinclairian tradition, whereas the major work of Palmer, the Second Interim Report on English Collocations

---

<sup>1</sup> <http://www.sketchengine.co.uk/>

(1933), remained largely inaccessible and, hence, forgotten. Nevertheless, the two approaches look at similar language data, but not at all from the same perspective. That words go together is hardly new, without speaking of collocation as such, Dr Johnson knew that such information would be vital in his dictionary when he pointed out that:

The syntax of this language is too inconstant to be reduced to rules, and can only be learned by the distinct consideration of particular words as they are used by the best authors. Thus, we say according to the present modes of speech, the soldier died of his wounds, and the sailor perished with hunger; and every man acquainted with our language would be offended with a change of these particles, which yet seem originally assigned by chance, there be no reason to be drawn from language why a man may not, with equal propriety, be said to die with a wound or perish of hunger. (Johnson 1747/2008) : 25)

In the Second Interim Report on English Collocations, Palmer made the first systematic effort to classify collocations. Most of his categories would now be termed lexical phrases or fixed expressions and idioms (Moon 1998) as he lists any all multiword formulae, largely by syntactic criteria. He defines collocation as:

a succession of two or more words that must be learnt as an integral whole and not pieced together from its component parts.

Although Palmer's book largely disappeared from circulation, it has been much quoted and his ideas greatly inspired the developing phraseological schools, notably in central and eastern Europe. His ideas certainly found a place in Hornby's Oxford Advanced Learner's Dictionary, first published in 1949. His ideas also helped shape a phraseological tradition developed by Hausmann and others who privileged syntactic criteria in the locating and listing of collocations for dictionary usage. Another tradition developed in Russia and led to Igor Mel'čuk's *Théorie Sens Texte* and the lexical functions, that include collocation. The latter breaks free of simple syntactic patterning to discuss the semantic roles played by collocations.

It could be said that the Palmerian phraseological tradition represents collocation as a static phenomenon in that collocations are found, either in texts or by intuition, and described. There is a tendency to seek what is fixed and invariable, rather than the everyday wealth of collocation found in corpora. Apart from Mel'čuk's lexical functions, little attempt is made to go beyond classifying through formal linguistic criteria. On the other hand, with Firth and the contextualist tradition, we have a much more dynamic approach based on the key notion that "You shall know a word by the company it keeps" (Firth 1957).

## Art for dictionaries sake: Comparing cultural outlooks <sup>1</sup> through dictionaries and corpora

The Firthian approach places collocation and meaning firmly within context. Outside of context there can be no real meaning as “words do not have meaning, meanings have words! (op.cit.).

Firth’s writings were largely literary in outlook, and have indeed inspired the stylistics analyses developed by Louw (1993, 2000) in semantic prosody and other analyses. However, Firth’s insights were taken up and adapted to computer analysis of corpora by John Sinclair, thereby founding an entirely new approach to collocation. Disarmingly simple to read, but extremely profound in its repercussions on language studies, Sinclair’s *Corpus, Concordance, Collocation* (1991) is probably by far the most important work in modern language studies placing as it does collocation at the centre of language.

Very many important studies have stemmed from Sinclair’s insights, in addition to semantic prosody, Hoey’s notion of lexical priming (2005) which seeks to explain how we combine words to express thought, and also how we learn language to start with can also be cited as bringing to the fore fundamental questions about language. In many ways, primings are akin to the norms and exploitations described by Hanks (forthcoming) in a revolutionary new approach to language and dictionary building. Sinclairian notions of collocation are also the basis of the two approaches used in this study; collocational Networks as the widening lexical environment of lexical items and collocational resonance, which describes the transfer of meaning from one textual environment to another. These two methodologies have been brought together in experimental dictionary building in the form of organic dictionaries. The aim of organic dictionary building (Williams and Millon) is to bring the phraseological and contextualist traditions together so that collocational networks can be used to explore data and provide a means for for headword extraction and navigation in a dictionary in an attempt to overcome the limits of alphabetisation. Networks provide a means of applying dynamic collocation and preparing the ground for the phraseological knowledge to be included in the dictionary.

### **6. The corpus picture. What is art?**

The first step will be to build a local network for art from the Ukwac corpus. This is done by using the statistical collocations using a word sketch. As can be seen in figure one, there are aspects of *art* and related multiword units that did not appear in the dictionaries discussed earlier.

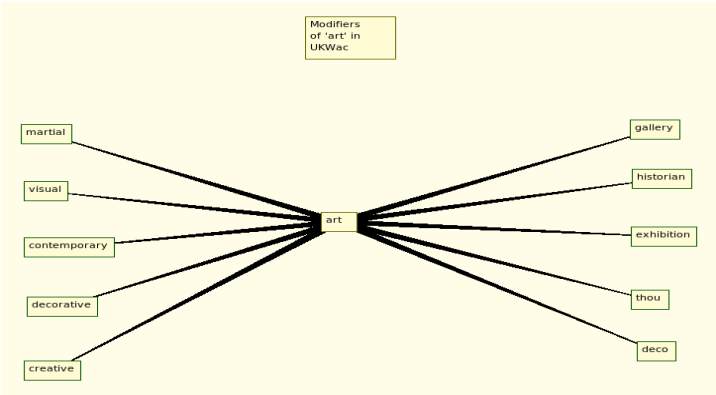


Figure one. Modifiers of art in UKwac

At a first level, we can see a series of lexical units that on the right-hand side are easy to enter in a dictionary in alphabetical format, provided we get over the dispersal problem, and and on the left those units which will require some other form of anchoring. In addition, we have the multiword unit *martial art(s)*, which do not fall under the creative functions seen so far. At a second level, each of these collocates could be used to start a network leading to en enriched lexicon. What are the collocates of martial, visual etc.? How do these link with other words in the growing headword list? This is very much the basis of an organic dictionary. One problem will be the effects of lemmatisation also causes difficulties, which makes it necessary to look at separate networks for the singular and the plural forms.

Art	museum	Arts	fine
	works		centre
	gallery		martial
	fine		visual
	contemporary		crafts

## Art for dictionaries sake: Comparing cultural outlooks 2 through dictionaries and corpora

	martial		sciences
--	---------	--	----------

Table one: Immediate collocates of *art* and *arts*

Some collocates will be common to both singular and plural forms, thus we talk of *martial arts* in general, but also a *martial art*. We also have a combined form with *arts and crafts*, a formulae found in the English dictionaries to describe a particular movement. *Art gallery*, *art historian* and *the fine arts* were not found in the dictionaries consulted, which means that these might need enriching. It can be noted that the definite article that appears with *fine arts* needs to be included as it clearly designates a particulate grouping of art types.. This is interesting in itself, but is even more so when we look at the French corpus using the same methodology and compare results (tables three and four).

Art	museum	Art	contemporain
	works		plastiques
	gallery		histoire
	fine		martiaux
	contemporary		oeuvres
	martial		musées

Table two: Immediate collocates of *art in English and in French*

Arts	fine	Arts	plastiques
	centre		martiaux
	martial		visuels
	visual		métiers
	crafts		beaux
	sciences		décoratifs

Table three: Immediate collocates of *arts in English and in French*

If we now start linking the findings from the English and French corpora, a number of units that are apparently easily translated can be found (table four). Thus, in bringing together the data from two different datasets, we

have not only enriched two monolingual dictionaries, but also built the basis of a bilingual one.

Fine arts	Beaux arts
Martial arts	Arts martiaux
Arts and crafts	Arts et métiers
Visual arts	Arts visuel
Art works	Oeuvres d'art
...	...

Table four: Simple translation equivalents

However, as with *culture*, this simple picture is too simple. Behind each translation equivalence lies significant cultural variations in interpretation. We are not necessarily discussing the same thing.

We can start with the apparently odd man out in table four, *martial arts*. This is a case of agreement, on the whole *martial arts* refers to Japanese combat skills that have become popular sports in the west. The underlying element is the notion of skill, rather than human creativity, skill at potentially neutralising other humans. The interesting point is that in our western culture, we have similar traditions of combat skills going back to ancient Greek, providing the inspiration for the Olympic games. Agreement between French and English on *martial arts* comes about because the concept is of relatively recent in Western Europe.

Another apparently simple translation equivalent is arts and crafts / arts et métiers. These both refer to skills, but artistic skills in producing beautiful objects or buildings. This is where priming comes in to play. In English, *arts and crafts* primes *movement* as collocate, which in turn conjures up images of William Morris and the architectural, decorative elements and furniture associated with this circle. This was renewal of the the notion of craftsmanship, amongst which book binding and carpentry were included. In French, *arts et métiers* refers to two entities, a museum and a prestigious engineering school. In the first case, the term collocates with *Musée*, and concerns the museum that houses a collection of scientific instruments belonging to the *Conservatoire National des Arts et Métiers* (National Conservatory of Arts and Industry). This *conservatoire* pre-dates the Arts and Crafts movement, and also has very different origins and terms of reference. The second primed reference is to the engineering school and refers directly to where a given engineer was trained. The full

## Art for dictionaries sake: Comparing cultural outlooks 2 through dictionaries and corpora

name of the institution was the *École nationale supérieure d'arts et métiers*, it is now called *Arts et Métiers ParisTech*, but to most French people, it is still *Arts Et Métiers*. The skills of fine furniture making and carpentry do not enter under their terms of reference as the school trains scientists and engineers. Simple translation, but totally different terms of reference.

The *Fine Arts / Beaux arts* formulae brings us fully to human creativity. *Fine Art* concerns the creation of objects that serve only for aesthetic appreciation and require visual appreciation, hence *visual art* as another related term. *Beaux Arts*, like *Arts et Métiers* tends to collocation with a higher education institution, the *École Nationale Supérieur des Beaux Arts*, but also to large municipal art galleries. Part of what is taught in the *Beaux arts* covers aspects of *Arts and Crafts*, but by no means all. Once again, surface similarity leads to different concepts.

The corpus then provides vital information as to both similarity and differences in usage. This information can then find its way into a bilingual dictionary that goes beyond simple translation equivalents. A corpus-based dictionary must show translation equivalents, but, to be truly usable, these require supplementing with a lot of cultural and encyclopaedic information.

### 7. The way forward

So how far have we come and what is left to be done? So far we have seen the advantages and drawbacks of paper dictionaries and also what the use of corpora can bring. We now need to see how new technological and the accumulated knowledge of centuries of dictionary making can be harnessed to take the art and craft of lexicography forward without sacrificing its values.

By having a corpus based dictionary, we have removed some of the problems encountered earlier in terms of missing data and updating senses in line with current usage. Defining traditions and cultural expectations behind what we expect from a dictionaries still differ, but we now do have a common base from which to build, one that is based on reality of usage and frequency of use. We have also a means not only of enriching a lexicon, but a tool to go beyond alphabetical organisation and the artificial groupings into imposed semantic categories. We have also found a means for comparing usage across languages and to locate real translation equivalents.

There remains however a problem, one that the use of comparative corpora has clearly shown. How can we be certain that the associated meanings and prosodies are the same in each language?

The answer is that we cannot. If we have different cultural expectations and backgrounds, we also impose different meanings, that is what cultural and language variety is about; a wealth of experience. Thus, if we cannot simply translate, we can explain. Prototypes are one tool to allow comparison, another one to come from the Sinclairian toolbox is the bridge dictionary, a dictionary designed to bridge gaps in understanding.

Bridge dictionaries (Williams 2008b) were a means to bridge the gap between languages by using the COBUILD dictionary as a hub. The idea is simple, a full sentence definition provides an understandable sentence explaining a meaning. The sentence can thus be translated as shown in the examples below for *nail*.

English

(1) A nail is a small piece of metal with a sharp end which you hit with a hammer in order to push it in to something.

Portuguese

COUNT N A nail é uma pequena haste de metal com um lado pontiagudo ena qual se bate com um martelo para crava-la em algo. Veja a figura

Czech

(1) Nail je malykousek kovu se s̃ picatym koncem, na který uhodí me klavidem,abychom ho do neceho zarazili.

Now imagine that we build a dictionary definition in the same way for *clou* (nail) in French, and then translate it. In this way we go beyond simply giving a list of words in the lexicon, we can endeavour to show how they can be used to give meanings in context. With such an approach, the real primings and expectations can be made clear showing the similarities and differences between, for example, *fine arts* and *beaux arts*, and all the other apparently simple equivalences that make up an encoding oriented bilingual dictionary.

## 8 Conclusion

Our starting point was the dictionary, essential reference tools that we have grown up with and learned to love. Being so close, we have not seen them grow old with a need of being rejuvenated without losing their soul. Dictionaries provide a wealth of knowledge, and that knowledge is largely cultural, so it is vital that users have the means to access this information and that it is kept up to date.

Dictionaries are human artefacts, so they vary enormously. There are no bad big commercial dictionaries, but they are not necessarily adapted to all

## Art for dictionaries sake: Comparing cultural outlooks 2 through dictionaries and corpora

situations. A first step in rethinking dictionaries is to give users the means to judge and draw the information they seek from the best source for their immediate need. The biggest challenge is to go how to go further in creating dictionaries for the 21<sup>st</sup> century. Paper dictionaries are not dead, far from it, but due to costs, many publishing houses are likely to push towards electronic dictionaries. One problem of this will be getting users to pay when there are free, but not necessarily good, dictionaries on the web. The difference will come from rethinking dictionaries and bringing in more of the of the information carried by a corpus.

Organic dictionaries are an attempt to let the corpus speak by building a word list from statistical significant relationships discovered in the corpus. This implies using collocational networks for the extraction and management of lexis, and then using collocational resonance to show language change through the use of prototypes to manage sense variation. The Bridge dictionaries were not a commercial success, but they did provide a unique means to transfer real meanings across language and across contexts. They were probably simply too early, many breakthrough require time before there full value is recognised, and lexicography is a remarkably conservative discipline. In the general rethink, the advantages of explaining rather than defining may become essential in building new bilinguals that will give access to the corpus and explained access to the underlying cultural environments.

Having a general rethink is necessary. English being used as a world language can give a false impression that the cultural assumptions across the world are ironed out. Dictionaries are eminently cultural tools that will have to show the meaning potentials of lexis in different environments. Beyond definitions of senses, we need to raise awareness of the cultural assumptions that are hidden within entries and bring forward the primings and prosodies underlying different Englishes, let alone their equivalences in other languages. It also means recognising that other languages carry different assumptions and reaffirming the wealth of language variety. The dictionaries of both major and lesser-used languages demonstrate the wealth of human experience, they are a storehouse of the knowledge that makes us human.

Dictionaries are a vast repository of human knowledge and experience. Users need to be taught to appreciate dictionaries, and demand more from them whilst accepting that they do not show reality but simply reflect a reality, like fine art.

## References

### *Dictionaries*

- Johnson, Samuel. 1755. *Dictionary of the English Language*  
 Merlet, Philippe. 2005, *Le Petit Larousse*. Paris: Larousse.  
 Soanes, Catherine. Stevenson, Angus. 2005. *Oxford Dictionary of English* (ODE). 2nd edition (revised).. Oxford. Oxford University Press.  
 Simpson John. 2009, *Oxford English Dictionary* OED. CD-ROM Version 4. Oxford: Oxford University Press  
 Sinclair, John. McH. 2003, *Collins Cobuild English Language Dictionary*. (COBUILD) 4th edition 2003. Glasgow: HarperCollins.  
 Wehmeier, Sally. 2005. *Oxford Advanced Learner's Dictionary*. (OALD). 7th edition. Oxford. Oxford University Press.  
 Rundell, Michael. 2005. *MacMillan English Dictionary for Advanced Learners* (MEDAL). Oxford: MacMillan

### *Others*

- Béjoint, Henri. 2010. *The Lexicography of English*. Oxford: Oxford University Press.  
 Church, Kenneth W. 2010; 'More is more' in de Schyver, G-M. 135-142.  
 Cowie, Anthony P. (ed) 2009. *The Oxford History of English Lexicography*. 2 vols. Oxford: Oxford University Press.  
 De Schryver, Gilles-Maurice. (ed) (2010). *A way with words: Recent advances in lexical theory and analysis, a Festschrift for Patrick Hanks*. Kampala: Menha Publishers.  
 Firth, John R. 1957. *Papers in Linguistics 1934-1951*. Oxford: OUP  
 Fontenelle, Thierry. 2008. *Practical Lexicography: A Reader*. Oxford: Oxford University Press.  
 Hanks Patrick. 1987. Definitions and Explanations. In Sinclair J. (ed) 1987. *Looking Up : An Account of the COBUILD Project in Lexical Computing*. London : Collins. 116-136.  
 Hanks, Patrick. 1994. Linguistic Norms and Pragmatic Exploitations or, Why Lexicographers Need Prototype Theory, and Vice Verse. *Papers in Computational Lexicography: Complex 94*. 89-113.  
 Hanks, Patrick. 2000. Do word meanings exist?. In Kilgarriff, Adam , Palmer, Martha (eds). 2000. *Senseval: Evaluating Word Sense*

## Art for dictionaries sake: Comparing cultural outlooks 2 through dictionaries and corpora

- Disambiguation Programmes. Computers and the humanities*, 34/1-2 pp. 205-215.
- Hanks, Patrick. (forthcoming). *Lexical Analysis: norms and exploitations*.
- Hoey, Michael. 2005. Lexical Priming: A New Theory of Words and Language. London: Routledge.
- Johnson, Samuel. 1747/2008. 'The plan of a dictionary of the English language'. In Fontenelle, Thierry (ed) 19-30.
- Louw, Williams. 1993. Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In M. Baker, 1993. *Text and Technology*. Amsterdam: John Benjamins, 157-76.
- Louw, Williams. 2000/2008. Contextual Prosody Theory: bringing Semantic Prosodies to Life. In Heffer, Christopher. & Sauntson, Helen. (eds) (2000) *Words in Context: A Tribute to John Sinclair on his Retirement*. CD-ROM: English Language Research Discourse Analysis Monograph No. 18. Reprinted in online journal *Texto* (2008): <http://www.revue-texto.net/index.php?id=124>.
- Moon Rosamund. 1998. *Fixed Expressions & Idioms in English*. Oxford : Clarendon Press.
- Palmer, Harold E. 1933. *Second Interim Report on English Collocations*. Tokyo: Kaitakusha
- Pruvost, Jean. (ed) 2005. *Les dictionnaires Larousse: Génèse et évolution*. Montréal: Les Presses de l'Université de Montréal.
- Pruvost Jean. 2006. *Les dictionnaires français: outils d'une langue et d'une culture*. Paris: Ophrys.
- Rundell, Michael. 2006. More than one way to skin a cat: Why Full Sentence Definitions have not been universally adopted. In Corini, E., Marelllo, Carla., Onesti, Cristina. (eds), *Proceedings of the XIIIth EURALEX International Congress*. Turin: Università di Torino, 323-338. Reprinted in Fontenelle, T. 197-209.
- Sinclair, J McH, (ed) 1987. Looking Up: an account of the COBUILD Project in Lexical Computing. London: Collins
- Sinclair, John McH. 2005. Corpus and Text: Basic principles. In Martin Wynne (ed.), 1-16.
- Williams Geoffrey. (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 Geoffrey. 2007. In search of representativity in specialised corpora: categorisation through collocation. In Teubert Wolfgang & Krishnamurthy Ramesh (eds). *Critical Concepts in Corpus Linguistics*. London : Routledge. Vol.3 : 324-346.

- Williams Geoffrey. 2008a. 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.
- Williams Geoffrey. 2008b 'A multilingual matter:Sinclair and the bilingual dictionary'. *International Journal of Lexicography*. Vol. 21:3. pp 255-266.
- Williams Geoffrey. (forthcoming) A cultivated audience: Comparable corpora and cross language collocation. In Bondi Marina. Bondi, M. Cacchiani S & Palumbo G eds. "Corpus Linguistics and Language Variation", special issue of RILA (Rassegna Italiana di Linguistica Applicata) Bulzoni.
- Williams Geoffrey and Millon Chrystel. 2009. The General and the Specific: Collocational resonance of scientific language. *Proceedings of Corpus Linguistics* 2009. <http://ucrel.lancs.ac.uk/publications/cl2009/#papers>
- Williams, Geoffrey., Piazza, Roberta;, and Giuliani, Delphine Forthcoming. Nation and Supernation: A tale of three Europes, in Bayley, Paul and Williams, Geoffrey (eds) *European Identity: What the Media Say*. Oxford: Oxford University Press.
- Wynne, Martin (ed.) (2005), *Developing Linguistic Corpora: A guide to good practice*. Oxford: Oxbow Books. Also online at <http://ahds.ac.uk/linguistic-corpora>, Last accessed November 2010.
- Yong, Heming & Peng, Jing. 2008. *Chinese Lexicography: A history from 1046 BC to AD 1911*. Oxford: Oxford University Press.

## Appendices

	ODE	OALD	COBUILD	MEDAL
Art	Art	Art	Art	Art
art	archaic	X	X	art
Art	Abbrev.	X	X	X
	2 entries	X	X	X
Art deco				
Art director	X		X	
	17 entries	6 entries	4 entries	5 entries

## Art for dictionaries sake: Comparing cultural outlooks <sup>2</sup> through dictionaries and corpora

Art form				
	1 entry	1 entry	2 entries	2 entries
Art gallery	X		X	
Art history			X	X
Art house				
	40 entries	29 entries	23 entries	24 entries
Art nouveau				
	2 entries	X	2 entries	X
arts and craft			X	
Arts and Crafts Movement			X	X
Arts council, the	X	X	X	
	X	2 entries	X	1 entry
Art therapy			X	X

Appendix one. A comparative table for art and related words

### Abstract

Dictionaries are repositories of cultural knowledge and assumptions, and yet their wealth is rarely exploited and understood by users who take them for granted and have rarely been taught to appreciate the wealth within. Corpus linguistic has contributed much to modern lexicography. It allows linguistics to explore language in context and explore the often surprising data that comes out of the text and that takes us beyond intuition.

This paper looks at two words that seem easily understood and easily interested. It starts by looking at how dictionaries are generally perceived and the underlying difficulties in handling data. It then goes on to discuss how ‘art’ is handled in different dictionaries in English and French and then what corpora in English and French reveal about this word in context. Finally the paper calls for a rethink of dictionaries so as to integrate the wealth of data held in corpora so as to provide a bridge between cultures.

**Bio data**

Geoffrey Williams is Professor of linguistics at the Université de Bretagne-Sud, France, director of the department of document management and currently Pro Vice Chancellor for International Relations. He is specialised in corpus linguistics and its applications in the field of special language lexicography. He has published widely internationally, notably in the International Journal of Corpus Lexicography, of which he is a member of the editorial board, and the International Journal of lexicography. He is currently President of the French association for applied linguistics (AFLA) and President of the European Association for Lexicography (EURALEX).