Testing the *sub*-test

An analysis of English *-ic* and *-ical* adjectives*

Stefan Th. Gries
University of Southern Denmark

This paper pursues two objectives, one linguistic and one methodological in nature. First, it is concerned with a corpus-based analysis of the degree to which pairs of *-ic*/*-ical* adjectives (e.g. *classic*/*classical*) are synonymous. Second, it investigates whether Church et al.'s (1994) *sub*-test can be fruitfully applied to this phenomenon. As to the first issue, I conclude that individual *-ic*/*-ical* adjectives can be located on a continuum of semantic similarity, with some being virtually completely synonymous and others being strongly differentiated; several semantic and distributional distinctions between members of adjective pairs are pointed out on the basis of distinctive collocates. As to the second question, I demonstrate on the basis of a simulation that the *sub*-test is conceptually adequate, but suffers from its asymptotic approach, which is why Fisher-exact is argued to be a more adequate diagnostic.

Keywords: derivational morphology, *-ic*/*-ical* adjectives, collocations, *t*-test, Fisher-exact test

1. Introduction

One out of many demanding tasks of lexicographers is to provide information on the degree to which different lemmas are semantically similar or dissimilar, information that is crucial to the formulation of definitions in dictionary entries and, perhaps even more so, to the development of thesauri. Although intuition and simple example collection have been the main methods in these areas, the recent past has witnessed a development from reliance on intuitions to more sophisticated corpus-linguistic techniques based on concordances and collocations. A considerable part of research in this area has been devoted to

International Journal of Corpus Linguistics 8:1 (2003), 31–61.
ISSN 1384-6655/ E-ISSN 1569-9811

identifying statistical measures of collocational strength. Many of these measures serve to identify words that co-occur significantly more often than would be expected by chance; others are used to identify words that serve to separate different words from each others (cf. Berry-Rogghe 1974; Church & Hanks 1990; Church et al. 1991; Dunning 1993; Weeber, Vos & Baayen 2000; to name but a few, and Manning & Schütze 2000: Chapter 5 for an overview).

While these and many publications concerned with (the identification of) significant collocations are widely quoted by corpus and computational linguists, another related study seems to have received much less attention, namely the *sub*-test of Church et al. (1994). This work by Church and his colleagues aims at analysing the distribution of semantically similar words. Assuming that semantic similarity is a gradient property, they operationalise the semantic similarity of two words W_1 and W_2 in terms of the distributional criterion of substitutability of W_1 and W_2 in a line of reasoning similar to that discussed in Miller and Charles (1991). More precisely, they replace "questions about synonymy (a semantic notion) by questions about textual substitutability (a distributional notion)", where substitutability is defined as "the phenomenon of two items appearing in discourse to occupy the same lexicosyntactic space" (Church et al. 1994: 169).[1] Then, the amount of overlap of W_1 and W_2 found is tested for significance with a variant of the *t*-test. Let me briefly illustrate the approach they take with reference to their example (*ask for* vs. *request*).

In order to determine the degree to which the verb *ask for* can be replaced by *request* in verb-object pairs (V-O pairs) in the 1988 AP corpus (comprising about 44m words), Church et al. first determined the number of significant verb-object collocations, which amounted to 75,115. Then, they identified all significant collocates in the object position of each of the two verbs, yielding 85 and 59 significant objects for *ask for* and *request* respectively as well as the number of significant collocates that occurred after both verbs, resulting in 28 cases. Thirdly, they determined the number of V-O pairs where W_2 (the object) is a significant collocate of *request*, amounting to 2,382 such pairs. For ease of understanding, these data are summarised in Table 1 in the form familiar from measures of collocational strength.

From Table 1, it follows that the expected amount of overlap is 2.69 ($\approx$ 3), and the question to be answered is whether 28 is significantly larger than 2.69 (or 3, for that matter). Church et al. (1994: 170) suggest to use the formula in (1),[2] according to which the observed amount of overlap is indeed significantly higher than the expected one.

Table 1. Applying the *sub*-test to *request* → *ask for*

	$N_{sign.\ V\text{-}O\ pairs}$ (where O = sign. coll. of *request*)	$N_{sign.\ V\text{-}O\ pairs}$ (where O ≠ sign. coll. of *request*)	Σ rows
$N_{sign.\ coll.\ of\ ask\ for}$	28	57	85
$N_{sign.\ coll.\ of\ other\ V\text{-}O\ pairs}$	2,354	72,676	75,030
Σ columns	2,382	72,733	75,115

$$t = \frac{28 - 2.7}{\sqrt{(28 \cdot (1 - 28 \div 85) + (2.7 \cdot (1 - 2,382 \div 75,115))}} \approx 5.47 \qquad (1)$$

Thus, *ask for* and *request* are distributed similarly (such that *request* can be substituted for *ask for* more often than expected) so the two words are probably semantically similar – Church et al. (1994: 169) are careful, however, to point out that "distributional evidence alone cannot be used conclusively to prove semantic regularities" and that the *sub*-test is also sensitive to co-hyponyms and antonyms.[3]

Although the general approach and the results of the *sub*-test as discussed by Church et al. are promising and instructive, on a more detailed level of scrutiny they are fraught with some minor shortcomings. First, it is unclear from the paper how the significant collocations entering into Table 1 are identified in the first place. This might seem to be splitting hairs, but it is nevertheless an important point to mention since it has been shown time and again that (i) different measures of collocational strength yield different results and (ii) MI (the only measures of collocational strength mentioned in Church et al. 1994) is often problematic since MI tends to overestimate the significance of rare events (cf. Manning & Schütze 2000: Section 5.4).[4]

This leads us to the second point, namely the use of the *t*-test as a means of determining whether the observed overlap deviates significantly from the one expected according to the null hypothesis. As is well-known, this test is an asymptotic test which can be problematic for the non-normally distributed and low-frequency data corpus/collocation data often instantiate.[5]

The most important point of critique, however, is that the authors do not provide any evaluation measure for the obtained results. By that I mean that one needs to demonstrate that significant overlap of significant collocates of two words W_1 and W_2 is in fact restricted to distributionally/semantically similar words rather than to all pairs of words (of one word class). For example,

Table 2. Applying the *sub*-test to *valid* → *unusual*

	$N_{sign.\ Adj\text{-}W2\ pairs}$ (where W_2 = sign. coll. of *valid*)	$N_{sign.\ Adj\text{-}W2\ pairs}$ (where $W_2 \neq$ sign. coll. of *valid*)	Σ rows
$N_{sign.\ coll.\ of\ unusal}$	15	129	144
$N_{sign.\ coll.\ of\ other\ Adj\text{-}W2\ pairs}$	11,761	223,282	235,043
Σ columns	11,776	223,411	235,187

the fact that *blue* and *expensive* both have *car* as a significant collocate would increase the proposed similarity index, but does of course not entail any interesting semantic commonality. Consider Table 2 for a more detailed example, namely the *sub*-test testing whether *valid* can replace *unusual* in the written part of the British National Corpus (cf. Section 3 below for technicalities).

Church et al.'s *t*-test shows that the number of 15 overlapping collocates is significantly larger than the 7 collocates that would be expected by chance ($t = 1.73$; $p \approx 0.041$) although the degree of semantic similarity is intuitively rather limited. Before accepting the *sub*-test, thus, one needs to perform several *sub*-tests and then either compute some evaluation measure such as, e.g., precision (i.e. the proportion of true cases of semantic similarity of those word pairs exhibiting significant overlap in one or both directions) or find some other way of proving the appropriateness of this test. Moreover, it would be interesting to see in what way, if any, the *sub*-test sheds light on the nature of near synonymy of words difficult to keep apart.

In the remainder of this study, I will discuss the following issues. First, to what extent can the above-mentioned shortcomings of the *sub*-test as suggested in Church et al. (1994) be remedied, and what are the consequences for the results one obtains? Second, does the precision of the *sub*-test merit its application to the lexicographic question of semantic similarity or overlap of words? In order to answer the first question, I will investigate the degree of distributional/semantic similarity of English adjectives ending in *-ic* and *-ical* (such as *economic/economical*). In order to answer the second question, I will compare the results of the *-ic/-ical* part of the study to a brief 'simulation study' on the degree of overlap of randomly chosen adjectives.

Accordingly, the paper is structured as follows. Section 2 provides a brief introduction to the phenomenon of *-ic/-ical* adjectives, pointing out some difficulties associated with these adjectives and explaining the motivation to analyse them with the *sub*-test. Section 3 will then explain in some detail how the *sub*-test was performed on such adjective pairs and what results were ob-

tained. Section 4 is devoted to determining the appropriateness of the *sub*-test: in particular, I test to what degree its result depends on whether Church et al.'s asymptotic version or an exact test is used (Section 4.1). Also, I will compare the *-ic*/*-ical* results to the results of a brief simulation of the above-mentioned kind (Section 4.2). Finally, Section 5 will summarise and conclude.

2. *-ic*/*-ical* adjectives: a short review

2.1 General issues

The class of English *-ic* and *-ical* adjectives is a very interesting phenomenon of English word-formation and puzzles learners of English as a foreign language and linguists alike. One problem (I will not address here) is concerned with the distribution of the suffixes, as there seem to be no apparent regularities governing which adjective root takes only *-ic* (e.g. *democratic, domestic, scientific, electronic, realistic, diagnostic* ...), which takes only *-ical* (e.g. *zoological, technical, physical, cosmological, umbilical* ...) and which takes both (e.g. *politic(al), economic(al), historic(al), electric(al)classic(al)* ...[6]).

In addition, there is no (fairly) uniformly accepted account of which suffix is associated with which meaning. On the contrary, different researchers have postulated different elements of meaning for the two suffixes, which makes it more difficult to agree (i) on the dimensions along which the adjectives differ in meaning and (ii) the degree to which a proposed dimension decreases the adjectives' similarity to each other. Consider Table 3 for an overview of suggested dimensions.

Table 3. Proposed meanings of the two suffixes

Author/source	*-ic*	*-ical*
Jespersen (1942:391)	quality and category	quality
Marchand (1969:242), Hawkes (1976:95)	semantically more direct connection to root substantive	semantically less direct connection to root substantive (wider senses)
Ross (1998:42)	specific	less specific/more general
Marsden (1985:30)	genuine	resembling/imitation
Marchand (1969:242), Fournier (1993:238)	scientific terms	wider common use
Kaunisto (1999), Gries (2001)	prefixed forms	

Then, there is the somewhat more difficult problem most central to our interest, namely the degree of meaning differentiation of adjectives belonging to the last of the three groups above. Opinions as to the (degree of) synonymy of the two adjectives making up an *-ic/-ical* pair differ wildly. However, the differences in degrees of synonymy are not completely arbitrary. Rather, there seems to be a continuum of semantic similarity (with synonymy as one extreme case; cf. Note 1) and different adjective pairs are located on different points of this continuum. For instance, there are adjective pairs where all researchers agree that the two forms are not synonymous (e.g. *politic* vs. *political*). Then, there are cases where researchers differ in their assessment of the degree of semantic similarity (e.g. *magic* vs. *magical*). Lastly, there are cases where researchers seem to agree that both forms by and large have the same meaning (e.g. *problematic* vs. *problematical*). (Cf. Kaunisto 1999: Section 2 and Gries 2001: Section 2 for comprehensive reviews). But a closer look at previous analyses shows that many of them suffer from a host of additional problems:

- Few researchers' findings seem to be based on empirical evidence.
- Some dimensions are formulated too vaguely to qualify for direct empirical analysis.
- If empirical evidence has been adduced, then it is sometimes register-specific only (Kaunisto 1999) or methodologically flawed (cf. Gries 2001: 77 on Marsden 1985).
- The implicit understanding of similarity or synonymy underlying some analyses is flawed because similarity is apparently considered symmetric, an approach that is questionable given many psychological findings concerning similarity (cf. e.g. Tversky 1977).
- The findings reported in some studies seem to be strongly influenced by prescriptive attitudes (cf. Fowler 1968 and contemporary dictionaries).

Therefore, the degree to which *-ic/-ical* adjectives in general or some particular *-ic/-ical* adjectives are semantically similar or even synonymous has remained largely unresolved.

2.2 The analysis of Gries (2001)

One of the most recent analyses of *-ic/-ical* adjectives is Gries (2001), applying the *sub*-test to the present question,[7] namely determining the semantic similarity of the two adjectives constituting an *-ic/-ical* pair. For 15 of the most frequent *-ic/-ical* adjectives,[8] I determined the degree of overlap of significant R1

collocates in the written part of the British National Corpus (i.e. in about 90m words), where the significance of a collocation was determined on the basis of a mixture of the Chi-square test and Dunning's (1993) log-likelihood test. Then, I summarised the relative frequencies of overlap in a two-dimensional diagram ($ESCO_2$), thereby at the same time accounting for the fact that similarity/substitutability need not be symmetric. Finally, the significant collocates and the overlap of selected adjective pairs were analysed and interpreted further on the basis of this diagrammatic representation and separating collocates (as determined by Church et al.'s 1991 *t*-test), yielding some interesting observations concerning semantic distinctions of the two adjectives and methodological consequences.

Gries (2001) is an obvious forerunner of the treatment of the issues addressed in the present work, but it is still deficient in several respects. The first drawback is of course that the interpretation of the overlap findings is based on findings which have not been tested for significance, which of course introduces the danger that the proposed semantic conclusions are based on statistically insignificant distributional data. Secondly, although the corpus size of Gries (2001) is by far the largest of all *-ic/-ical* analyses, it is not quite clear whether the chosen sample of such adjectives is representative and large enough to license the range of semantic and methodological conclusions put forward.[9] The fairly strong correspondence between my empirical findings and previous analyses of *-ic/-ical* adjectives is prima facie evidence for the proposed analysis, but it would definitely be desirable to revise and improve on this work.

Finally, just like Church et al. (1994), I did not address the issue of the diagnostic value of significant collocate overlap. This is perhaps less of a problem since I did not claim to have obtained significant results, but it is still undesirable to publish results based on a methodology whose merits and limits have remained unanalysed. Let us therefore now turn to the methodology for the present study.

3. Methods and results

The first step consisted of identifying the most relevant *-ic/-ical* adjective pairs for the analysis. On the one hand, it is of course sensible to investigate the most frequent pairs, but on the other hand it is also necessary to analyse pairs where both forms are reasonably frequent. To that end, I first generated a concordance of all words tagged as adjectives in the written part of the British

National Corpus (BNCw, amounting to 90m words) that ended in *-ic* or *-ical*. The resulting list of about 1,500 words was then scanned for pairs of *-ic*/*-ical* adjectives, yielding a list of 57 pairs with frequencies for each member of a pair (listed in Appendix A). From this list of *-ic*/*-ical* adjective pairs, I identified those pairs in which each adjective occurred more than 20 times, and these 47 pairs (bold-typed in Appendix A) were used in the analysis to be described shortly.[10]

For the *sub*-test, however, more is necessary than just the collocates of *-ic*/*-ical* adjective pairs – one also needs the number of significant collocates of all adjectives (corresponding to the number of all significant V-O pairs in Table 1 above). Therefore, as the next step, all words in the BNCw that were tagged as adjectives (AJ0) were extracted. This concordance was then edited in order to filter out cases where the first word to the right (the R1 collocate) of the italicised adjective is unlikely to be directly related to the adjective of interest:

1. *classical* (*that is* ...), i.e. cases where the adjective precedes left or right brackets;
2. *symmetric* </p>, i.e. cases where the adjective is paragraph-final;
3. *lucky*<c PUN>&hellip, i.e. cases where the word of interest has been elided;
4. *available*<c PUN>, i.e. cases where the adjective is used sentence-finally, etc.[11]

The application of these and other operations resulted in a list of 1,758,462 concordance lines containing at least one adjective. After eliminating all cases where the adjective was not followed by a second word, I generated a frequency list of all bigrams (each consisting of an adjective and its R1 collocate) in order to obtain the collocation frequencies that are necessary for the calculation of measures of collocational strength. However, I decided to include only those collocations in the analysis which occurred at least three times since collocations that occur only once or twice are, on average, unlikely to (i) be significant in the first place (given the large corpus size) and (ii) contribute reliable information to the question of substitutability. The resulting list of bigrams contained 297,111 collocations (consisting of an adjective followed by another word and occurring at least three times in the BNCw) and their frequencies.

In the next step, I determined the overall frequencies of the words in these collocations in the BNCw and computed two measures of collocational strength: MI (cf. Church & Hanks 1990; Church et al. 1991) and $2 \log \lambda$ (cf. Dunning 1993). MI was only used to identify the bigrams that occurred *more* often than would be expected by chance (i.e. those for which MI > 0). Of all

Table 4. Applying the *sub*-test to *analytic* → *analytical*

	$N_{sign.\ Adj\text{-}W2\ pairs}$ (where W_2 = sign. coll. of *analytic*)	$N_{sign.\ Adj\text{-}W2\ pairs}$ (where $W_2 \neq$ sign. coll. of *analytic*)	Σ rows
$N_{sign.\ coll.\ of\ \textit{analytical}}$	9	53	62
$N_{sign.\ coll.\ of\ other\ Adj\text{-}W2\ pairs}$	2,311	232,814	235,125
Σ columns	2,320	232,867	235,187

Table 5. Applying the *sub*-test to *analytical* → *analytic*

	$N_{sign.\ Adj\text{-}W2\ pairs}$ (where W_2 = sign. coll. of *analytical*)	$N_{sign.\ Adj\text{-}W2\ pairs}$ (where $W_2 \neq$ sign. coll. of *analytical*)	Σ rows
$N_{sign.\ coll.\ of\ \textit{analytic}}$	9	8	17
$N_{sign.\ coll.\ of\ other\ Adj\text{-}W2\ pairs}$	9,123	226,047	235,170
Σ columns	9,132	226,055	235,187

those cases, I focussed on the 235,187 bigrams for which 2 log λ was larger than 6.64 (corresponding to the Chi-square threshold for a significance level of 1%; cf. Dunning 1993:66).

As the final step, from the list of all 235,187 significant bigrams all collocations starting with one of the 94 adjectives of interest were extracted in order to determine the degree of significant collocate overlap of each adjective pair. Table 4 and Table 5 illustrate this procedure for *analytic* and *analytical*, which had 17 and 62 significant R1 collocates respectively.

We find that 9 of the 62 significant collocates of *analytical* (i.e. 14.5%) are also significant collocates of *analytic* while 9 of the 17 collocates of *analytic* (i.e. 52.9%) are also significant collocates of *analytical*. Less technically, *analytic* and *analytical* share 9 significant collocates. This shows that the distributional/semantic similarity of the two adjectives is indeed asymmetric: *analytic* is much more similar to *analytical* than vice versa because *analytic* can be replaced by *analytical* more often than the other way round. This interpretation is supported by checking these findings for significance: the expected frequencies in Table 4 and Table 5 are both 1 (rounded) and the resulting *t*-scores are 2.91 ($p < 0.01$) and 3.78 ($p < 0.001$) respectively. Figure 1 summarises the results for all 47 adjective pairs in the $ESCO_2$ format proposed in Gries (2001). In this diagram,

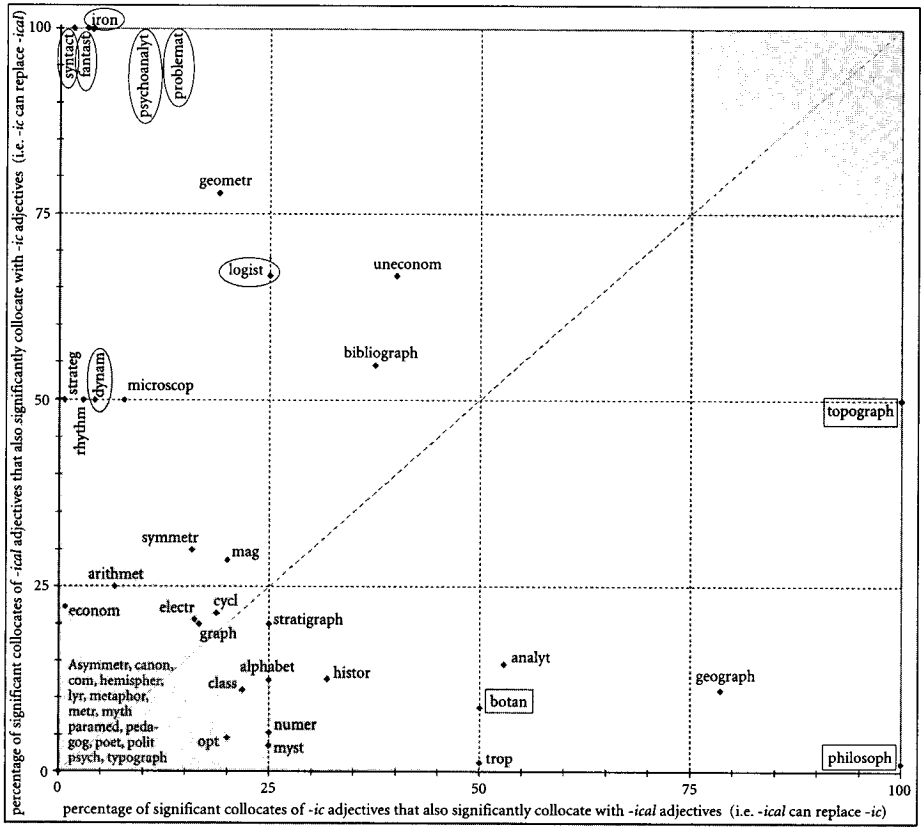

Figure 1. $ESCO_2$ and *sub*-test (*t*-test) for *-ic*/*-ical* adjectives

- the 13 adjective roots in bold type (e.g. *geometr* and *geograph*) indicate that the amount of overlap was significantly higher than the expected one in both directions;
- the 3 adjective roots in a rectangle (*botan*, *philosoph*, *topograph*) indicate that the amount of overlap was significantly higher than the expected one along the dimension represented on the x-axis;
- the 7 adjective roots in an ellipsis (e.g. *syntact* and *logist*) indicate that the amount of overlap was significantly higher than the expected one along the dimension represented on the y-axis.

In Gries (2001: 88–89, 89–92), I illustrated

- how $ESCO_2$ exemplifies and corresponds to previous analyses of *-ic/-ical* adjectives on the basis of some adjective pairs located at extreme positions in the diagram (namely *politic(al)*, *analytic(al)*, *economic(al)* and *problematic(al)*);
- how the empirical data and the analysis produce more precise characterisations of some adjective pairs (namely *logistic(al)*, *symmetric(al)*, *numeric(al)* and *magic(al)*).

The present results for these adjectives are fairly similar and equally supportive,[12] but since the earlier study failed to test the overlap for significance, it is worth comparing at least some of the earlier results and discussion with the present ones. As to *politic(al)*, there is no overlap at all, which supports previous findings about the strong semantic and distributional distinction of *politic* and *political*. With respect to *analytic(al)*, my earlier finding and the practice of dictionaries to consider the words fairly synonymous are corroborated since, as can be seen in Figure 1, the overlap is bidirectionally significant.[13] As to *economic(al)*, previous findings are again supported since the degree of meaning differentiation between *economic* ('pertaining to economy') and *economical* ('money-saving') is reflected in the fact that the overlap we find is small (and in fact smaller than expected by chance). Also, the *sub*-test shows that the unexpectedly strong tendency of *problematical* being replaceable by *problematic* is highly significant ($t = 10.5$; $p < 0.001$). Finally, for *logistic(al)*, I concluded that *logistic* can be more often substituted for *logistical* since *logistic* can be used in the contexts of mathematics and transportation whereas *logistical* can only be used in the latter sense. The present results show that this difference in directionality is indeed significant since *logistical* can be used where *logistic* can, but not necessarily vice versa.

Let us now look at some additional cases in more detail. Take, for instance, *alphabetic(al)*, which is an example which is treated differently by reference works. All reference works have an entry for *alphabetical*, but then some (e.g. CoCD and CCED) define *alphabetical* as "arranged according to the normal order of the letters in the alphabet", but neither mention *alphabetic* in the entry for *alphabetical* nor do they have an entry for *alphabetic* while others (e.g. CEDT) at least mention *alphabetic* as a synonym of *alphabetical* in the latter term's entry. CaED, then, devotes one entry to both adjectives, which states "pertaining to the alphabet, arranged alphabetically". The OED, finally, gives the following definitions, where I find it difficult to see the difference between *alphabetic* (2) and the first two possibilities of *alphabetical* (1):

alphabetic † 1. Arranged in order of the alphabet. Obs.
2. Of, pertaining to, or by means of an alphabet; or by letters representing simple sounds.

alphabetical 1. Of, pertaining to, or in order of the alphabet.
† 2. fig. Literal, strict. Obs. rare.
3. = alphabetic 2.

According to Figure 1, however, we find that both adjectives do exhibit only very moderate overlap, which does not reach standard levels of significance and rather indicates a distinctness of the adjectives' distributions/meanings. The distribution is summarised in Figure 2, where the number of significant collocates of each adjective (ordered according to collocational strength) and the degree of overlap are indicated approximately by the sizes of the (differently-shaded) rectangles. The collocates in bold-type are distinctive collocates for the preceding adjective (cf. Note 13). The difference is fairly obvious, though not recorded by all reference works: *alphabetic* is used for denoting the specific kind of writing sign (the alphabet) whereas *alphabetical* is exclusively used for denoting the process or result of sorting according to the alphabet. True, the overlapping collocate *order* suggests that this meaning also figures in *alphabetic*, but (i) even then *alphabetic* would have an additional sense that is not associated with *alphabetical* and (ii) note that, although *order* occurs significantly with both adjectives (contradicting the OED's entry according to which this use should be obsolete), it is a distinctive collocate for *alphabetical*, which suits the remaining distributional facts very well. In short, *alphabetic(al)* exemplifies a case where some reference works have not provided an adequate description of actual usage which can, however, easily provided by the present way of analysis. In this respect, note also that the difference between *alphabetic* and *alphabetical* cannot be couched conveniently in the terms of any of the dichotomies proposed in the literature.

Another pair of adjectives we can look at only briefly is *geographic(al)*. Again, from current reference works we learn that both words are synonymous:

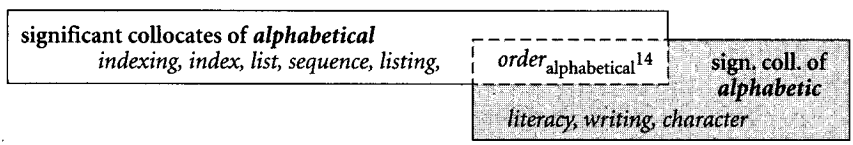

Figure 2. Significant and distinctive collocate sets for *alphabetic(al)*

"*geographic* means the same as *geographical*; a formal and rather old-fashioned word" (CoCD: s.v. *geographic*) and "*[g]eographical* or *geographic* means concerned with or relating to geography" (CCED: s.v. *geographic* and *geographical*); this is echoed by the OED. A look at the corpus data shows that the situation is not that simple (although lack of space prevents us from investigating in detail the more than 100 significant collocates). *Geographic* is less frequent than *geographical* (although not "somewhat rare" as the OED suggested) and a glance at Figure 1 shows that the overlap is in fact bidirectionally significant, supporting the reference works just quoted in two ways: the two adjectives are distributionally very similar, and *geographical* can replace *geographic* more often than vice versa, which is in turn the reason why *geographic* is defined via *geographical* and not vice versa. A closer look (especially at the distinctive collocates), however, also reveals some distinctions: for instance, *geographic* is only used attributively (all its significant R1 collocates are nouns)[15] whereas *geographical* is also used predicatively and/or in con-/disjunction with other attributively used adjectives (having values of 2 log λ of 16.98 and 12 for *and* and *or* respectively). Also, the distinctive collocates of *geographic* are common, simple words (namely *information*, *market*, *pole*, *magazine*, *races*) whereas those of *geographical* are much less common as well as morphologically more complex words (e.g. *mobility*, *locations*, *variations*, *distribution*, *inequalities*, *proximity* etc.); this stands in stark contrast to Marchand's (1969:242) and Fournier's (1993:248) rule of thumb mentioned above in Table 3. While it is difficult to lay down the observations into watertight usage rules for, e.g., learners of English as a second language, the differences are quite apparent and demonstrate that previous lexicographic treatment leaves some matters unresolved.

As the next example, let us consider *arithmetic(al)*. Some previous references and studies have concluded that both adjectives are completely synonymous (CoCD, CEDT, OED, Ross 1998:43), but according to Figure 1, the two adjectives do not exhibit significant overlap. Consider Figure 3.

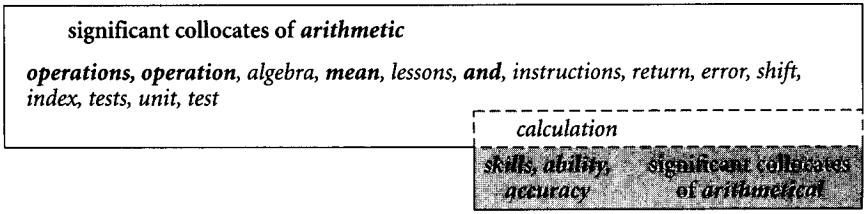

Figure 3. Significant and distinctive collocate sets for *arithmetic(al)*

Again, two findings are worth a brief mention. First, we see that it is again the *-ic* adjective alone that is used predicatively (cf. *and* as its significant and distinctive collocate, but also Note 15). More interesting, however, is that the significant (and distinctive) collocates of *arithmetic* are semantically different from those of *arithmetical* such that their connection to the adjective is less direct. The significant collocates of *arithmetical* mostly refer to properties of human beings doing X whereas the significant collocates of *arithmetic* rather tend to characterise the nature of X directly. Put differently, one needs *arithmetical skills* to carry out an *arithmetic operation*; one needs *arithmetical abilities* to compute an *arithmetic mean*. On a very general level, this pattern might be considered to instantiate the direct (*-ic*) vs. less direct (*-ical*) pattern postulated by Marchand (1969:242). While more (detailed) work is necessary to tease the regularity out of the data in some more detail, the two comments just made were mentioned by no reference work at all, and the OED's entry even uses the phrase "arithmetical operation" in its definition of *arithmetic*, although we have seen that *operations* is in fact a distinctive collocate of *arithmetic*.[16]

We now turn to *optic(al)*. According to CoCD and CCED, *optic* means 'relating to eyes or to sight' (s.v. *optic*) but *optical* is not mentioned in *optic*'s entry. *Optical*, then, is defined as follows: 'Optical instruments, devices, or processes are concerned with vision, light, or images' (s.v. *optical*). In other words, the entries do not establish any relation between the two adjectives (unlike in many other cases of *-ic*/*-ical* pairs), but they loosely imply a distributional difference in that *optical* seems to be restricted to inanimate referents (esp. technical devices); loosely because the word *vision* in the definition of *optical* can, but need not, be understood as involving animate referents. The OED, by contrast, provides similar definitions (implying more overlap, though) and relates both adjectives such that two sub-entries for *optic* refer to *optical* and one of *optical* refers to *optic*. Again, we need to turn to the exact results of the corpus analysis to shed some light on these conflicting claims and results. Figure 1 indicates that *optic* and *optical* share significantly more collocates (in both directions) than would be expected by chance; for the exact distribution of collocates, consider Figure 4.

Three results are worth mentioning in that they relate to previous findings. First, Figure 1 and Figure 4 illustrate the superior coverage of the OED in this case: the fact that the observed overlap is significant reflects (i) the strong similarity of the definitions of both adjectives and (ii) the fact that at least the OED considers the adjectives so similar that one (*optic*) is defined in terms of the other (*optical*). Second, the direction of the OED's definition (i.e. that *op-*

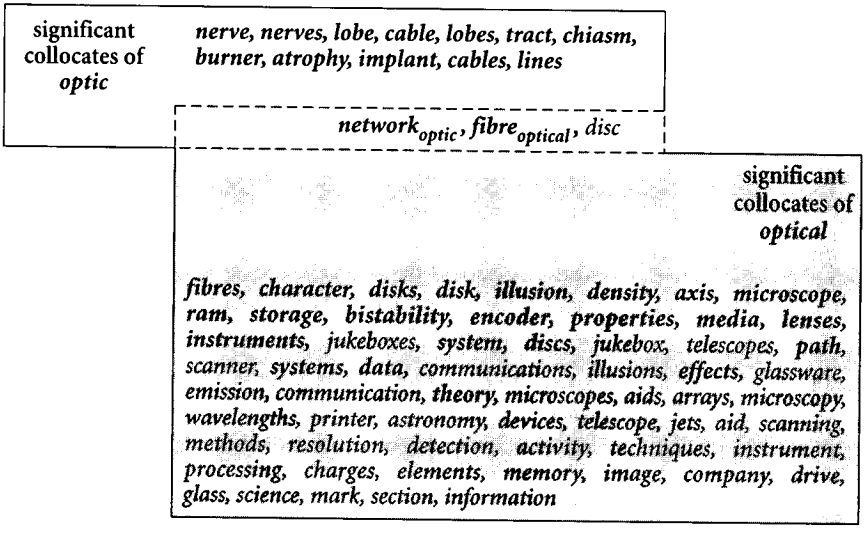

Figure 4. Significant and distinctive collocate sets for *optic(al)*

tic is more defined via *optical* than vice versa) is reflected, too, since Figure 1 and Figure 4 indicate that *optic* has less of an identity of its own (since *optical* can more often (in terms of relative frequencies) be substituted for *optic* than vice versa). Finally, although the degree of overlap is bidirectionally significant, the distribution of distinctive collocates supports the entries for *optic(al)* in CoCD and CCED, whose more pronounced distinction between the adjectives is clearly manifested in the data. On the other hand, while the distinction between human- or medicine/biology-related collocates (those of *optic*) and physical-device-related collocates (those of *optical*) is at least implied in CoCD and CCED it cannot be subsumed under any of the more general claims concerning *-ic*/*-ical* adjective pairs (cf. above Table 3) and, thus, seems to be specific to *optic(al)*. One might argue, in fact, that *arithmetic(al)* and *optic(al)* exhibit conflicting tendencies: while the former adjective pair displays a tendency to have the less strongly human-associated collocates co-occur with the *-ic* adjective, the latter has it exactly the other way round (cf. the above discussion of the data in Figure 3). While this analogy is far from perfect, it demonstrates once more that it is unlikely that we will ultimately find a unified basis for semantically distinguishing between all *-ic* adjectives and the corresponding *-ical* adjectives and that diachronic findings (cf., e.g., Kaunisto 2001) can shed light

Table 6. Reference works on *botanic(al)*

Adjective	Source	Entry
botanic	CoCD	"means the same as botanical"
	CCED	"means the same as botanical"
	CEDT	"botanical or botanic; of or relating to botany and plants"
	CaED	"of or pertaining to botany"
	OED	"Pertaining to the science or study of plants, to botany. (Now mostly superseded by botanical, exc. in names of institutions founded many years ago, as 'The Royal Botanic Society', 'The Botanic Gardens'.)"
botanical	CoCD	"is used to describe things relating to the scientific study of plants"
	CCED	"Botanical books, research, and activities relate to the scientific study of plants"
	CEDT	"botanical or botanic; of or relating to botany and plants"
	CaED	"of or pertaining to botany"
	OED	"Concerned with the study or cultivation of plants, pertaining to botany."

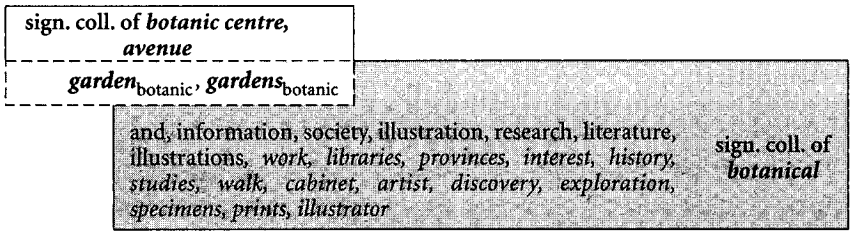

Figure 5. Significant and distinctive collocate sets for *botanic(al)*

on many if not all of these issues (since, in this case at least, all combinations of *optic(al)* and its R1 collocates are extremely recent).

Finally, let us briefly inspect *botanic(al)*. Table 6 summarises what references say about these adjectives and, on that basis, one would conclude that the adjectives are synonymous. Figure 1, by contrast, tells a different story: *botanical* can replace *botanic* significantly more often than expected, but *botanic* can replace *botanical* in an only insignificant number of cases. In other words, there must be a distributional/semantic difference, and this difference is fairly obvious (though difficult to put into words) once the significant and distinctive collocates are investigated.

The distinctive collocates of *botanic* are human-made locations characterised by the visible presence of (large amounts of) concrete plants etc. whereas the relations of the distinctive collocates of *botanical* to plants are less direct in that they are mostly concerned with representations or mental abstractions of concrete plants. In other words, *botanic(al)* is another instance of Marchand's (1969:242) and Hawkes's (1976:95) direct vs. less direct pattern even though the exact way of being 'less direct' is different from that of *arithmetic(al)*.

Let me briefly summarise what I believe to be the key findings of this section. We have addressed the question of how synonymous the members of *-ic/-ical* adjective pairs are with two different though naturally related techniques:

i. the $ESCO_2$ analysis and the *sub*-test (investigating the amount, directionality and significance of substitutability of *-ic/-ical* adjectives), and
ii. the distinctiveness of the sets of significant collocates by means of Church et al.'s (1991, 1994) *t*-test.

As a result, we have been able to

a. improve upon my previous results by discussing some findings in more detail;
b. demonstrate that several previous accounts to establish the suffixes' general meanings fail to account for adjective-specific properties: while the distinction direct vs. less direct proposed by Marchand (1969) and Hawkes (1976) intuitively seems to be an appropriate metaphor in some cases, it is nevertheless neither restrictive nor precise enough (covering both 'properties of humans doing X vs. properties of X' in the case of *arithmetic(al)* and 'concrete vs. abstract' in the case of *botanic(al)*);
c. illustrate the extent to which contemporary dictionaries lack important usage specifications for particular adjectives by pointing out distributional/semantic regularities that have hitherto gone unnoticed.

All these results are prima facie evidence for the utility of the *sub*-test in general and its applicability to the present issue in particular. However, the introductory section has already pointed out some caveats which need to be addressed in order for the $ESCO_2$ analysis and the *sub*-test (avoiding and using a test for significance respectively) to qualify as a valid measure of lexicographic analysis. These methodological issues will be dealt with in the following section.

4. Evaluation and improvement of the *sub*-test

4.1 The asymptotic test (*t*-test) vs. an exact test (Fisher exact test)

The previous section has shown how the *sub*-test can be applied to the semantic/distributional differentiation (or lack of it) of *-ic/-ical* adjectives. Let us recall, however, that one problem of the *sub*-test already mentioned in the first section is that it is not an exact test and that, although the corpus is very large, it is very likely to produce unreliable results given the small and skewed frequencies of the collocates involved. While the identification of collocations in the present study has been based on the log-likelihood test well known to be applicable to small frequencies, we have still used the *t*-test for (i) the identification of significant overlap and (ii) the identification of distinctive collocates. This section is therefore concerned with the question of if and how the results might be altered by using a statistically more appropriate test rather than the *t*-test proposed by Church et al. (1994: 168–171).

As is well known, the exact test that is probably most appropriate for the identification of significant contingency tables such as Table 1 above is the Fisher-exact test (cf., e.g., Pedersen 1996; Manning & Schütze 2000: 189; Weeber, Vos & Baayen 2000), which is in turn based on the hypergeometric distribution. It is applied to contingency tables in order to determine the cumulative probabilities of finding the observed result and all more extreme deviations (where extreme means 'in the direction of the H_1'[17]). In the example of Table 1, for instance, the exact probability of this result is $3.437 \cdot 10^{-21}$ and the sum of this probability and all more extreme ones (i.e. for all possible tables with more overlapping collocates and identical marginal totals is $3.67 \cdot 10^{-21}$; i.e. the observed distribution is highly significant and for all practical purposes impossible to obtain by chance). In other words, in this (very extreme) case used by Church et al. to exemplify their technique, Church et al.'s *t*-test and the Fisher-exact test do in fact yield identical results. Unfortunately, however, the 94 *sub*-test results contain many cases where the observed overlap is much smaller, which leads to the expectation that asymptotic tests will not yield reliable results for the degree of differentiation of *-ic/-ical* adjectives. Consider Table 7 for a comparison of the results of 94 *t*-tests and 94 Fisher-exact tests (47 adjective pairs and 2 directions for each pair).

As is immediately obvious, the results are strikingly different: when the statistically more appropriate Fisher-exact test is used the number of significant overlap results rises from 36 cases to 56 cases.[18] This difference is according to

Table 7. *Sub*-test results (*t*-test vs. Fisher-exact) for 47 pairs of *-ic/-ical* adjectives

outcome	significant (lack of) overlap		insignificant (lack of) overlap		Σ rows
test	*t*	F-E	*t*	F-E	
observed > expected	36	56	29	9	65
observed < expected	1	1	28	28	29
Σ columns	37	57	57	37	94

exact binomial tests highly significant, which is problematic in two respects: on the one hand, it undermines the *t*-test version of the *sub*-test (though of course not its general logic) by showing how strongly the asymptotic test deviates from the exact test. On the other hand, if we do accept the results of the exact test, the results seem counter-intuitive since suddenly nearly all of the cases where the overlap is larger than expected (where sometimes obs = 1 and exp = 0) are cases of significant collocate overlap, which contradicts our intuitions as well as prior lexicographical practice. In other words, it seems as if either the *sub*-test is, once conducted statistically most appropriately, much too sensitive, producing positive results for even the slightest overlap, or the pairs of *-ic/-ical* adjectives are on average indeed so similar to each other. The following section will be concerned with this issue.[19]

4.2 *-ic/-ical* adjectives vs. random adjectives

So far, we have concerned ourselves with the *sub*-test using two different statistical approaches, an asymptotic and an exact one. While the results seemed encouraging at first, we must not take the *sub*-test's results at face value without also looking at its diagnostic value; this is especially important after we have seen that, for the Fisher-exact test, even minimal overlap can already produce a significant result. Exploring this issue is the purpose of the present section.

Unfortunately, it is very difficult to establish a standard against which to judge the performance of the *sub*-test: In order to avoid circularity, we can of course not use (our) intuitions concerning *-ic/-ical* adjectives since this is precisely the object we want to investigate by means of the *sub*-test. Also, it is doubtful that naïve native speaker judgements (cf. Marsden 1985:31–32) will be useful since it will be difficult to (i) guarantee that the subjects are unaware of the purpose of the test and (ii) one can never be sure as to whether it is really semantic similarity that determines the subjects' ratings. What is possible, however, is to determine whether the *sub*-test is capable of distinguishing between cases where we would definitely expect some degree of similarity (as

with *-ic/-ical* adjectives, which, after all, share the same root) and cases where we would not (as with pairs of randomly chosen adjectives). More precisely, we have seen above that out of 65 cases where the observed overlap is larger than the expected one, the result was significant in as much as 55.4% and 86.2% for the *t*-test and Fisher exact respectively – if we find an even remotely similar percentage of significant overlap for random adjectives, we know that the *sub*-test is much too sensitive and, thus, worthless.

To that end, I carried out a simulation. I first constructed a pseudo-random sample of adjectives: for each of the 94 *-ic/-ical* adjectives, I chose randomly either another adjective with the same frequency or the (or another) adjective with the closest frequency. For instance, *political* is the most frequent member of a pair of *-ic/-ical* adjectives, occurring 29,630 times. The adjective in the BNCw with the frequency closest to 29,630 is *young*, occurring 28,963 times, which thus replaces *political* in the simulation. As a second example, *politic* occurs 31 times in the BNCw, as does the randomly chosen *smokeless*, which replaces *politic*. Thus, in order to simulate the two comparisons of *politic* and *political* (i.e. one in each direction), with respect to frequency similar adjectives *young* and *smokeless* were tested for (significant) overlap; consider the list of pseudo-randomly collected (frequency-controlled) adjectives in Appendix B for the remaining cases. Finally, for each of the 47 pairs of pseudo-randomly chosen adjectives, I determined the overlap of significant collocates and computed the *sub*-test (with both the *t*-test and the Fisher-exact test); the result is summarised diagrammatically in Figure 6.

As is obvious, these results are markedly different from those obtained for the *-ic/-ical* adjectives: the number of adjective pairs not exhibiting any significant collocate overlap at all is much higher and the number of cases where the overlap is both higher than expected and significant is much lower. For more precise, test-specific results, however, consider Table 8.

On a very general level, we find that the *-ic/-ical* adjectives differ strongly from the random adjectives: while the random adjectives yielded only 18 cases (=19.1%) where the overlap was larger than expected, we found 65 such cases (=69.1%) for *-ic/-ical* adjectives; i.e. the overall degree of distributional/semantic similarity of *-ic/-ical* adjectives is significantly higher. But let us now compare the results in Table 8 to those in Table 7 by looking at Figure 7, where "$\neq$" and "="mean "is not" and "is significantly different from" respectively (according to binomial tests).[20]

The difference between *-ic/-ical* and random adjectives is again borne out by the fact that in Figure 7 the *t*-test results and the Fisher-exact test results

Table 8. *Sub*-test results (*t*-test vs. Fisher-exact) for 47 random adjective pairs

outcome	significant overlap		insignificant (lack of) overlap		Σ rows
test	*t*	F-E	*t*	F-E	
observed > expected	5	8	13	*10*	18
observed < expected	0	0	76	*76*	76
Σ columns	5	8	89	*86*	94

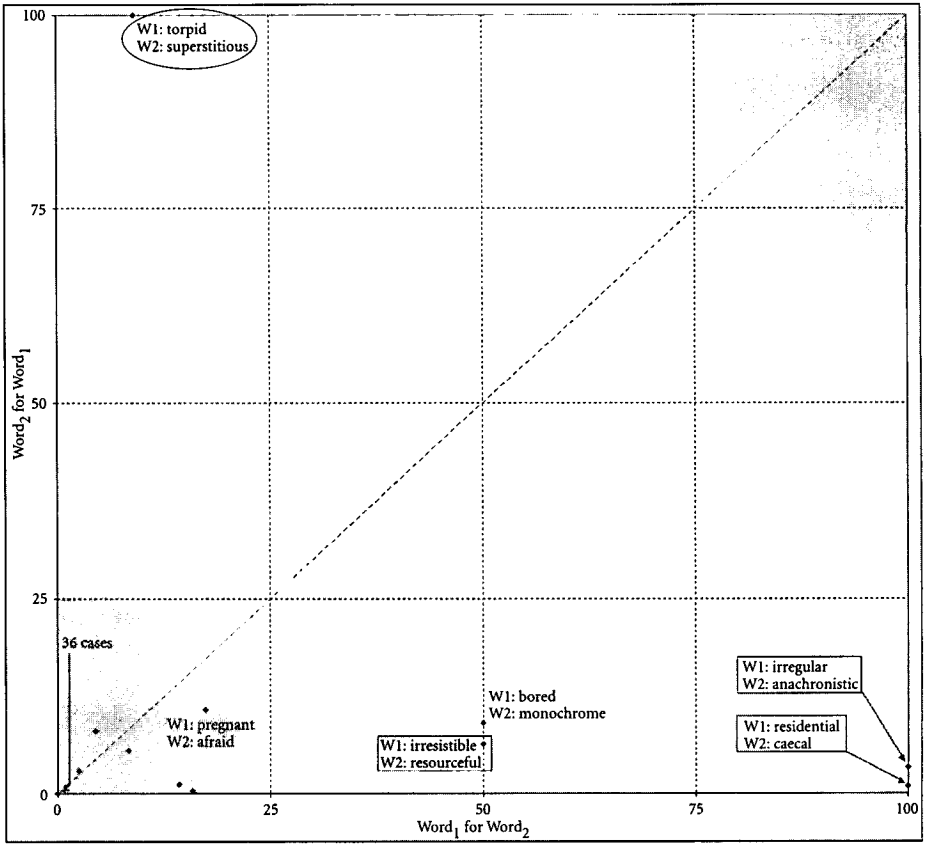

Figure 6. $ESCO_2$ results for 47 pairs of random adjectives[21]

within each test group (i.e. along the vertical dimension of Figure 7) differ significantly. Once we look at the horizontal dimension of Figure 7, however, we find an interaction between test type (*t*-test vs. Fisher-exact test) and adjective type (random vs. *-ic*/*-ical*): we already know from Table 7 that the *t*-test and

t_{random} ($^{5}/_{18}$)	$=^{22}$	Fisher-exact$_{random}$ ($^{8}/_{18}$)
$\neq^{23}$		$\neq$
$t_{\text{-}ic/\text{-}ical}$ ($^{36}/_{65}$)	$\neq$	Fisher-exact$_{\text{-}ic/\text{-}ical}$ ($^{56}/_{65}$)

Figure 7. *-ic/-ical* vs. random adjectives × *t*-test vs. Fisher-exact tests (obs. overlap > exp. overlap)

the Fisher-exact test yield considerably different numbers of significant results for *-ic/-ical* adjectives, but for random adjectives, the two tests yield similar results: in fact, the ratio of significant *t*-tests and Fisher-exact tests is identical: $^{36}/_{56} \approx^{5} /_{8} \approx 1.6$.

Given these results, we can return to the question posed at the end of the previous section: is the *sub*-test too sensitive or is it an appropriate measure (and *-ic/-ical* adjectives are really so similar to each other)? On the basis of the present results, I tend to suggest that an analysis of significant collocate overlap is in fact an appropriate technique for the analysis of distributional/semantic similarity. While the large number of significant overlap cases of *-ic/-ical* adjectives from the Fisher-exact test invites the inference that the *sub*-test is overly sensitive, the number of overlap of pairs of random adjectives is considerably lower and, thus, demonstrates that not all pairs of words from the same word class induce equally many significant results. Thus, the conclusion must be that the members of *-ic/-ical* adjective pairs are, on the whole, indeed so much more similar to each other and the diagnostic value of the $ESCO_2$ analysis is supported after having been put to the test. This is, of course, the 'desired' result, but one that Church et al. (1994) and Gries (2001) have not cared to provide, so the overall utility of this technique was still in doubt (cf. also the following section).

5. Conclusion

The present study pursued two objectives, a linguistic and a methodological one. As to the former, I hope to have demonstrated that

i. *-ic/-ical* adjectives can be placed on a two-dimensional continuum of semantic similarity, ranging from distributionally/semantically very similar cases to distributionally/semantically very dissimilar cases;

ii. particular adjective pairs can be located on this two-dimensional continuum on the basis of the overlap of significant R1 collocates, which is sensitive enough to identify both semantic as well as syntactically relevant distributional distinctions (e.g. human-related vs. non-human related collocates and attributive vs. predicative usage respectively);
iii. it is unlikely that a unified distinction of *-ic/-ical* adjective pairs will be detected as we have seen that the distinctions between the members of some pairs is largely contingent on the adjectives' semantics.

Given lack of space, only a few adjectives could be discussed, and the discussions offered here can only provide a starting point for further analyses. Consider, for instance, a more detailed analysis of *-ic* and *-ical* adjectives in terms of the distinction of attributive vs. predicative use. Wulff (unpubl. manuscript) analysed 14 of the most frequent *-ic/-ical* adjectives in the written parts of the BNC parts H and K in three steps. First, for each adjective, 50 occurrences from the corpus data were picked out randomly and checked for attributive or predicative usage (if an adjective occurred less than 50 times, all occurrences were analysed). Second, an additional sample of randomly chosen 300 adjective occurrences were checked analogously to obtain a rough estimate of the general ratio of attributive vs. predicative uses of adjectives; in her sample, 232 (77.3%) and 68 (22.7%) cases of attributive and predicative usage were found respectively.[24] Third, she tested whether the observed usage ratio of *-ic/-ical* adjectives matches the expected one (of the random adjectives). Her results are summarised in Table 9, where plusses/minuses in parentheses indicate whether the observed usage is more/less frequent than the expected one. Wulff (n. p.) concludes

> there is a very strong tendency for the *-ic* and *-ical* adjectives to be used attributively: [...] The 28 tests yielded 20 significant results, of which all but one show that the attributive use is even more frequent than the 77.33% to be expected. The exception is *politic*, where all 7 uses are predicative.

While the data base is probably still too limited, we already find that *economic* and *economical* as well as *lyric* and *lyrical* behave quite differently, which is little surprising for the former, but more so for the latter since contemporary reference works do not characterise the latter pair accordingly. As the most obvious refinements of analysis, I would therefore consider it necessary to widen the scope of analysis by not only focussing on R1 collocates, but by also including additional information such as part of speech tags in the immediate environment, syntactic parses etc. in order to identify further differ-

Table 9. *-ic/-ical* adjectives vs. random adjectives: Attributive vs. predicative usage

Adjective	attributive	predicative	$p_{binomial}$
botanical, classic, cyclic, economic, electrical, graphic, historical, political	50 (+)	0	<0.001
lyric	39 (+)	0	
logistical	36 (+)	0	
botanic, historic, logistic, magic	49 (+)	1	
numeric	33 (+)	1	0.0018
classical, electric, numerical	48 (+)	2	<0.001
graphical	47 (+)	3	0.0016
analytic	33 (+)	4	0.056 ns
analytical, cyclical	43 (+)	7	0.093 ns
magical	42 (+)	8	0.17 ns
symmetric	37 (–)	13	0.337 ns
lyrical	24 (–)	11	0.15 ns
economical, symmetrical	27 (–)	23	<0.001
politic	0 (–)	7	

ences. Also, other distributional curiosities await explanation. For instance, what is the motivation, if any, for the fact that the significant collocates of *-ical* adjectives are longer (in letters) than those of *-ic* adjectives ($t_{Welch} = 6.26$; $df = 3982$; $p < .0001$)?

On a slightly more theoretical level, the present study also contributes to the issue of whether substitutability or co-occurrence are more adequate measures of semantic similarity and, ultimately, synonymy. On the basis of their experimental evidence, Miller and Charles (1991:22–23) conclude that substitutability is superior to co-occurrence as a predictor of semantic similarity. However, both Gries (2001) and the present study demonstrate that even the simple co-occurrence approach yields interesting and telling results in this difficult area. Once the *sub*-test is augmented by the additional kinds of information mentioned above, it might well be the case that co-occurrence strategies towards semantic similarity fare much better than has previously been assumed. This would therefore be an interesting area for further research.

With regard to the methodological point, I believe the following conclusions are warranted. On the one hand, it has once again become clear that the difference between asymptotic tests and exact tests can be so strong as to change an analysis's outcome altogether. In the light of the massive differences obtained for the *-ic/-ical* adjectives, I believe that the gain in computational ease is outweighed by the risk of obtaining highly skewed results and, given to-

day's computational resources, the computational shortcut of asymptotic tests should be abandoned wherever possible; this holds for measures of collocational strength, distinctive collocates and the *sub*-test alike. On the other hand, however, it seems as if the question of whether the observed overlap is significantly greater than expected should not be overestimated since the kind and degree of semantic similarity can only be discussed satisfactorily anyway once the more qualitative analysis of distinctive collocates and direction of substitutability are applied to the adjectives under investigation. A significance level can, thus, serve as an indicator of which adjective pairs to investigate first, but nothing semantically crucial hinges on the fact of whether the finding of one overlapping collocate is significant for one adjective and insignificant for another. In other words, not everything needs to be tested for significance – but if one's question does in fact merit a distinction between significant and insignificant findings, then exact tests are doubtlessly called for in order to increase the likelihood of solid results. One obvious proposal in this connection (already pointed to in Note 19) is to test to what extent the identification of distinctive collocates is influenced by the choice of test.

Hopefully, the present study will increase the number of works devoted to the complex issue of *-ic/-ical* adjectives (or near synonyms in general, for that matter) in order to shed light on some probably only seemingly arbitrary patterns of usage. A particularly straightforward continuation of the line of research advocated here is to automate the *sub*-test identification of words having largely overlapping sets of collocates for the lexicographical purposes mentioned in the introduction. This research might produce, as a side effect, further methodologically relevant findings.

Notes

* I am very grateful for Heike Wagner's (University of Hamburg) computational assistance: without her Perl scripts, this paper would have taken much more time to complete. I also thank Anatol Stefanowitsch (University of Bremen) and Stefanie Wulff (University of Hamburg) for useful discussions and Constanze Bühner (Southern Denmark University) for her assistance. Finally, I am grateful for some useful comments by the IJCL reviewers, which helped to put some things into perspective. Naturally, I alone am responsible for any remaining shortcomings.

1. Synonymy is thus viewed as an extreme point on a continuum of semantic similarity (cf. also Miller & Charles 1991:2).

2. Note in passing that Church et al. do not round the figure for the expected numbers (2.69) to an integer (i.e. to 3). Although this is of course the usual practice, rounding would, strictly speaking, probably make more sense conceptually. I only mention it here because in some of the data below using a decimal as opposed to rounding to an integer is decisive for exceeding or failing to exceed the standard level of significance.

3. The latter fact, also discussed at length in Charles and Miller (1989) and Miller and Charles (1991), was also alluded to by one of the IJCL reviewers. This is no real weakness of the test, however, for two reasons. First, co-hyponyms do by definition exhibit a considerable degree of semantic similarity. Second, while it has been argued that antonyms are negatively similar to each other (Miller & Charles 1991:25), a different approach would be to argue that two antonyms are in fact similar to each other, differing only with respect to the dimension of which they denote opposite poles.

4. Church et al. (1994) do also mention the *t*-test, but the version of the *t*-test they discuss does not measure collocational strength as such but rather whether one of two words (e.g. *strong* and *powerful*) co-occurs with another word (e.g. *support*) significantly more often than the other.

5. Note in passing that this point of critique also applies to the *t*-test for the identification of distinctive collocates mentioned in Note 4 and used in Gries (2001); cf. Section 4.1 for an alternative.

6. For quantitatively more comprehensive information, cf. Gries (2001:102–103, Note 1).

7. More precisely, I developed an analytic technique called ESCO (Estimation of Significant Collocate Overlap) which corresponds to the *sub*-test without, however, testing the amount of overlap for significance. This is partially due to the fact that, at the time of development, I was unfortunately unaware of Church et al.'s *sub*-test (cf. Gries 2001:103, Note 6) and motivated the ESCO technique independently on the basis of Tversky's (1977) contrast model and Biber's (1993) analysis of polysemous words; early forerunners of Church et al.'s (1994) work on substitutability include Harris (1954) and Clark (1968).

8. The adjective pairs analysed were *analytic(al)*, *classic(al)*, *comic(al)*, *economic(al)*, *electric(al)*, *geometric(al)*, *graphic(al)*, *historic(al)*, *logistic(al)*, *lyric(al)*, *magic(al)*, *numeric(al)*, *politic(al)*, *problematic(al)* and *symmetric(al)*.

9. While the frequency of these adjectives is probably the most straightforward way to operationalise their representativity (in a particular corpus), it is not the only one. For instance, instead of relying on the most frequent adjectives for analysis, an equally possible approach would be to (also) consider the degree of dispersion of the adjectives in the corpus both in general and across different genres/registers. While these different measures will surely often coincide, they need not necessarily do so.

10. The adjective pair *mechanic(al)* was excluded since already a cursory look at the corpus data showed that many cases of *mechanic* that were tagged as an adjective were in fact (incorrectly tagged) nouns.

11. Other editing operations include the conversion of character entities (e.g. "dollar" → "$"), the deletion of tags, sentence numbers etc.

12. The deviations of the present results to those of Gries (2001) are due to the fact that (i) I tested only those bigrams for significance that occurred more than twice (while Gries [2001] tested those occurring more than once) and (ii) I used a slightly different measure of collocational strength. While the deviations are minor, they nevertheless give an idea of how strong the influence of such mere technicalities can in fact be. (The most striking difference is the present disproportionately higher number of adjective pairs without overlap.)

13. A look at each adjective's distinctive collocates (i.e. collocates which occur significantly more often with one adjective that the other, according to Church et al.'s *t*-test) suggests a meaning difference mentioned neither in Gries (2001) nor most reference works, namely *analytic* seems to be preferred for the humanistic philosophical meaning (e.g. *statements, tradition, philosophers, study, thinking* etc.) whereas *analytical* tends to be used more often with several natural sciences' expressions (e.g. *chemistry, results, laboratory, equipment, measurement* to name but a few). Whether this tendency is decisive or not requires further research.

14. Subscripts of overlapping collocates show with which adjective collocates occur significantly more often.

15. This observation needs to be taken with a grain of salt since sentence-final predicative use would not have survived the editing process described at the beginning of Section 3. Since this caveat applies to all adjectives alike, however, it only influences the quantitative dimensions of the findings, but not qualitative conclusions.

16. One of the IJCL reviewers took issue with my comments on *arithmetic(al)*, pointing out that "[i]t is true that [the] BNC does have *arithmetic operation* only (32x) and no arithmetical one, but there are, instead, similar a-al calculations, manipulation and processes to be found. Hence, it may be just scarcity of data (only 64 occurrences of a-al)." I agree completely that my results may be due to scarcity of data. However, I felt I should base my claims solely on the actual evidence drawn from the particular corpus chosen to investigate without regard to results one might obtain on the basis of additional/different data in order to avoid speculations for which I do not have empirical evidence.

17. Since one is mostly concerned with finding *a* or more cases of $word_1$ and $word_2$, usually, only the one-tailed probability is computed (as is also the case with the *t*-test discussed above); I followed this practice.

18. The probability to obtain 56 or more significant results out of 65 (from the Fisher-exact test) if we would expect 36 out of 65 (from the *t*-test) is 0 – likewise, the probability to obtain 36 or fewer significant results out of 65 (from the *t*-test) if we would expect 56 out of 65 (from the Fisher-exact test) is 0, too.

19. I have already mentioned above that the identification of distinctive collocates by means of Church et al.'s (1994) *t*-test also suffers from the drawback addressed in this section. When I replaced this variant of the *t*-test by an exact test (the binomial test, where the observed frequencies of collocates are checked against the probabilities of the node words), I obtained an at first sight surprising result: While the exact test for the *sub*-test led to a strong increase in significant distributions, the exact test for distinctive collocates led to a drastic decrease of distinctive collocates: More precisely, nearly all collocates labelled as distinctive by the

binomial test are also distinctive by the *t*-test, but (i) the binomial test mainly picks out only the most extreme cases and (ii) in very few cases, the binomial test labels collocates as distinctive that the *t*-test has not identified already.

20. A similar figure can be constructed for all 94 cases rather than only those where obs. overlap > exp. overlap. It is, however, identical to Figure 7 in all relevant respects, which is why I will not discuss it here.

21. The notation is as in Figure 1 above. No difference is made between results of *t*-tests and Fisher-exact tests.

22. For instance, the probability to obtain 5 or more cases of significant overlap out of 18 cases where the observed overlap is larger than expected (the *t*-test result) when we would expect to find 8 out of 18 (the Fisher-exact test result) is 0.117 ns; likewise the probability to obtain 8 or more cases of significant overlap out of 18 cases where the observed overlap is larger than expected (the Fisher-exact test result) when we would expect to find 5 out of 18 (the *t*-test result) is 0.097 ns.

23. That is, the probability to obtain 36 or more cases of significant overlap (with observed>expected) out of all 94 cases (the *t*-test result for *-ic/-ical* adjectives) when we would expect to find 5 out of 94 (the *t*-test result for random adjectives) is 0.

24. Although based on only a comparatively small sample, her results are strikingly similar to those of David Lee (personal communication).

References

Berry-Rogghe, G. L. M. (1974). Automatic Identification of Phrasal Verbs. In J. L. Mitchell (Ed.), *Computers in the Humanities* (pp. 16–26). Edinburgh: Edinburgh University Press.

Biber, D. (1993). Co-occurrence Patterns among Collocations: A Tool for Corpus-Based Lexical Knowledge Acquisition. *Computational Linguistics, 19*, 531–538.

Charles, W. G., & G. A. Miller (1989). Contexts of Antonymous Adjectives. *Applied Psycholinguistics, 10*, 357–375.

Church, K. W., & P. Hanks (1990). Word Association Norms, Mutual Information, and Lexicography. *Computational Linguistics, 16* (1), 22–29.

Church, K. W., W. Gale, P. Hanks, & D. Hindle (1991). Using Statistics in Lexical Analysis. In U. Zernik (Ed.), *Lexical Acquisition: Exploiting On-line Resources to Build a Lexicon* (pp. 115–164). Hillsdale, NJ: Lawrence Erlbaum.

Church, K. W., W. Gale, P. Hanks, D. Hindle, & R. Moon (1994). Lexical Substitutability. In B. T. S. Atkins & A. Zampolli (Eds.), *Computational Approaches to the Lexicon* (pp. 153–177). Oxford/New York: Oxford University Press.

Clark, H. H. (1968). On the Use and Meaning of Prepositions. *Journal of Verbal Learning and Verbal Behavior, 7*, 421–431.

Cobuild on CD-ROM (1995). Glasgow: Harper Collins Publ. [CoCD]

Collins English Dictionary and Thesaurus (1993). Glasgow: Harper Collins Publ. [CEDT]

Dunning, T. (1993). Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics, 19*, 61–74.

Collins Cobuild E-Dict (1998). Glasgow: Harper Collins Publ. [CCED]

Fournier, J.-M. (1993). Motivation savante et prononciation des adjectifs en *-ic* en anglais contemporain. *Faits de Langues, 1*, 235–240.

Fowler, H. W. (1968). *A Dictionary of Modern English Usage.* 2nd edition. London: Oxford University Press.

Gries, St. Th. (2001). A Corpus-linguistic Analysis of English *-ic* and *-ical* Adjectives. *ICAME Journal, 25*, 65–108.

Harris, Z. (1954). Distributional Structure. *Word, 10*, 146–162.

Hawkes, H. (1976). *-ic* and *-ical*: The Terrible Twins. *Lenguaje y Ciencias, 16*, 91–102.

Jespersen, O. (1942). *A Modern English Grammar. On Historical Principles – Part VI (Morphology).* London: George Allen & Unwin Ltd. [OJ]

Kaunisto, M. (1999). *Electric/Electrical* and *Classic/Classical*: Variation between the Suffixes *-ic* and *-ical. English Studies, 4*, 343–370.

Kaunisto, M. (2001). Nobility in the History of Adjectives Ending in *-ic* and *–ical.* In R. Brend, A. K. Melby & A. Lommel (Eds.), *LACUS Forum XXVII: Speaking and Comprehending* (pp. 35–46). Fullerton, CA: LACUS.

Marchand, H. (1969). *The Categories and Types of Present-Day English Word-Formation. A Synchronic-Diachronic Approach.* 2nd edition. München: Beck.

Manning, Chr. D., & H. Schütze (2000). *Foundations of Statistical Natural Language Processing.* 4th printing with corrections. Cambridge, MA: The M.I.T. Press.

Marsden, P. H. (1985). Adjective Pairs in *-ic* and *-ical*: Towards a Systematic Description of Current Usage. *Lebende Sprachen, 30*, 26–33.

Miller, G. A., & W. G. Charles (1991). Contextual Correlates of Semantic Similarity. *Language and Cognitive Processes, 6*, 1–28.

Oxford English Dictionary on CD-ROM (1994). Version 1.15, 2nd ed. Oxford: Oxford University Press. [OED]

Pedersen, T. (1996). Fishing for Exactness. *Proceedings of the SCSUG 96 in Austin, TX* (pp. 188–200).

Quirk, R., S. Greenbaum, G. Leech, & J. Svartvik (1985). *A Comprehensive Grammar of the English Language.* London/New York: Longman.

Ross, N. J. (1998). The *-ic* and *-ical* Pickle. *English Today, 14*, 40–44.

Snell, M. (1972). *German-English Prose Translation.* München: Hueber.

The Cassell Encyclopaedia Dictionary (1990). London: Cassell. [CaED]

Tversky, A. (1977). Features of Similarity. *Psychological Review, 84*, 327–352.

Weeber, M., R. Vos, & H. Baayen (2000). Extracting the Lowest-Frequency Words: Pitfalls and Possibilities. *Computational Linguistics, 26*, 301–317.

Wulff, St. (unpublished manuscript). *-ic* and *-ical*: A Corpus Analysis. University of Hamburg.

Appendix A. *-ic/-ical* adjectives in the written part of the BNC

Adjective	N in BNCw	Adjective	N in BNCw	Adjective	N in BNCw
politic	31	problematic	734	microscopic	229
political	29630	problematical	126	microscopical	30
economic	22936	ironic	692	psychoanalytic	231
economical	472	ironical	85	psychoanalytical	26
historic	2022	geometric	583	pedagogic	129
historical	5329	geometrical	190	pedagogical	125
electric	2760	poetic	681	asymmetric	135
electrical	2057	poetical	56	asymmetrical	110
classic	1614	symmetric	352	logistic	142
classical	3175	symmetrical	323	logistical	93
strategic	2694	comic	493	canonic	26
strategical	22	comical	128	canonical	183
mechanic	273	mystic	118	conic	10
mechanical	1875	mystical	499	conical	187
geographic	269	rhythmic	499	uneconomic	145
geographical	1560	rhythmical	69	uneconomical	48
genetic	1734	endoscopic	485	topographic	49
genetical	16	endoscopical	17	topographical	136
magic	913	cyclic	277	stratigraphic	58
magical	835	cyclical	187	stratigraphical	109
tropic	24	botanic	164	elliptic	13
tropical	1659	botanical	283	elliptical	115
dynamic	1473	psychic	419	anarchic	108
dynamical	94	psychical	27	anarchical	15
philosophic	66	metric	367	hemispheric	67
philosophical	1251	metrical	78	hemispherical	40
graphic	630	lyric	104	typographic	41
graphical	629	lyrical	258	typographical	61
optic	185	mythic	97	obstetric	84
optical	912	mythical	232	obstetrical	6
analytic	219	bibliographic	199	paramedic	36
analytical	781	bibliographical	127	paramedical	33
fantastic	932	arithmetic	227	haemorrhagic	50
fantastical	46	arithmetical	57	haemorrhagical	7
numeric	215	metaphoric	74	druidic	8
numerical	684	metaphorical	192	druidical	14
syntactic	836	alphabetic	41	synodic	5
syntactical	34	alphabetical	223	synodical	5

Appendix B. List of random adjectives for the simulation performed in Section 4.2

young	cumulative	irregular	packed	subsidiary
main	danubian	irresistible	pliable	suggestive
afraid	determined	irrevocable	post-graduate	superstitious
grand	disadvantaged	laborious	pregnant	systematic
residential	distal	left-handed	promotional	technological
widespread	distinguishable	lighthearted	proverbial	thoughtful
additive	double-sided	meiotic	rebellious	thunderous
advancing	editing	mesopotamian	recyclable	tiresome
adversarial	equine	monochrome	reedy	torpid
allowable	filial	monotonous	relentless	ugly
amazed	frenetic	moveable	renewed	unconnected
anachronistic	grant-aided	mughal	resourceful	undergraduate
baltic	gruelling	murky	running	unexpected
bored	imperfect	newsworthy	satisfied	unfortunate
caecal	inclusive	niggling	scared	uninteresting
childlike	infinite	nominal	scattering	unrepentant
collective	innovatory	observant	smokeless	vacant
compensatory	intermolecular	off-peak	socratic	wishful
considerate	irrational	oppressed	submissive	