

THE VERY IDEA! A CORPUS-BASED COMPARISON OF IDEA, CONCEPT AND NOTION AND THEIR FORMAL EQUIVALENTS IN SPANISH

Christopher S. Butler
Swansea University
csbutler@ntlworld.com

The present article uses corpus-based analysis in order to explore the similarities and differences shown by the nouns IDEA, CONCEPT and NOTION in English and, secondarily, between these nouns and their formal equivalents IDEA, CONCEPTO and NOCIÓN in Spanish. Three areas of investigation are highlighted: the frequencies of the three nouns across four registers of each language; the adjectival collocations into which the three words enter; and the occurrence of some idiomatic constructions built around the words under scrutiny. The results are of interest not only in their own right but also in the context of applications to, for instance, translation and the teaching and learning of English and Spanish.

Key words: corpus, contrastive linguistics, register, collocation, English, Spanish

1. Introduction: aims and methodology

The aim of this article is to exploit the power of corpus-based analysis in order to reveal differences, as well as similarities, between lexical items which appear to be closely related.¹ More concretely, I shall examine some properties of the nouns IDEA, CONCEPT and NOTION in English and between these nouns and their formal equivalents IDEA, CONCEPTO and NOCIÓN in Spanish, including their distribution across four registers of

¹ This article is based on a presentation given at the 5th International Conference on Contrastive Linguistics, Leuven, July 2008. I am grateful to members of the audience at that conference, also to anonymous reviewers, for their comments. The work reported here was done under the auspices of two projects: 'Estudio comparativo de la interfaz gramática-discurso en lengua inglesa con especial referencia a la coherencia y a la subjetividad' (A comparative perspective on the grammar-discourse interface in English, with special reference to coherence and subjectivity), based in Santiago de Compostela and financed by the Spanish Ministry of Education and Science (HUM2007-62220); and 'Corpus linguistics and online contrastive analysis (English-Spanish)', carried out by Research Group 930175 of the Universidad Complutense de Madrid with a research grant from the Comunidad Autónoma de Madrid and the Universidad Complutense.

each language: spoken, fiction, news and academic writing.² The main focus will be on the English data, the Spanish data being examined rather more cursorily.

If we look at dictionary definitions of these items, we often find that they are circular. Consider, for example, the following definitions, taken from two online dictionaries:

Compact Oxford Dictionary online³

idea

• **noun** **1** a thought or suggestion about a possible course of action. **2** a mental impression. **3** a belief. **4 (the idea)** the aim or purpose.
— origin Greek, ‘form, pattern’.

concept

• **noun** **1** an abstract idea. **2** an idea to help sell or publicize a commodity.
— ORIGIN Latin *conceptum* ‘something conceived’.

notion

• **noun** **1** a concept or belief. **2** an impulse or desire. **3** a vague awareness or understanding. **4 (notions)** chiefly N. Amer. items used in sewing, such as buttons and pins.
— ORIGIN Latin, ‘idea’.

Collins English Dictionary PRO online⁴

idea NOUN

- 1 any product of mental activity, thought
- 2 a scheme, intention, or plan
- 3 the thought of something ◇ the idea excites me
- 4 a belief or opinion
- 5 a vague notion, inkling ◇ they had no idea of the severity of my injuries
- 6 a person's conception of something ◇ his idea of integrity is not the same as mine
- 7 aim or purpose ◇ the idea is to economize on transport
- 8 (Philosophy) (in Plato) a universal model of which all things in the same class are only imperfect imitations

concept NOUN

an abstract or general idea ◇ one of the basic concepts of quantum theory

notion NOUN

- 1 an idea or opinion
- 2 a whim

² In accordance with convention, small capitals are used to indicate lemmas rather than word forms.

³ <http://www.askoxford.com/dictionaries>, accessed 21.7.2008

⁴ <http://www.collinslanguage.com/shop/english-pro.aspx>, accessed 21.7.2008

The word CONCEPT is defined in terms of IDEA, and NOTION in terms of CONCEPT (in the Compact Oxford Dictionary) or IDEA (in the Collins English Dictionary PRO), suggesting a certain degree of equivalence, although there are some pointers towards differences too (e.g. the emphasis on abstraction for CONCEPT, and the insubstantial element for NOTION implied by the use of ‘vague’ in the Compact Oxford Dictionary and ‘whim’ in the Collins English Dictionary PRO). Indeed, it is easy to find corpus examples in which the three words in each language appear to be used more or less interchangeably:

- (1) The idea of a pioneer was a very important one for early Christians. (BNC ABV 1450)⁵
- (2) The notion of a network is fundamental to hypertext. (BNC CGA 23)
- (3) The concept of a university doctorate has always been clear. (BNC HPN 9)
- (4) *La idea de un “socialismo liberal” es una contradicción en los términos.*
‘The idea of a “liberal socialism” is a contradiction in terms.’ (*Corpus del Español*, Section 19-N, España ABC)
- (5) *El concepto de defensa estratégica ha sufrido un profundo cambio durante el siglo XX.*
‘The concept of strategic defence has suffered a profound change during the 20th century.’ (*Corpus del Español*, Section 19-A, Encarta: Sistemas de defensa)
- (6) *La noción de Dios que me proponen las religiones me resulta ininteligible.*
‘The notion of God that religions propose is unintelligible to me.’ (*Corpus del Español*, section 19-F, Manuel de historia)

In each of these cases, the noun under scrutiny is in the syntactic context *the idea/concept/notion of a* NOUN, or *la idea/el concepto/la noción de (un(a))* NOUN. However, as we shall see, corpus analysis demonstrates that there are interesting differences, as well as similarities, across the three lexemes in each language, and between languages.

For English, the British National Corpus (World Edition) was accessed through Mark Davies’s BNC website (see fn. 5). In the present study, only four sections of the corpus were used, namely those containing material from spoken language, newspaper writing, fictional writing and academic writing. For Spanish, the Corpus of Español, again available online (see fn. 5), was used. This corpus contains material from the 12th to the 20th century, but only the 20th century component, again consisting of spoken language, news, fiction and academic writing, from both Spain and Latin America, was taken for analysis. The composition of the corpora is given in Table 1. Concordances and lists of collocates were produced using the software provided at each of the websites.

⁵ Examples marked BNC are taken from the British National Corpus (World Edition) (see <http://www.natcorp.ox.ac.uk/>), while those marked *Corpus del Español* are taken from the Spanish corpus compiled by Mark Davies, available at <http://www.corpusdelespanol.org/>.

Corpus	Words (millions)				
	Total	Spoken	Fiction	News	Academic
BNC World	100	9.96	15.91	10.47	15.33
Corpus del Español	100 (approx. one fifth from 1900s)	5.68	5.72	5.72	5.71

Table 1: The corpora used in the study

2. Frequencies

First let us look at the frequencies of the three nouns in each language, both as a whole and in the four separate registers. The data for English are given in Table 2.

Lemma	Total	Spoken	Fiction	News	Academic
IDEA	16987 [328.76]	3343 [335.64]	5071 [318.75]	2144 [204.78]	6429 [419.37]
CONCEPT	4666 [90.30]	263 [26.40]	151 [9.49]	211 [20.15]	4041 [263.60]
NOTION	2856 [55.27]	117 [11.75]	338 [21.24]	166 [15.85]	2235 [145.79]

Figures in square brackets are frequencies per million words.

Table 2: Frequencies of the lemmas IDEA, CONCEPT and NOTION in the spoken, fiction, newspaper and academic sections of the British National Corpus (World Edition)

The lemma frequencies were calculated by summing those for the separate singular and plural forms. It can be seen that IDEA is much more frequent, in terms of occurrences per million words, than the other two nouns, and that CONCEPT is more frequent than NOTION. In all four registers, IDEA is the most frequent; the least common is NOTION for spoken, news and academic, but CONCEPT for fiction. The academic register has the highest frequency of each of the three nouns; IDEA is least frequent in news writing, CONCEPT in fiction and NOTION in spoken language. Analysis of the distribution of raw frequencies across the four registers and three nouns by means of the chi square test reveals that the differences are highly significant ($\chi^2 = 4458$, $df = 6$, $p = 0.0000$) and that, taking the three nouns in relation to each other, IDEA occurs less frequently in the academic register and more in the other three categories than would be predicted from the null hypothesis⁶ of no association between noun and register, while CONCEPT and

⁶ The null hypothesis in a statistical test is that there is no difference between the values under test (here, the distribution of different categories across registers), in the populations from which the samples were drawn, the differences observed in the samples being due to sampling variation. Usually the investigator wants to reject the null hypothesis in favour of the alternative hypothesis that there is indeed a difference between the population values under study.

NOTION occur far more frequently in academic language than the null hypothesis would predict and less frequently in the other three registers, particularly fiction.

Comparable data for Spanish are shown in Table 3. Once more, in terms of normalised frequencies, IDEA is the most common overall, and CONCEPTO is more frequent than NOCIÓN. All four registers individually mirror this overall distribution of frequencies. IDEA is highest in spoken language, lowest in news, while CONCEPTO occurs most often in academic language and least in fiction, and NOCIÓN most frequently in fiction and academic language and least in news. Chi square analysis of the raw frequencies again gives a highly significant result ($\chi^2 = 569$, $df = 6$, $p = 0.0000$), and shows that when the three words are compared among themselves, IDEA is somewhat less common in academic writing and news than would be expected from the null hypothesis but more frequent in spoken language and fiction, while CONCEPTO is rather more frequent than expected in academic language and news but less common in spoken language and especially in fiction, and NOCIÓN is more frequent in fiction and less in spoken language and news.

Lemma	Total	Spoken	Fiction	News	Academic
IDEA	5800 [254.16]	1798 [316.55]	1406 [245.80]	1211 [211.71]	1385 [242.56]
CONCEPTO	1910 [83.70]	424 [74.65]	111 [19.41]	503 [87.94]	872 [152.71]
NOCIÓN	263 [11.52]	56 [9.86]	89 [15.56]	41 [7.17]	77 [13.49]

Figures in square brackets are frequencies per million words.

Table 3: Frequencies of the lemmas IDEA, CONCEPTO and NOCIÓN in the spoken, fiction, newspaper and academic sections of the *Corpus del Español*

The most salient points of comparison between the two languages are as follows:

- In the four registers taken together, the frequency ordering is IDEA > CONCEPT > NOTION, or the equivalent, for both languages.
- In terms of normalised frequencies, NOCIÓN is considerably less frequent than NOTION, and this is due to the relatively high occurrence of NOTION in English academic texts.
- When the three nouns are compared among themselves, IDEA in both languages shows a lower relative frequency in academic writing than expected from the null hypothesis of equality of distribution of the three items, whereas CONCEPT and CONCEPTO show an association with academic language, as does NOTION in English. Both CONCEPT and CONCEPTO are particularly infrequent in fiction.

Overall, then, there are striking similarities between the two languages as far as frequency is concerned, with some more minor differences. From the associations between noun and register, we may hypothesise that in both languages IDEA, although it

occurs frequently in all four registers, is a more general word favouring non-technical registers, while CONCEPT(O) is associated with the more specific mental constructs which are important in the language of professional activities. English NOTION appears to associate more with professional registers than Spanish NOCIÓN.

3. Adjectival collocation

With these preliminary results in mind, let us turn to a second type of property of the three nouns in each language. We might expect that useful information about the usage of each item would be revealed by a study of the adjectives with which it associates. In order to test this, the ability of the software to generate frequency data on collocates selected by word class was exploited. For English, word forms tagged as adjectives in the BNC and occurring at the position (L1) one to the left of the head noun were tabulated in descending order of collocational frequency. Lists were obtained for both the singular and the plural form of each noun, and a list of collocates for each lemma drawn up by including any collocate which occurred in the top 20 in the frequency list for the singular form, the plural form, or both. The distribution across lemmas can be seen in Table 4. Let us begin with the collocates which are common to all three nouns.

IDEA, CONCEPT, NOTION	IDEA only	CONCEPT only	NOTION only	IDEA, CONCEPT	IDEA, NOTION	CONCEPT, NOTION
whole	good	legal	accepted	Marxist	clear	traditional
new	bad	central	simple		brilliant	scientific
general	faintest	key	romantic		vague	fundamental
very	better	broader	popular			
basic	great	difficult	objective			
other	excellent	limited	simplistic			
different	rough	elusive	wider			
original	wrong	British	confused			
abstract	nice	particular	complex			
religious	fair	important	preconceived			
modern	splendid	relevant	strange			
theoretical	political	historical	conservative			
	certain	linguistic	current			
	radical	psychological	false			
	dominant		intuitive			
	fresh		older			
	interesting					
	western					
	clever					

Table 4: Distribution of adjectival collocates at L1 for English data

One interesting observation is the occurrence of *whole* and *very* in the list. In order to illustrate the characteristics of the combinations of these adjectives with our three nouns, I give below an unbiased sample of five concordance lines for each combination, selected by generating a concordance for the combination and then taking every n^{th} one, where n is the total divided by 5 and rounded down (for *whole idea*, the most frequent combination, 100 lines were generated and every 20th taken).

1. them? Well Bill stood up and was a bit negative really to the whole idea of parish constables, farm watch and everything else, it was quite disappointing There
2. and Mary Rose began to think about drapes and colour schemes. The whole idea was too romantic and the awful prospect of the Ryan farmhouse or its equivalent vanished
3. to her lips. "But if you're a feminist, the whole idea of masculine verb endings including the feminine is a complete joke," she protested
4. 1983, pp. 55-7): The report goes on to repudiate the whole idea of a "no first use" pledge, which might enable "the Soviets
5. only in recent years was there a widespread and fundamental questioning of the whole idea of dictatorship, and a reevaluation of peaceful, democratic forms of political change,
6. ways. Absolutely. Of course this is something for the future, the whole concept of hydrogen economy is well developed. When the economics become right, when the
7. be tough minded or sombre in mood to acknowledge the truth that the whole concept of A Retreat, in the older senses of the phrase, is now redundant
8. areas to coincide with National Foster Week, and a change in the whole concept of foster care. Family placement team members will be manning stalls at Alton market
9. not just upon the Christianity that precedes the reference, but on the whole concept of locality as somehow opposed to cosmopolitanism. We do well to remember that many
10. additional handicap. With a simple statement we have done away with the whole concept of deafness itself. Children are either capable of speaking or not; those who
11. these things. Do you think it's going to ultimately change the whole notion of publishing? Do you think that perhaps in due course publishing will move into
12. seating, designed primarily with public spaces in mind. If the whole notion of Designers' Saturday is altogether too much for you, or if your designer
13. and fewer dead-ends and irreversible choices than in the past, and the whole notion of "stages" has to be interpreted in the light of the individual's self-perception and the
14. inappropriate use of acute beds by younger patients to be raised. The whole notion of bed blocking seems to imply that older people enter hospital and then wilfully continue
15. and if necessary abandon a role altogether. Equally important, the whole notion of set role behaviour may come to seem problematic. In many situations, actors
16. anything, but --;" "Pregnant!" I screeched. The very idea ! The thought of the two of them fucking was bad enough; Lewis impregnating
17. director?" "Certainly not!" She was scandalised by the very idea , especially since the MD was a delightfully avuncular man, certainly old enough to
18. still convinced was operating in the club? Galvanised into action by the very idea , she rose abruptly to her feet, striving to keep her expression impassive as
19. was success in producing it. It is, rather, that the very idea of a body of knowledge about the world, of the sort we have now
20. This is precisely what the apparatus is designed to achieve. The very idea of the variable draws upon an analogy with the highly abstract structure of mathematics in
21. for entitlement to registration, since it is an indissociable principle of the very concept of establishment. 62. Part (c) of question (2),
22. process under review?" It may well be the case that the very concept of adjudication as applied to disputes between private individuals will have to be modified in

23. dialectical possibilities of expansion and compression in a double logic. Thus the very concept of totalization, distinguished from totality, must always be refused its prospective closure,
24. of red objects are included in the set. The account presupposes the very concept, redness, the acquisition of which it is meant to explain. It is
25. a phenomenological analysis seems to result from the logical pressures built into the very concept of ontological existent. As a result of these pressures we are forced to modify
26. he would look entirely at home among the mountains --; but the very notion of spending time with him was enough to horrify her. She met his eyes
27. , it encounters an overwhelming obstacle. It overlooks the fact that the very notion that the market is capable of legitimating power is widely acknowledged as being open to
28. been to lead poets to trust their ear implicitly, thus discrediting the very notion of measure. The closeness of Pound's and Eliot's collaboration when they were
29. to revise an existing literary theory, but in order to make possible the very notion of such a thing. One may hear echoes of this kind of project in
30. from syntactic theory; but to make things even more difficult, the very notion of semantic equivalence is a disputed one. Some linguists argue that such equivalence does

There are some clear generalisations we can make here. 24 out of the 30 examples contain the framework *the + whole/very + idea/concept/notion + of + nominal*, the nominal (which is sometimes a gerund) usually having an indefinite or zero article (note that in the one example, 20, where the definite article occurs, it has a generic meaning). Other possibilities include a structure in which the noun is postmodified by a *that* clause rather than an *of*-phrase (27), a definite phrase with *idea* standing alone (*The very idea!*, as in 16), a non-postmodified NP with the definite article acting as subject or object (2), or a prepositional phrase with *by* (17, 18). The function of the adjective is to emphasise, or focus upon, the mental construct, either by reference to its entirety (*whole*) or its inherent nature (*very*). But perhaps the most striking aspect of this set of examples is the presence, in the majority of cases, of something which clearly shows a negative attitude on the part of the speaker or writer. In some instances this takes the form of a lexical item or phrase which itself has a negative meaning: examples are *negative, too romantic, awful prospect, a complete joke, repudiate, questioning, redundant, too much for you, wilfully, problematic, scandalised, refused, horrify, discrediting, disputed*. In other examples where there are no strongly negative lexical items, the context still creates a negative impression: *the whole concept of foster care requires a change; the whole concept of deafness has been done away with; the whole concept of locality is somehow opposed to cosmopolitanism* (note the *somehow*); *the very concept of adjudication may have to be modified*; and so on. In yet other cases, consultation of the corpus itself reveals that the context beyond that given in the concordance line gives an indication of negative attitude: for instance, in the first example for *whole notion*, the later context shows that what is being moved into is a clearly undesirable situation in which nothing gets published.

This kind of phenomenon has been discussed under the label of *semantic prosody*. Since this concept has been the subject of considerable debate in the corpus linguistics

literature, it is worth exploring a little here. The term is usually attributed to a paper by Louw (1993), who defines it as “[a] consistent aura of meaning with which a form is imbued by its collocates” (Louw 1993: 157). Louw, however, refers to a personal communication dating from 1988, in which John Sinclair makes use of this label, using the term *prosody* in a similar way to Firth, who used the word in reference to phonological features which range across segmental boundaries. Partington (2004: 131) traces the origin of the study of semantic prosody even further back, to Sinclair’s (1987) demonstration that the items HAPPEN and SET IN are strongly associated with unpleasant events. Louw (1993: 172-73) warns us that “it should not be too readily assumed that semantic prosodies can be recovered with ease from corpora”, and that “[i]t may well turn out to be the case that semantic prosodies are less accessible through human intuition than most other phenomena to do with language”.

As Hunston (2007: 250) points out in her excellent discussion of views of semantic prosody, there are quite major differences in the way in which the term is used by different authors. We have seen that early work by Sinclair identified a negative prosody associated with certain lexical items in English. The examples given by Louw also centre around ‘good’ and ‘bad’ prosodies. Partington, who defines semantic prosody as “the spreading of connotational meaning beyond single word boundaries” (1998: 68), clearly sees the connotational meanings involved as “favourable or unfavourable” (1998: 66), and this is also true of later discussion (Partington 2004). Stubbs (1995) demonstrates that the lexical item CAUSE has a strong negative prosody, and later (Stubbs 2001: 65) that PROVIDE is associated with positive evaluation.⁷ On the other hand, Sinclair, in later work, recognises the possibility of semantic prosodies which are not simply ‘good’ or ‘bad’, but rather more specific in their interpretation. For instance, a sequence of words containing an indication of visibility plus a preposition (normally *with* or *to*, depending on the word class of the visibility word) plus *the naked eye* is found, in concordances, to be associated with a specific semantic prosody of difficulty, and the phrase *true feelings* with one of reluctance (Sinclair 2004: 30-36).

Mention of Sinclair’s later work brings us to another important difference in approaches to semantic prosody. As noted by Hunston (2007: 250), Partington’s discussion is concerned largely with the prosodies associated with individual words: in Partington (2004), these are individual members of the semantic set of words including, and related to, HAPPEN. Partington does, however, state that “the evaluative meaning of phrases involving SET IN is [...] communicated through the whole unit (consisting of item plus environment) rather than residing in just one of its parts” (2004: 132). Sinclair takes this idea further, associating semantic prosodies with an ‘extended unit of meaning’ which, in the case of *naked eye* is the whole complex of visibility + preposition + *the naked eye*, as we saw above.

Whitsitt (2005) sees the concept of semantic prosody as inherently problematic. He first isolates three different types of definition of the term. He paraphrases Louw’s

⁷ Stubbs (2001: 66) comments that *pragmatic prosody* would be a better term than *semantic prosody*, and he himself actually uses the designation *discourse prosody* in order to emphasise not only that this phenomenon is pragmatic, relating to speaker attitude, but also that it has an important function in the creation of discourse coherence.

definition in the following way: “if several different words all sharing the same semantic trait are frequently used with another word, meaning will be passed, over time, from that group of words to the other word” (Whitsitt 2005: 284). He notes that Sinclair’s comment that “[a] semantic prosody [...] is attitudinal, and on the pragmatic side of the semantics/pragmatics continuum” (Sinclair 2004: 34) casts doubt on the appropriateness of the term itself, and that Stubbs, in renaming the phenomenon as *pragmatic prosody* or *discourse prosody*, appears to agree. Finally, he discusses Partington’s use of the term, which equates it with connotation, and notes that Louw has rejected this usage, stating that “semantic prosodies are not merely connotational” (Louw 2000: 50), but that “[Louw] seems unable to offer any convincing argument demonstrating the difference between semantic prosody and connotation” (Whitsitt 2005: 286). The remainder of Whitsitt’s discussion is devoted to a detailed examination of Louw’s original treatment of semantic prosody, and finds it wanting. Among the aspects which Whitsitt finds most unsatisfactory is the claim that semantic prosodies carry over from one context to another, in such a way that a word or phrase may be interpreted positively or negatively even in contexts which do not themselves give any evidence of positive or negative evaluation. We shall, however, see some evidence for this idea later in our discussion.

Hunston, on the basis of careful analysis of concordance data, points out that even when some item frequently co-occurs with collocates indicating positive or negative evaluation, this is not so in all cases, but is a statistical tendency. She concludes: “... it may be argued that ascribing semantic prosody to a word is over-simplistic. The attitudinal meaning of a word may be altered by its immediate phraseology. The concepts of positive and negative evaluation may themselves be [*sic*] over-simplification, and in any case rest crucially on a notion of point of view” (Hunston 2007: 256). Hunston herself leans towards Sinclair’s view that semantic prosodies characterise the consistent discourse function of an extended unit of meaning consisting of the central item and a set of co-occurrences, and that if the surrounding phraseology changes, so may the semantic prosody.

Relating this view to our examples with *whole/very* plus *idea/concept/notion*, we may say that users of English commonly use this combination as part of an overall unit of meaning, whose precise limits may be difficult to define, which frequently contains expressions intended to have negative import, and whose function is to express the opinion that there is some problem associated with the mental construct under description. However, speakers and writers of English are also free to use such combinations simply to focus on the idea, concept or notion, without implying any associated negativity.

Before we leave the issue of modification by *whole/very*, it is interesting to note its distribution across registers, as shown in Table 5.

Sequence	Total	Spoken	Fiction	News	Academic
<i>whole idea</i>	108 [2.09]	34 [3.41]	37 [2.33]	14 [1.34]	23 [1.50]
<i>whole concept</i>	44 [0.85]	14 [1.41]	6 [0.38]	6 [0.57]	18 [1.17]
<i>whole notion</i>	12 [0.23]	2 [0.20]	0 [0.00]	3 [0.29]	7 [0.46]
<i>very idea</i>	56 [1.08]	1 [0.10]	34 [2.14]	1 [0.10]	20 [1.30]
<i>very concept</i>	23 [0.45]	0 [0.00]	1 [0.06]	0 [0.00]	22 [1.44]
<i>very notion</i>	40 [0.77]	1 [0.20]	10 [0.63]	3 [0.29]	25 [1.63]

Table 5: Frequencies of *whole/very* + noun in the 4 registers in BNC World

We see that on the basis of normalised frequencies, *whole idea* occurs in all four registers, being relatively more common in spoken language, while *whole concept* is found predominantly in spoken and academic language. *Whole notion* is rare: 7 of the 12 examples are in academic language. The combinations with *very* are almost confined to the academic and fiction registers: *very idea* is highest in fiction, *very concept* is found, with one exception, only in academic language, and *very notion* is also most frequent in this register.

Let us now move on to examine the other frequent adjectival collocates which are common to the three nouns. These can be seen as falling into a number of semantic classes:⁸

- generality: *general, basic*
- difference: *different, other*
- recency: *new, modern, original*
- domain: *religious, theoretical*
- abstraction: *abstract*

The whole set of shared collocates can be seen as reflecting the similarities in meaning among the three nouns.

Table 4 shows that there are a few collocates which are shared by two of the nouns. IDEA and NOTION share adjectives indicating clarity or the lack of it (*clear, vague*), also a positive evaluative collocate (*brilliant*); IDEA and CONCEPT share an adjective (*Marxist*) indicating domain; CONCEPT and NOTION have in common adjectives of domain

⁸ Adjective classifications proposed in the literature (e.g. Dixon 2005; Rachidi 1989; Lee 1994; Hundsnurher and Splett 1982) were considered, but found not to facilitate the analysis in that they did not contain categories which were essential for an insightful interpretation of the results. I therefore decided to follow the principle commonly adopted in corpus linguistics, that categorisations should be allowed to emerge from the data themselves. Like most semantic categorisations, that of adjectives is not totally neat and tidy, and it is inevitable that some examples can be classified in more than one way.

(*scientific*) and importance (*fundamental*), and also one (*traditional*) which we might treat under recency or the lack of it.

It is also instructive to look at those adjectival collocates which are specific to individual nouns. What stands out most clearly in the list for IDEA is the predominance of evaluative terms, both positive (*good, better, great, excellent, nice, splendid, interesting, fresh, clever, accepted*) and negative (*bad, wrong*). There are also specific adjectives of generality (*rough, fair*), importance (*dominant*), domain (*political*), provenance (*western*) and degree (*faintest*), this last being part of idiomatic constructions which I shall deal with later. Finally, there are two adjectives specific to the list for IDEA which are more difficult to assign to clear semantic classes (*radical, certain*). As far as collocates found only in the list for CONCEPT are concerned, it is clear that domain adjectives are important (*legal, historical, linguistic, psychological*), together with those of importance (*central, key, important, relevant*) and generality or the lack of it (*broader, particular, limited*). Two collocates (*difficult, elusive*) indicate difficulty, and one (*British*) provenance. For NOTION, there are collocates of recency or the lack of it (*current, older*), generality (*wider*), difficulty or the lack of it (*complex, simple*) and objectivity (*objective*). However, the most striking feature is a clear preference for adjectives of negative evaluation (*simplistic, confused, preconceived, strange, false* and possibly *conservative*). If we examine the whole of the BNC, rather than just the four registers focused on in this article, we find additional negatively evaluative strong collocates of NOTION: *absurd, mistaken, fancy*. Furthermore, there are other adjectives in the list (e.g. *romantic, popular, intuitive*) which, although they are not in themselves inherently negative, are easily interpreted as negative in some, though not all, of the contexts in which they are used together with NOTION. All four examples with *romantic* (with a small initial 'r', rather than with a capital, referring to a period of literature or art) are of this type:

has. I think there's been surprisingly little er perhaps it's a slightly **romantic notion**, but I think it's got something to do with, brain is the first organ to truly "die". It's a **romantic notion** I know --; but to me, with his system idling yet the brain forgo your claim, do you think Rohan would allow it? What a **romantic notion** you must have of him." "We've fallen in love with proximity to an active audience. However, there is the problem of the **romantic notion** of pure art devoid of social responsibility. Some people engaging in the issue

For *popular notion* most of the concordance lines are too limited to show the effect clearly, so expanded contexts are provided below:

This chapter examines the currently **popular notion** of "transition to adulthood" in order to illustrate the complex and contentious nature of integration into adult life. The **popular notion** of a complete fusion of Christianity and Aristotle, obstructive to all criticism, can not stand. However, despite the publicity given to the more extravagant claims about the impact of new technology on the level of unemployment, and the **popular notion** that the silicon chip is a job destroyer, a survey, published in 1979, of some 400 documents on the effect of the new information technologies on employment showed "how little foundation there

is to existing studies, half of which are by pessimists (often with a trade union background)

The **popular notion** that words do not hurt you is by no means endorsed among women survivors of male violence

Reform as the central aim of the penal system was a highly **popular notion** in the 1950s and 1960s: penological thought was dominated by "the rehabilitative ideal".

What we appear to have here is an instance of the phenomenon, discussed earlier, whereby an item, in this case NOTION, which frequently collocates with lexical items or longer stretches of language indicating a particular type of evaluation (here, negative), can sometimes induce an evaluative interpretation even of associated items which are not themselves inherently attitudinal. As we have seen, Hunston (2007: 266) warns us that this is only a tendency, since there are occasions on which this attitudinal meaning transfer does not occur.

We may summarise the findings on adjectival collocation as follows:

- Certain types of adjective (emphatic/focusing, generality, difference, recency, domain, abstraction) are common to all three nouns, reflecting the close similarity among them.
- IDEA collocates strongly with evaluative adjectives, positive and negative.
- CONCEPT is associated with adjectives indicating domain, importance and difficulty.
- NOTION associates strongly with negative evaluative adjectives, and may even exhibit a transfer phenomenon such that adjectives without an inherently negative component, such as *romantic* or *popular*, can at times be interpreted negatively in this context. These facts fit in well with the suggestion, in the dictionary definitions cited at the beginning of this paper, that notions are somewhat insubstantial, inconsequential mental constructs.

Broadly, the information derived from adjectival collocation supports and adds detail to that which we obtained from frequency considerations. Ideas are mental constructs which are part of our everyday lives, and are often evaluated either positively or negatively with respect to an intended application (e.g. *a good/brilliant idea* is one which the person describing it thinks is likely to be successful in relation to some particular plan or situation). Concepts, on the other hand, are mental constructs which are predominantly associated with fields of knowledge such as science, religion or philosophy, and are frequently commented on in terms of their importance or difficulty. Finally, notions are mental constructs which tend to be viewed negatively. Despite these differences, there are contexts in which the three nouns appear to be in free variation, though it is possible that close examination of an even larger database would reveal subtle differences even here.

We now pass to a rather more cursory examination of the Spanish data on adjectival collocation. Here, lists were obtained for adjectives occurring at either the position one to the left of the noun (L1) or one to the right (R1), for the singular and plural forms of lemmas. This reflects the difference in word order between Spanish and English: in

Spanish, although the default situation is for an adjective to follow a noun, there are circumstances in which the order may be reversed. An adjective was included in the list for a lemma if it appeared in the 20 most frequent collocates for the singular form, the plural form, or both. There is a problem here in that, since NOCIÓN is not common in the corpus, individual adjectival collocates are of too low a frequency for valid conclusions to be drawn. We shall therefore look only at IDEA and CONCEPTO. The distribution of collocates across the two lemmas is shown in Table 6.

Collocates which are common to both nouns fall into the following categories, which overlap strongly with those for the three English nouns:

- generality: *general* ('general'), *básico* ('basic')
- importance: *central* ('central'), *fundamental* ('fundamental')
- difference: *diferente*, *distinto* ('different')
- domain: *filosófico* ('philosophical'), *religioso* ('religious'), *estético* ('aesthetic')
- recency: *nuevo* ('new')
- evaluation: *bueno* ('good')
- other: *propio* ('own' or 'very'); *cierto* (with plural = 'certain')

IDEA, CONCEPTO	IDEA only	CONCEPTO only
nuevo	fijo	antiguo
general	claro	tradicional
básico	solo	erróneo
bueno	original	clásico
central	inicial	actual
fundamental	malo	alto
diferente	descabellado	lechuguino
propio	grande	esencial
filosófico	puta	científico
religioso	mínimo	principal
cierto	remoto	espiritual
distinto	aproximado	matemático
estético	genial	hindú
	político	genético
	liberal	diverso
	previo	elemental
	revolucionario	cultural
	ilustrado	complejo
	musical	moderno
	socialista	
	avanzado	
	romántico	
	menor	

Table 6: Distribution of adjectives at L1 + R1 in Spanish (IDEA, CONCEPTO only)

If we now look at the adjectival collocates which are specific to IDEA, we see that there are a few which are evaluative, either positively: *grande* ('big'), *genial* ('brilliant'),

avanzado ('advanced'), and perhaps *ilustrado* ('erudite') or negatively: *malo* ('bad'), *descabellado* ('crazy'), but these are by no means as numerous as for English IDEA. There are also specific adjectives of generality: *aproximado* ('approximate, rough'), recency: *original* ('original'), *inicial* ('initial'), *previo* ('previous'), domain: *político* ('political'), *musical* ('musical'), *socialista* ('socialist'), clarity: *claro* ('clear') and degree: *menor* ('least'), *mínimo* ('smallest'), *remoto* ('remote'),⁹ *puta* (lit. 'prostitute'), these last being part of idiomatic expressions (see later discussion). But there are also other collocates which do not fall into the main groupings: *fijo* ('fixed'), *solo* ('single'), *liberal* ('liberal'), *revolucionario* ('revolutionary'), *romántico* ('romantic'). Collocates unique to the list for CONCEPTO include ones of domain: *matemático* ('mathematical'), *científico* ('scientific'), *cultural* ('cultural'), *genético* ('genetic'), *espiritual* ('spiritual'), recency or the lack of it: *antiguo* ('old'), *tradicional* ('traditional'), *actual* ('current'), *moderno* ('modern'), and possibly *clásico* ('classical'), importance: *esencial*, *elemental* ('fundamental'), *principal* ('main'), difference: *diverso* ('different'), provenance: *hindú* ('hindu'), complexity/difficulty: *complejo* ('complex'). What marks CONCEPTO out as different from English CONCEPT, however, is the occurrence of evaluative adjectives, both positive and negative: *erróneo* ('erroneous'), *alto* ('elevated'), *lechuguino* ('dandified').¹⁰

Summarising the salient differences between the two languages with respect to adjectival collocation with the three nouns in each case, we may say that the Spanish nouns are more similar in their usage, overall, than the English ones. In particular, all three Spanish nouns associate with evaluative items, but this tendency is much weaker than for English IDEA and NOTION, and we have seen that English CONCEPT shows no predilection for such items. Interestingly, a similar general conclusion was reached in an earlier study of three adverbs, *basically*, *essentially* and *fundamentally*, and their formal equivalents in Spanish: once more, Spanish did not differentiate so clearly among the three items as did English (Butler in press). It is possible to speculate, on the basis of these two studies, that there might be a tendency towards a higher degree of precision or specificity in English lexis than in Spanish; however, this is an area which clearly requires considerably more research.

4. Idiomatic usages

Finally, we turn to some idiomatic usages in the two languages.¹¹ We find English sequences consisting of forms of HAVE plus *no* and one of the three nouns under discussion (e.g. *I have no idea*). Combinations with *idea*, which express the meaning of not knowing something, are particularly frequent in fiction, followed by news, spoken

⁹ In the contexts in which it appeared with *idea*, *remoto* did not indicate lack of recency.

¹⁰ In view of the low frequency of collocations with NOCIÓN, these were extracted from the Leeds internet corpus of Spanish (<http://corpus.leeds.ac.uk/internet.html>). Preliminary investigation suggests that although there is some collocation of NOCIÓN with negatively evaluative adjectives, this tendency is much less pronounced than in the case of English NOTION.

¹¹ Clearly, the idiomatic constructions discussed here are only a small subset of the ones into which the nouns under study enter. I have focused here on certain constructions which have restricted variability, and which are common to at least two of the nouns.

and academic language, as shown in Table 7. The other two nouns occur with smaller frequencies in this construction, and are used in a somewhat more literal way, to indicate that the subject does not possess a particular mental construct.

Collocation	Total	Spoken	Fiction	News	Academic
HAVE <i>no idea</i>	873 [16.90]	64 [6.43]	638 [107.95]	143 [13.66]	28 [1.83]
HAVE <i>no concept</i>	10 [0.19]	3 [0.30]	2 [0.34]	3 [0.29]	2 [0.13]
HAVE <i>no notion</i>	15 [0.10]	0 [0.00]	13 [2.20]	0 [0.00]	2 [0.13]

Figures in square brackets are frequencies per million words.

Table 7: Frequencies of the lemma HAVE + *no* + noun in the 4 sections of BNC World

Sequence	Total	Spoken	Fiction	News	Academic
<i>the faintest idea</i>	68 [1.32]	13 [1.31]	49 [3.08]	3 [0.29]	3 [0.20]
<i>the faintest notion</i>	2 [0.04]	0 [0.00]	2 [0.13]	0 [0.00]	0 [0.00]
<i>the slightest idea</i>	16 [0.31]	1 [0.10]	14 [0.88]	0 [0.00]	1 [0.07]
<i>the slightest notion</i>	1 [0.02]	0 [0.00]	1 [0.06]	0 [0.00]	0 [0.00]
<i>the foggiest idea</i>	8 [0.15]	4 [0.40]	4 [0.25]	0 [0.00]	0 [0.00]
<i>the foggiest notion</i>	1 [0.02]	1 [0.10]	0 [0.00]	0 [0.00]	0 [0.00]
<i>the remotest idea</i>	7 [0.14]	2 [0.20]	5 [0.31]	0 [0.00]	0 [0.00]
<i>the least idea</i>	12 [0.23]	0 [0.00]	11 [0.69]	0 [0.00]	1 [0.07]
<i>the vaguest idea</i>	7 [0.14]	0 [0.00]	4 [0.25]	3 [0.29]	0 [0.00]
<i>the vaguest notion</i>	3 [0.06]	0 [0.00]	3 [0.19]	0 [0.00]	0 [0.00]

Figures in square brackets are per million words. No other combinations were found.

Table 8: Sequences of the form *the* + superlative adjective + *idea/notion* occurring in the 4 sections of BNC World

Closely related are expressions of the form HAVE + *n't* + *the* + superlative of adjective + noun, as in *I haven't the faintest idea*. In terms of meaning, these are reinforced versions of the expressions we have just considered, such as *I have no idea*. As shown in Table 8, *idea* is the most frequent noun in this construction, and occurs with *faintest*, *slightest*, *foggiest*, *remotest*, *vaguest* and *least* as the superlative adjective, *faintest* being the most common. *Notion* occurs with *faintest*, *slightest*, *foggiest*, *vaguest*,

but *concept* is not found at all in this construction. These constructions are most frequent in fiction and (less so) spoken language.¹² They were very infrequent in news and non-existent in academic language. A detailed account of such sequences, situated within the framework of Systemic Functional Linguistics, can be found in Tucker (1996).

Related expressions are found in the Spanish corpus (see Table 9). There are several expressions with *idea*: *no* + TENER ('have') + *ni idea*, *la (más) mínima idea*, *ni puta idea*, *la más remota idea*, etc. Of the most frequent, *ni idea* occurs most in spoken language, *la menor idea* in spoken language and fiction. There are also rare occurrences of *la menor/mínima noción*, but no occurrences of this construction were found with *concepto*.

Sequence	Total	Spoken	Fiction	News	Academic
<i>ni idea</i>	75 [3.29]	57 [10.03]	11 [1.92]	7 [1.22]	0 [0.00]
<i>la menor idea</i>	50 [2.19]	22 [3.87]	23 [4.02]	5 [0.87]	0 [0.00]
<i>ni puta idea</i>	8 [0.35]	7 [1.23]	1 [0.18]	0 [0.00]	0 [0.00]
<i>la más mínima idea</i>	6 [0.26]	3 [0.53]	2 [0.35]	1 [0.18]	0 [0.00]
<i>la mínima idea</i>	1 [0.04]	1 [0.18]	0 [0.00]	0 [0.00]	0 [0.00]
<i>la más remota idea</i>	4 [0.18]	0 [0.00]	3 [0.52]	1 [0.18]	0 [0.00]
<i>ni remota idea</i>	1 [0.04]	1 [0.18]	0 [0.00]	0 [0.00]	0 [0.00]
<i>idea remota</i>	2 [0.09]	2 [0.35]	0 [0.00]	0 [0.00]	0 [0.00]
<i>la menor noción</i>	2 [0.09]	1 [0.18]	0 [0.00]	1 [0.18]	0 [0.00]
<i>la mínima noción</i>	1 [0.04]	0 [0.00]	1 [0.18]	0 [0.00]	0 [0.00]

Figures in square brackets are per million words. No other combinations were found.

Table 9: Idiomatic expressions with *idea/noción* in the *Corpus del Español*

¹² The high frequency in fiction of a sequence whose idiomaticity might lead us to expect an association with spoken language recalls the comment of Biber *et al* (1999: 509) that emotive use of *poor*, as in *poor little bastards*, is frequent in fiction. It is as if writers of fiction unknowingly over-emphasise what they take to be typical conversational features in an effort to make their dialogues seem natural.

5. Conclusion

In conclusion, we may say that close corpus-based analysis has shown that IDEA, CONCEPT and NOTION share a core of meaning which is reflected in their apparent interchangeability in some contexts, and in their frequent collocation with certain semantic groupings of adjectives. However, the study also reveals a number of differences in behaviour which suggest that ideas, as conceptualised in English, are mental constructs which belong essentially to the area of how we conduct our everyday lives and are frequently evaluated, positively or negatively, in terms of their suitability for some particular application. The term *notion* tends to be used preferentially for mental constructs which are negatively evaluated, while concepts are characterised by their abstraction and belong primarily to the realm of mental activities involved in professional and technical areas. The Spanish data reveal a picture which is similar in many ways, but there are indications that the three Spanish nouns are more similar than the English ones in their range of usage. At a more general level, this study has shown that detailed studies of corpus materials focusing on collocational behaviour offer powerful means of discriminating the fine detail of lexical usage, and can provide information which is of interest not only within descriptive and contrastive linguistics itself, but also in applied areas such as translation and advanced language learning.

Works cited

- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan, eds. 1999: *Longman Grammar of Spoken and Written English*. London: Longman.
- Butler, Christopher S. In press: “‘Basically Speaking’: a Corpus-Based Analysis of three English Adverbs and their Formal Equivalents in Spanish”. María L. A. Gómez-González, J. Lachlan Mackenzie and Elsa González-Álvarez, eds. *Current Trends in Contrastive Linguistics: Functional and Cognitive Perspectives*. Amsterdam and Philadelphia: Benjamins.
- Dixon, Robert M. W. 2005: *A Semantic Approach to English Grammar*. 2nd ed. Oxford: Oxford UP.
- Hundsnurher, Franz and Jochen Splett 1982: *Semantik der Adjektive im Deutschen: Analyse der semantischen Relationen*. Opladen: Westdeutscher Verlag.
- Hunston, Susan 2007: ‘Semantic Prosody Revisited’. *International Journal of Corpus Linguistics* 12.2: 249-268.
- Lee, Sun-Muk 1994: *Untersuchungen zur Valenz des Adjektivs in der deutschen Gegenwartssprache*. Frankfurt am Main: Lang.
- Louw, Bill 1993: ‘Irony in the Text or Insincerity in the Writer? — the Diagnostic Potential of Semantic Prosodies’. Mona Baker, Gill Francis and Elena Tognini-Bonelli, eds. *Text and Technology: In Honour of John Sinclair*. Amsterdam: Benjamins. 157-76.
- Louw, Bill 2000: ‘Contextual Prosodic Theory: Bringing Semantic Prosodies to Life’. Chris Heffer and Helen Saunston, eds. *Words in Context: A Tribute to John Sinclair on his Retirement*. English Language Research Discourse Analysis Monograph 18. University of Birmingham. [CD-ROM]
- Partington, Alan 1998: *Patterns and Meanings*. Amsterdam: Benjamins.
- Partington, Alan 2004: “‘Utterly Content in Each Other’s Company’: Semantic Prosody and Semantic Preference”. *International Journal of Corpus Linguistics* 9.1: 131-56.
- Rachidi, Renate 1989: *Gegensatzrelationen im Bereich deutscher Adjektive*. Tübingen: Niemeyer.
- Sinclair, John M. 1987: *Looking Up*. London: Collins.

-
- Sinclair, John M. 2004 (1996): 'The Search for Units of Meaning'. *Textus* IX: 75-106. Reprinted in John Sinclair (2004) *Trust the Text: Language, Corpus and Discourse*. London: Routledge. 24-48.
- Stubbs, Michael 1995: 'Collocations and Semantic Profiles: on the Cause of the Trouble with Quantitative Studies'. *Functions of Language* 2.1: 23-55.
- Stubbs, Michael 2001: *Words and Phrases: Corpus Studies of Lexical Semantics*. New York: Blackwell.
- Tucker, Gordon H. 1996: 'So Grammarians haven't the Faintest Idea: Reconciling Lexis-Oriented and Grammar-Oriented Approaches to Language'. Ruqaiya Hasan, Carmel Cloran and David G. Butt, eds. *Functional Descriptions: Theory in Practice*. Amsterdam and Philadelphia: Benjamins. 145-78.
- Whitsitt, Sam 2005: 'A Critique of the Concept of Semantic Prosody'. *International Journal of Corpus Linguistics* 10.3: 283-305.

Received 17 July 2008

Revised version accepted 9 September 2008

Christopher S. Butler is Honorary Professor at Swansea University, UK. His research interests are in the comparison of functional theories of language and their relationship with formal and with cognitive and/or construction-based theories, and in the use of quantitative corpus-based techniques in the study of English and Spanish. Recent publications include *Structure and Function: A Guide to Three Major Structural-Functional Theories* (Benjamins, 2003) and 'Cognitive adequacy in structural-functional theories of language'. *Language Sciences* 30.1: 1-30, 2008.

Address: Applied Linguistics, School of Arts, Swansea University, Singleton Park, Swansea SA2 8PP. Tel. +44 (0)1792 602540, Fax +44(0)1792 602545. E-mail cbutler@ntlworld.com (the latter used for all correspondence.)