Co-occurrences of Antonymous Adjectives and Their Contexts

John S. Justeson*
Slava M. Katz*
IBM T. J. Watson Research Center

Charles and Miller propose that lexical associations between antonymous adjectives are formed via their co-occurrences within the same sentence (the co-occurrence hypothesis*), rather than via their syntactic substitutability (the* substitutability hypothesis*), and that such co-occurrences must take place more often than expected by chance. This paper provides empirical support for the co-occurrence hypothesis, in a corpus analysis of all high-frequency adjectives and their antonyms and of a major group of morphologically derived antonyms (e.g.,* im-possible, un-happy*). We show that very high co-occurrence rates do appear to characterize all antonymous adjective pairs, supporting the precondition for the formation of the association; and we find that the syntactic contexts of these co-occurrences raise the intrinsic associability of antonyms when they do co-occur. We show that via one of these patterns, mutual substitution within otherwise repeated phrases in a sentence, the co-occurrence hypothesis captures the generalizations that were the basis for the substitutability hypothesis for the formation of antonymic associations.*

1. Antonymic Association

Much current research in linguistics is concerned with textual or discourse bases for linguistic structure; within lexical semantics, such research is directed at particular lexical relations and at correlations between syntax and semantics. This paper addresses the textual underpinnings of antonymy between predicative adjectives, following up research reported by Charles and Miller (1989).

Antonymy is a special lexical association between word pairs. That it is lexical and not simply semantic follows from the fact that different words for the same concept can have different antonyms; for example, *big-little* and *large-small* are good antonym pairs, but *large-little* is not. The classic work on associations among adjectives, and between antonymous adjectives in particular, is by Deese (1964, 1965), analyzing the results of stimulus-response word-association tests.[1] Charles and Miller (1989) argue, contrary to more complex psycholinguistic theories, that the primary source of these associations is a tendency they hypothesize for antonyms to co-occur within the same sentences in discourse. This paper supports and extends their hypothesis.

Deese's work on word association for adjectives was based on a list of the 278 most frequent adjectives in the Thorndike-Lorge (1944) count, those having a frequency of 50 per million or more. Thirty-four pairs of these words had a reciprocal property: each member of the pair was the most frequent response to the other on word-association

* Natural Language Group, IBM T. J. Watson Research Center, P.O. Box 704, Yorktown Heights, New York 10598

1 When we use the term *association*, we use it in the general sense of associative pairing, not for response frequencies in word-association tests.

tests. These pairs were all antonymic (e.g., *good-bad, big-little, large-small*). Essentially reciprocal were 5 additional antonym pairs overlapping with these 34: *easy-hard* (cf. *hard-soft*), *heavy-light* (cf. *light-dark*), *left-right* (cf. *right-wrong*), *long-short* (cf. *short-tall*), and *new-old* (cf. *old-young*). For these pairs, member *A* has *B* as its most frequent response, while *B* is the second most frequent response to *A* (*A*'s most common response being its antonym in the original list of 34 pairs).

A general finding of word-association studies is that the part of speech of the response usually agrees with that of the stimulus. Deese's data are consistent with this, and suggest further that at least the most frequent antonymous adjectives are associated more directly with each other than with other adjectives. Deese (evidently) did not find antonymic responses to be the most frequent for some 200 of the 278 adjectives he studied. These were generally of lower frequency than those that were each other's most common responses; Deese (1962, 1964:347) showed that the most common responses to lower-frequency adjectives are typically nouns that those adjectives often modify, as in *administrative decision*, cited by Deese and actually used in the Brown Corpus (see Section 2.1). In addition, the words with reciprocal responses were almost all morphologically simple and etymologically native (i.e., of Old English origin).

In human contingency judgment studies (see Shanks and Dickinson 1987 for a recent review), the formation of associations among events is found to depend upon temporally close occurrence of the associated events. Lexical associations, and in particular the word pairs linked as stimulus and response in word-association tests, have thus been accounted for by positing frequent temporal proximity of lexical access of the words involved; Charles and Miller contrast two such accounts, the *substitutability* hypothesis, introduced by Ervin-Tripp (1961, 1963), and their own *co-occurrence* hypothesis.

Because responses to higher-frequency stimuli in association tests tend to agree in part of speech, and because parts of speech are defined linguistically by mutual substitutability within a context, psycholinguistic accounts have sought explanations for antonymic associations as side-effects of part-of-speech associations; this is an economical kind of account, since part-of-speech agreement requires an explanation in any event. The key idea of the substitutability hypothesis is that the context in which one word occurs in a sentence may call to mind other words syntactically substitutable in that context, so that substitutable words are activated mentally in close temporal proximity. Under such a hypothesis, the greater the syntactic and semantic appropriateness of a substitutable word in a context, the greater its mental activation. Part-of-speech agreement often suffices for syntactic appropriateness; and, intuitively, the use of an adjective that characterizes the position of a noun's referent in one region of a semantic dimension seems sensible mainly when another region cannot be ruled out as inconsistent. Accordingly, antonymous adjectives should frequently be called to mind in each other's environment.

Charles and Miller (1989) dispute the substitutability of antonymous adjectives, showing that their sentential contexts normally leave only one member of the pair plausible. They suggest instead the simplest form of associative explanation, that the forms in question do in fact tend to co-occur in close temporal contiguity — in particular, that antonymous adjectives occur more frequently than expected by chance in the same sentence. As supporting evidence they provide counts showing a higher than expected number of sentences containing both members of each of the antonym pairs *big-little, large-small, public-private,* and *strong-weak*;[2] they conclude that it is sentential

2 They also hypothesize that adjective pairs that are semantically opposite but not antonymic, one

co-occurrence alone that is responsible for the associative pairing of antonymous adjectives. This paper confirms that, for all adjectives frequent enough to judge, antonymous adjectives do co-occur within the same sentence much more often than is expected by chance. On the other hand, something like the substitutability hypothesis must also be involved: when antonyms do co-occur, they usually substitute for one another in clausal or phrasal contexts that are otherwise word-for-word repetitions of each other, apart from pronominalization or ellipsis in some cases. We propose that, while training for antonym association takes place via co-occurrence, it is their substitution in repeated contexts that makes this training particularly effective.

2. Co-occurrences among Antonyms

This section shows that adjectives do indeed tend to occur in the same sentence as their antonyms far more frequently than expected by chance. Using the 1,000,000-word Brown Corpus, Charles and Miller (1989) tested this hypothesis for four pairs of antonyms: *big-little, large-small, strong-weak,* and *public-private,* the first three pairs being among Deese's list of reciprocally-associated antonyms. We verify their results for these four pairs, and extend the demonstration to encompass all of the Deese antonyms. On a class basis, we extend it to encompass all pairs of antonymous adjectives for which at least one member has an adjectival frequency comparable to the Deese antonyms and to a major group of morphologically antonymous adjectives. For all three groups of adjective pairs, antonyms turn out to co-occur sententially roughly 10 times as often as expected by chance. These results are obtained mainly from a version of the Brown Corpus tagged by part of speech, and have been checked against an untagged 25,000,000-word corpus of general literature.

2.1 Test Corpora

Our work is based primarily on the Brown Corpus, a database containing 1,000,000 words of English text balanced across 15 general categories, divided into 500 text extracts of about 2,000 words each; for a detailed description, see Francis and Kučera (1982). In the version we use, all words are tagged essentially by part of speech. Thus, although the word *back* is used not only as an adjective but also as a noun, verb, and adverb, we are able to recover exclusively adjectival uses of the word from the corpus. The tagging system actually employs some distinctions finer than simple part of speech, so certain adjectives receive special tags; for example, *first* and *last* are tagged as ordinal numbers rather than as adjectives. On the other hand, some part-of-speech labels are used more broadly than under most conventional grammatical descriptions. For example, past participles of verbs are tagged as such in both verbal and adjectival use; thus, *married* is not marked as an adjective in the corpus although it does occur adjectivally (e.g., *a married couple*). For simplicity, consistency, and replicability, we consider only those adjectives marked with the standard adjective tag in the corpus.

To test hypotheses concerning sentential co-occurrence of words, the corpus must be divided into sentences. The Brown Corpus is divided by a special sentence-ending tag into a sequence of "pseudo-sentences." We eliminate those pseudo-sentences that are labeled as "headlines," since they rarely consist of full sentences. Quantitatively

member being a synonym of an antonym of the other, are not directly associated, and suggest that these "indirect antonyms" have no special tendency to co-occur sententially. They confirm this suggestion for four pairs they take to be good indirect antonyms: *powerful-weak, strong-faint, big-small, large-little.* In none of these cases do they find a substantial excess of sentential co-occurrences over what is expected by chance. For further discussion see Gross, Fischer and Miller (1989).

this has little effect; their inclusion would add less than 1% to the 54,717 sentences that remain. In a sample checked for incorrect sentence divisions, we found that all were cases of premature division, none of sentence joins; these errors are mainly list elements, separated by semicolons, that are treated as sentences, and subject-verb segments that are separated from quoted speech by either exclamation points or question marks.

As a result of these errors, some of the ratios we report of observed to expected numbers of sentential co-occurrences of antonyms may be somewhat higher or lower than they should be. We have made no attempt to correct for any inaccurate sentence divisions, finding that any increase is too slight to affect our results (and any decrease will bias against the co-occurrence hypothesis). If the reported ratios are higher than they should be, preliminary analysis shows that the increase is probably by no more than 3%; if they are lower (which is less likely in general), the decrease is always by more than 3% since the largest number of co-occurrences in our sample is 28. The co-occurrence hypothesis calls for high observed/expected ratios; a 3% increase in these ratios does not significantly bias in favor of the hypothesis, since the ratios we calculate are all more than 2:1 and average around 10:1.

We have also checked our main results on the APHB Corpus, a much larger but grammatically untagged corpus of 25,000,000 words, obtained from the American Publishing House for the Blind and archived at IBM's Watson Research Center. It consists of stories and articles from books and general circulation magazines, such as *Reader's Digest*, *Datamation*, and *Fortune*. Sentence separation is less reliable than in the Brown Corpus, and words are not tagged by part of speech. In addition, since the corpus is untagged, computations of the expected numbers of co-occurrences of antonymic adjectives are necessarily inflated except when nonadjectival use of both adjectives is rare. We used this corpus primarily to verify the results derived from the Brown Corpus, and for problems requiring a sample size much larger than 1,000,000 words.

2.2 The Deese Antonyms

For 4 of the antonym pairs identified by Deese (1964) we have no evidence because at least one member of the pair does not occur with the standard adjectival tag in the Brown Corpus: *alone-together*, *few-many*, *first-last*, and *married-single*. These have no effect on our counts. For the remaining 35 pairs we have determined the number of sentences containing each member of the pair; the number containing both members; the number expected to contain both members; the ratio of the observed to expected co-occurrences; the rate of sentential co-occurrence; and the probability that as many or more sentences would contain both members as were found to. The results, given in Table 1, show an overwhelming excess of observed over expected numbers of co-occurrences. The following sentences are typical of the antonym co-occurrences:

> *The group sets the styles in clothing, the kind of play engaged in, and the ideals of* ***right*** *and* ***wrong*** *behavior.*
>
> *Soil redeposition is evaluated by washing* ***clean*** *swatches with the* ***dirty*** *ones.*
>
> *Originals are not necessarily* ***good*** *and adaptations are not necessarily* ***bad.***

Overall, antonym co-occurrence takes place in more than 8.6 times as many sentences as expected,[3] and co-occurrences are found for 30 out of the 35 pairs, in spite

3 This estimate is probably too low. For one pair (*old-new*) the expected number of sentences (10.4) amounts to 39% of the total number (27.0) expected, and it happens to have the lowest ratio of

Table 1
Deese's adjective pairs and their sentential co-occurrences in the tagged Brown Corpus. *Sentential occurrences* of an adjective is the number of sentences in which the adjective occurs in the corpus; this number is given in columns 1 and 3, once for each member of an antonym pair. *Observed* is the number of sentences in which both adjectives occur; *expected* is the number expected to have both adjectives by chance; *ratio* is the ratio of observed to expected co-occurrences; *rate* $1/n$ indicates that one sentence out of n that have the less frequent adjective produces a co-occurrence with its antonym; and *probability* is the probability of observing by chance as many or more co-occurrences than are actually observed.

sentential occurrences of individual adjectives				sentential co-occurrences				
				observed	expected	ratio	rate	probability
85	active	11	passive	2	0.01709	117.0	1/5.5	1.30×10^{-4}
55	alive	157	dead	2	0.15781	12.7	1/27.5	1.10×10^{-2}
25	back	76	front	3	0.03472	86.4	1/8.3	5.79×10^{-6}
125	bad	682	good	16	1.55802	10.3	1/7.8	5.06×10^{-12}
316	big	273	little	12	1.57662	7.6	1/22.8	8.18×10^{-8}
146	black	243	white	22	0.64839	33.9	1/6.6	2.84×10^{-27}
3	bottom	69	top	0	0.00378	–	–	–
46	clean	36	dirty	1	0.03026	33.0	1/36.0	2.98×10^{-2}
136	cold	119	hot	7	0.29578	23.7	1/17.0	2.23×10^{-8}
147	dark	61	light	5	0.16388	30.5	1/12.2	6.89×10^{-7}
83	deep	14	shallow	0	0.02124	–	–	–
52	dry	45	wet	2	0.04277	46.8	1/22.5	8.54×10^{-4}
109	easy	150	hard	0	0.29881	–	–	–
63	empty	215	full	1	0.24755	4.0	1/63.0	2.20×10^{-1}
35	far	16	near	1	0.01023	97.7	1/16.0	1.02×10^{-2}
30	fast	48	slow	1	0.02632	38.0	1/30.0	2.60×10^{-2}
89	happy	32	sad	1	0.05205	19.2	1/32.0	5.08×10^{-2}
150	hard	59	soft	3	0.16174	18.5	1/19.7	5.87×10^{-4}
107	heavy	61	light	1	0.11929	8.4	1/61.0	1.13×10^{-1}
407	high	137	low	20	1.01904	19.6	1/6.9	3.95×10^{-20}
6	inside	38	outside	0	0.00417	–	–	–
347	large	504	small	26	3.19623	8.1	1/13.3	4.33×10^{-16}
122	left	231	right	28	0.51505	54.4	1/4.4	1.27×10^{-40}
508	long	187	short	12	1.73613	6.9	1/15.6	2.21×10^{-7}
60	narrow	113	wide	2	0.12391	16.1	1/30.0	6.92×10^{-3}
1001	new	569	old	28	10.40936	2.7	1/20.3	3.07×10^{-6}
569	old	357	young	17	3.71243	4.6	1/21.0	2.83×10^{-7}
101	poor	69	rich	7	0.12736	55.0	1/9.9	5.80×10^{-11}
40	pretty	20	ugly	0	0.01462	–	–	–
231	right	112	wrong	8	0.47283	16.9	1/14.0	2.91×10^{-8}
40	rough	35	smooth	1	0.02559	39.1	1/35.0	2.53×10^{-2}
187	short	55	tall	1	0.18797	5.3	1/55.0	1.72×10^{-1}
1	sour	62	sweet	1	0.00113	882.5	1/1.0	1.13×10^{-3}
189	strong	29	weak	3	0.10017	29.9	1/9.7	1.39×10^{-4}
63	thick	90	thin	1	0.10362	9.7	1/63.0	9.86×10^{-2}

of the small size of the corpus. For most of the 30 antonym pairs exhibiting sentential co-occurrences, the numbers of co-occurrences are statistically significant: 25 are

observed to expected co-occurrences among all pairs for which any co-occurrence was observed. Excluding this pair as an outlier yields an overall ratio of about 11.0 times as many sentences as expected with the co-occurrence.

significant at the .05 level; 20 at the .01 level; and in 18, the probability of obtaining so many co-occurrences is less than 10^{-4}. The occurrence of nonsignificant results rises as the expected number of co-occurrences declines. This suggests that the nonsignificant results are attributable in part to sample size. Indeed, testing in our larger but untagged corpus of 25,000,000 words, all 35 pairs do yield highly significant numbers of co-occurrences.[4]

Antonym co-occurrence is a mass phenomenon affecting the entire Deese list; it is not due mainly to the contributions of a few cases with idiosyncratically high co-occurrence rates as a result of their having some special property. We (over)control for the possibility of undue effect of high co-occurrence numbers from particular antonym pairs by computing the probability that 30 or more of the 35 pairs would co-occur in at least one sentence; in this computation, high numbers of co-occurrences have no effect. This probability turns out to be negligible, 7.1×10^{-23}. A simulation provided a picture of the distribution of the number of pairs having at least one co-occurrence by chance; only 9.4 antonym pairs were expected to exhibit a co-occurrence, and in none of a million trials did even 20 pairs do so.

To get an idea of the acquisitional consequences of these results it is useful to consider the *rates* of co-occurrence being found. The maximum number of potential sentential co-occurrences is the minimum of the number of sentences containing each member of the antonym pair; calculating the rate of co-occurrence as the proportion of observed to possible co-occurrences, we find an overall rate of one co-occurrence every 14.7 sentences. At such rates, the language learner is repeatedly exposed to training for the association.

The co-occurrence rates are higher if one controls for word sense. For example, only 24 of the 62 sentences containing the adjective *sweet* refer to taste or smell and thus readily contrast with *sour*; the others refer to attitudes, personalities, and music. This is most apparent in the case of the five adjectives that occur twice in the table, each with two different antonyms depending on the sense. Taking account of this, the low rate of 1 occurrence of *short* per 55 sentences containing *tall* rises to 1 occurrence of *tall* per 6 sentences containing *short*, since only 6 of the 187 sentences containing *short* use it to refer to height. Similarly, in only 310–321 sentences does *old* relate to newness, as opposed to youth;[5] restricted to these sentences, the co-occurrence rate increases to at least one occurrence of *new* per 11.1–11.5 sentences containing a relevant sense of *old*, and the observed/expected ratio from 2.7 to a more typical 4.8–4.9. The co-occurrence rate for *old-young* also rises, to one per 12.2–12.8 sentences, and the observed/expected ratio rises to 12.0–12.6.

2.3 Other High-Frequency Antonym Pairs

In addition to the words in Deese's list, we tested for sentential co-occurrence a set of antonym pairs one of whose members had a frequency of at least 50 in the 1,000,000-word Brown Corpus but that were not among the Deese antonyms. This set was constructed as follows. From a list of all adjectives with a frequency of at least 50, all

4 The lower rates in the case of low-frequency adjectives cannot produce statistically significant results in a corpus of only 1,000,000 words. Twelve of the 15 pairs that failed to produce statistically significant numbers of co-occurrences in the Brown Corpus were among the 14 (40% of) pairs having the lowest co-occurrence rates in the 25,000,000-word corpus. This is exacerbated by a weaker concentration of 13 of these 15 pairs among the 21 (60% of) pairs having the lowest minimum adjective frequencies.

5 The range is due to ambiguity in assignment of certain cases of *old*, as in *He came to a stretch of* ***old*** *orange groves, the trees dead, some of them uprooted, and then there was an outlying shopping area, and tract houses.*

adjectives were removed that did not seem to have any clear antonym, or that were comparative or superlative. For the remainder, one or more antonyms were placed on a candidate list. This list was given to a professional lexicographer, who confirmed most of the judgments and revised others. The original list, expanded by the additions but not reduced by the deletions made by the lexicographer, was then reviewed by a linguist specializing in lexical semantics. Finally, the list as expanded by her judgments was similarly critiqued by an elementary school teacher.

The set of 35 adjective pairs judged to be good lexical antonyms by all three reviewers we are confident in labeling as antonymic; analysis was restricted to these, at the expense of possibly excluding a handful of plausible candidate pairs.[6] Eleven of these antonym pairs[7] are *morphological antonyms*, that is, one member is derived from the other by a prefix of negation, *in-* (also in the forms *il-*, *im-*, *ir-*) or *un-*. For a word pair whose antonymy is morphologically marked, acquisition of the antonymy relation for that pair does not require an associative hypothesis involving those particular words. We therefore treat morphological antonyms separately (Section 2.4), and exclude them from the analysis of high-frequency, morphologically arbitrary antonym pairings. This leaves a set of 24 new antonym pairs to test, at least one of whose members occurs adjectivally in no fewer than 50 sentences in the Brown Corpus (see Table 2). Two of the 24 pairs on the list have no effect on the results. Because past participles in adjectival use are not labeled in the corpus as adjectives, we lack data for the pairs *open-closed* and *theoretical-applied*. For the 22 pairs for which the corpus provides data, 14 co-occur sententially. The overall co-occurrence rate (once per 13.3 sentences) is comparable to that (14.7) for the Deese adjectives; there is no significant difference as judged by the χ^2 statistic. Again, word-by-word statistics confirm about half of the pairings to be statistically significant. Again, nonsignificant results are for pairs with very low expected co-occurrences, and their sentential co-occurrence rate is similar to that for the significant (but more frequent) pairs. Finally, the result is again not due to any concentration of the observed cases in a few vocabulary items: for most pairs the number of co-occurrences is quite small, and the probability that as many or more than 14 of the pairs would exhibit at least one co-occurrence is negligible, 2.1×10^{-10}. Simulation shows that the number of pairs expected to show a co-occurrence was just 2.9.

A high rate of co-occurrence for antonyms is therefore validated not only for those adjectives that Deese showed to have the strongest reciprocal associative structure, but also for antonyms involving frequent adjectives in general.

2.4 Morphological Antonyms

This section addresses morphological antonyms, adjective pairs in which one member is derived from the other by an affix of negation. In particular, we treat those morphological antonyms in which the derivation is by any of the most productive prefixes of negation (*a-*, *ab-*, *an-*; *dis-*; *il-*, *im-*, *in-*, *ir-*; *un-*; and *non-*).[8] Morphological antonyms require separate treatment because their lexical antonymy is recoverable morphologically, hence their lexical association can be acquired differently from that between nonmorphological antonyms. We show here that morphological antonyms show the same tendency for greater than chance numbers of co-occurrences that character-

6 In our judgment, the best candidates among those not accepted by all three of our judges were *conservative-liberal, open-shut, primary-secondary, religious-secular, safe-dangerous,* and *typical-atypical.*

7 *Able-unable; complete-incomplete; common-uncommon; fair-unfair; legal-illegal; personal-impersonal; practical-impractical; responsible-irresponsible; safe-unsafe; useful-useless;* and *usual-unusual.*

8 We exclude morphological affix pairs that are themselves antonymous, e.g., *-ful* and *-less, pro-* and *anti-*.

Table 2
Other high-frequency adjective pairs and their sentential co-occurrences in the tagged Brown Corpus. *Sentential occurrences* of an adjective is the number of sentences in which the adjective occurs in the corpus; this number is given in columns 1 and 3, once for each member of an antonym pair. *Observed* is the number of sentences in which both adjectives occur; *expected* is the number expected to have both adjectives by chance; *ratio* is the ratio of observed to expected co-occurrences; *rate* $1/n$ indicates that one sentence out of n that have the less frequent adjective produces a co-occurrence with its antonym; and *probability* is the probability of observing by chance as many or more co-occurrences than are actually observed.

sentential occurrences of individual adjectives				sentential co-occurrences				
				observed	expected	ratio	rate	probability
21	absent	223	present	0	0.08559	–	–	–
62	ancient	184	modern	4	0.20849	19.2	1/15.5	5.93×10^{-5}
122	beautiful	20	ugly	0	0.04459	–	–	–
82	broad	60	narrow	1	0.08992	11.1	1/60.0	8.61×10^{-2}
55	busy	10	idle	0	0.01005	–	–	–
58	complex	158	simple	5	0.16748	29.9	1/11.6	7.63×10^{-7}
48	cool	64	warm	1	0.05614	17.8	1/48.0	5.47×10^{-2}
2	crooked	55	straight	0	0.00201	–	–	–
160	difficult	109	easy	3	0.31873	9.4	1/36.3	4.11×10^{-3}
26	dull	71	sharp	0	0.03374	–	–	–
240	early	128	late	12	0.56143	21.4	1/10.7	5.83×10^{-13}
28	false	220	true	5	0.11258	44.4	1/5.6	9.15×10^{-8}
287	general	113	specific	4	0.59270	6.7	1/28.3	3.05×10^{-3}
52	inner	28	outer	6	0.02661	225.5	1/4.7	2.03×10^{-13}
14	loud	59	soft	0	0.01510	–	–	–
101	lower	66	upper	9	0.12183	73.9	1/7.3	5.86×10^{-15}
217	major	43	minor	6	0.17053	35.2	1/7.2	1.96×10^{-8}
64	maximum	36	minimum	4	0.04211	95.0	1/9.0	9.74×10^{-8}
49	negative	71	positive	11	0.06358	173.0	1/4.5	2.18×10^{-22}
6	noisy	62	quiet	0	0.00680	–	–	–
174	private	264	public	13	0.83952	15.5	1/13.4	3.89×10^{-12}
50	sick	13	well	0	0.01188	–	–	–

izes nonmorphological antonyms. Typical sentences containing these co-occurrences are

Plato feels that man has two competing aspects, his ***rational*** *faculty and his* ***irrational.***

Choose carefully between ***contributory*** *or* ***non-contributory*** *pension plans.*

Judgment of antonymy is not nearly as subjective an issue in the case of morphological antonyms, so we composed the list to be tested ourselves. We identified 662 adjectives in the Brown corpus as being marked negative by one of these prefixes and as having a base that also functions adjectivally. For 27 of these pairs (e.g., *famous-infamous*), our judgment was that their morphological antonymy does not entail semantic oppositeness;[9] these spurious pairs were excluded from consideration. For nearly

9 Clear cases rejected on these grounds are *ionic-anionic; pathetic-apathetic; septic-aseptic; stringent-astringent; trophic-atrophic; graceful-disgraceful; material-immaterial; memorial-immemorial; passive-impassive; pertinent-impertinent; perturbable-imperturbable; different-indifferent; famous-infamous; fluent-influent;*

half of the 635 remaining antonym pairs we have no evidence in the Brown Corpus, since the unnegated member of the pair fails to occur. Our analysis therefore concerns the 346 semantically and morphologically antonymous adjective pairs for which each member occurs in the corpus.

Overall, adjectives on this list have very low frequencies; in fact, 36 (10%) of the morphologically positive and 154 (45%) of the morphologically negative adjectives occur only once in the corpus. With frequencies so low, both the expected and observed sentential co-occurrences of these antonyms are too low to test the co-occurrence hypothesis on a word-by-word basis, so we test them instead on a class basis. In this respect, the excess of observed over expected co-occurrences again turns out to be highly significant. The overall ratio of observed to expected co-occurrences, 34.5 (74/2.1), is much higher than the estimates of 8 to 11 times the expected number of co-occurrences suggested by the higher-frequency sets of antonym pairs. Furthermore, 48 of these antonyms co-occur in at least one sentence in the corpus. Computation of the probability of observing co-occurrences for at least 48 of the 346 pairs is intractable, so we estimate it using simulation; only 2.0 antonym pairs are expected to co-occur, and in only one of 1,000,000 trials did as many as 11 pairs co-occur. The probability of 48 or more pairs co-occurring by chance is therefore negligible; since the tail of the distribution drops rapidly, that probability is certainly far less than 10^{-6}. In summary, the co-occurrence of morphologically-expressed antonyms as a group is highly significant. Note that the frequencies are so low for most adjectives that there can be no training by textual co-occurrence for most morphological adjective pairs during a human lifetime, so the class phenomenon is acquisitionally pertinent at most to the morphological pattern.

2.5 Co-occurrence Rates

The analysis above shows that antonyms co-occur sententially far more often than expected by chance. It demonstrates this both on a pair-by-pair basis, for adjective pairs whose members are frequent enough, and on a class basis, for each group of antonyms we investigated.

There is substantial variation among antonym pairs, for the Deese group and for the other high-frequency adjectives, in the rates of sentential co-occurrence of antonyms; among pairs for which fairly stable estimates can be made (explicitly, those having 10 or more co-occurrences), co-occurrences take place an average of once every 12.5 sentences, with a range from once per 4.4 to once per 22.8 sentences. Nonetheless, both of the antonym groups involving high-frequency adjectives as well as the (mostly low-frequency) morphological antonyms have quite similar *overall* co-occurrence rates: once per 14.7 sentences for the Deese group, once per 13.3 sentences for the other high-frequency antonyms, and once per 18.2 sentences for the morphological antonyms. These rates do not differ significantly; comparing all three yields a χ^2 of 4.3947 with 2 degrees of freedom ($p \simeq 11\%$).[10] Furthermore, there is some constancy even to the

human-inhuman; terminable-interminable; conscionable-unconscionable; founded-unfounded. Perhaps more debateably excluded were *measurable-immeasurable; calculable-incalculable; credible-incredible; definable-indefinable; subordinate-insubordinate; respective-irrespective; canny-uncanny; clean-unclean; easy-uneasy.*

10 Separating the morphological adjectives from the others as possibly different in kind, the difference between the Deese group and the other antonym pairs involving high-frequency adjectives yields a χ^2 of 0.6959 with one degree of freedom ($p \simeq 40\%$); grouping these two against the morphological antonyms yields a χ^2 of 3.6663 with one degree of freedom ($p \simeq 5.6\%$). This second difference is on the borderline of statistical significance, suggesting that nonmorphological antonym pairs may co-occur at a slightly higher rate than do morphological adjectives.

variations in co-occurrence rate characterizing specific antonym pairs: rates in the APHB Corpus are in good agreement with those in the Brown Corpus, after adjusting for the inclusion of nonadjectival instances of some of the words (e.g., *little* is used adverbially about twice as often as it is used adjectivally) and for a slightly lower overall co-occurrence rate in the larger corpus. Co-occurrence rates therefore seem to be a relevant way of characterizing the co-occurrence phenomenon; they are also intuitively clear, being sample conditional probabilities for the occurrence of an adjective given the occurrence of its less frequent antonym.

In contrast, the three groups show substantially different overall ratios of observed to expected co-occurrences: 8.6 for the Deese group, 23.5 for the other high-frequency antonyms, and 34.5 for morphological antonyms. These differences, given the essentially constant co-occurrence rates, are due to differences in the overall frequencies of adjectives in the three groups. Let N be the total number of sentences in the corpus, $n_{1,i}$ the number containing the more frequent adjective in pair i, $n_{2,i}$ the number containing the less frequent adjective in that pair; and let $f_{m,i} = n_{m,i}/N$ be the corresponding relative frequencies, where $m = 1$ or 2. Then for any given antonym pair i, the rate r_i and the ratio ρ_i are related as $r_i = f_{1,i}\rho_i$. Using this relation, simple algebraic manipulation shows that the overall rate $\bar{r}$ and the overall ratio $\bar{\rho}$ are related as $\bar{r} = \bar{f}_1\bar{\rho}$, where $\bar{f}_1 = \sum_i f_{1,i}(f_{2,i}/\sum_j f_{2,j})$ (the average of the higher antonym frequencies weighted by the lower antonym frequencies). Antonym sets with systematically higher $f_{1,i}$ distributions (which are positively correlated with the $f_{2,i}$ distributions) therefore have lower observed/expected ratios. This again suggests that the relatively constant co-occurrence rates are a more appropriate way to characterize the co-occurrence phenomenon.

3. Syntactic Contexts of Co-occurrences

Charles and Miller (1989) demonstrated that both the sentential and the noun phrase contexts of an adjective differ from those of its antonym in sentences in which only one or the other appears; in such sentences, antonymic adjectives are not readily substitutable for one another. However, Charles and Miller did not address sentences in which antonyms do co-occur. This section examines these sentences for the antonyms discussed in Section 2. We find that, in sentences containing both members of an antonym pair, the antonymic adjectives are usually syntactically paired, and in these cases they are commonly found in conjoined phrases that are identical or nearly identical, word for word, except for the substitution of one antonym for the other:

> *That was one more reason she didn't look forward to Cathy's visit,* ***short*** *or* ***long****;*
>
> *Under normal circumstances, he had a certain bright-eyed all-American-boy charm, with great appeal for* ***young*** *ladies,* ***old*** *ladies, and dogs.*
>
> *There was* ***good*** *fortune and there was* ***bad*** *and Philip Spencer, in handcuffs and ankle irons, knew it to be a truth.*

The Brown Corpus contains 229 sentences with co-occurrences of both members of at least one of the Deese antonym pairs. Six of these sentences have members of two pairs, as in

> *Pre-decoration, low-cost molds, and the freedom to form* ***large*** *and* ***small,*** ***thick*** *and* ***thin*** *materials make plastics tailor-made for the industry.*

and two have two instances of the same pair. This yields 237 sentence/co-occurrence tokens to classify syntactically. A classification of antonym co-occurrences according to surface syntactic similarity is given in Table 3. The sentences in which antonyms co-occur have diverse structures, as well as diverse structural positions for the antonyms. In spite of the diversity, we find a strong trend for the antonyms to occur in syntactically parallel and usually lexically identical structures. Eighteen co-occurrences we rate as "accidental," 219 as involving direct contrast.[11] Excluding the accidental cases, 63% (139/219) of antonym co-occurrences are in lexically identical structures. In 42% (58/139) of these co-occurrences, the antonyms themselves are simply conjoined:[12]

*She felt **cold** and **hot**, sticky and chilly at the same time.*

39% (54/139) of them occur in repeated noun or prepositional phrases, word-for-word identical apart from substitution of the antonyms for each other along with optional deletion or pronominalization of some repeated words:

*...one of **low** anionic binding capacity and one of **high** anionic binding capacity.*

Table 3
Syntactic contexts of antonym co-occurrences in the tagged Brown Corpus:
Nonparenthesized numbers are for sentential co-occurrences; parenthesized numbers are for those that occur in immediately conjoined phrases of the type specified, and in conjoined larger structures containing such phrases, respectively.

syntactic context	sentential co-occurrences			
	Deese	other frequent	morphological	random
adjective conjunction adjective	58(58 + 0)	22(22 + 0)	30(30 + 0)	12(12 + 0)
identical noun phrases	49(16 + 17)	14(3 + 6)	7(3 + 2)	1(1 + 0)
identical prepositional phrases	5(3 + 2)	2(2 + 0)	1(1 + 0)	0(0 + 0)
identical head nouns	18(3 + 10)	8(3 + 0)	8(4 + 1)	1(1 + 0)
identical predicates	9(7 + 1)	0(0 + 0)	1(1 + 0)	2(2 + 0)
other	80(0 + 47)	36(0 + 26)	27(0 + 13)	217(0 + 81)
subtotal	**219(87+77)**	**82(30+32)**	**74(39+16)**	**233(16+81)**
accidental	18	3	0	–
total	**237**	**85**	**74**	**233**

11 "Accidental" is a cover term meant to convey the absence of direct semantic contrast in the uses of the two adjectives. Most of these cases appear simply fortuitous, e.g., *The work uses the **old** eighteenth century tradition of giving the part of a **young** inexperienced youth to a soprano*, and *It is possible that especially **large** anacondas will prove to belong to subspecies limited to a **small** area*. More difficult to classify are five or six examples in which the use of the polar terms does not involve their semantic contrast, though deliberate selection of the adjectives based on their antonymy is plausible or probable, e.g., *That cold, **empty** sky was **full** of fire and light*, and *Its **high** impact strength, even at **low** temperatures, resists chipping, cracking, and crazing, according to DuPont.*

12 Many antonym pairs overwhelmingly favor a particular order of presentation when they co-occur in the special patterns. For example, all 16 instances of *good* co-occurring with *bad*, which all occur in adjective-conjunction-adjective or in conjoined identical short noun phrases, are in the order *good, bad*; similarly, in all 30 cases of immediately conjoined morphological antonyms in our sample, the unmarked-marked order is exceptionless. In general, the more frequent adjective precedes its less frequent antonym and, correlatively, the unmarked precedes the marked where markedness clearly applies. This is consistent with Fenk-Oszlon's (1989) results concerning order and frequency in frozen phrases consisting of two conjoined words: in 84% of pairs, the higher-frequency member precedes the lower.

Together, these two patterns amount to 51% (112/219) of all co-occurrences.

Fully 164 (75%) of these 219 co-occurrences appear in conjoined syntactic structures. The pattern is even more characteristic of the 139 identical repeated phrases, 117 (84%) appearing in conjoined structures. Eighty-seven (63%) of the 139 identical phrases appear in an even more striking pattern, the phrases themselves occurring in immediate conjunction with each other, comprising 40% of the total 219 co-occurrences.

More than half of the remaining 80 co-occurrences are similar to the 139 in lexically identical structures in that they are found in highly parallel and strongly contrastive phrases. For example, 16 of the co-occurrences could have been grouped with the identical structures except that two contrasting nouns also substitute for each other, as in

The old shop adage still holds: a ***good mechanic*** *is usually a* ***bad boss.***

Frequently he must work long hours in the ***hot sun*** *or* ***cold rain.***

The ***pain*** *seems* ***short*** *and the* ***pleasure*** *seems* ***long.***

This special group occurs in conjoined structures at the same rate as the lexically identical group. Even the less tightly parallel examples exhibit quite strong contrastive parallelism, e.g.,

For example, a boy may inherit ***a small jaw*** *from one ancestor and* ***large teeth*** *from another.*

In their search for what turned out to be the ***right breakfast china*** *but the* ***wrong table silver****, they opened every cupboard door...*

There was ***nothing specifically wrong*** *with Edythe, but there was* ***absolutely nothing right*** *about her either.*

We have separately classified the other antonyms involving high frequency adjectives and the morphological antonyms according to these categories (see corresponding columns in Table 3). For the high-frequency group, there is no statistically significant difference in the proportions of cases that occur as an antonym pair joined by conjunction, lexically identical phrases with antonym substitution, or conjoined phrases. For morphological antonyms, antonym pairs joined by conjunction are almost twice as frequent, accounting for 42% (31/74) of the co-occurrences. Controlling for this difference, the proportion of co-occurrences in otherwise identical phrases and the proportion in conjoined phrases is the same as for the other two groups.

Analysis of a sample of random adjective co-occurrences shows that the distribution of antonymous adjectives across sentence structures is atypical of adjective pairs generally; see Table 3. To determine this, we selected 250 sentences at random from among those having at least two adjectives, and randomly selected two adjectives from each of those sentences. Seventeen sentences were removed from this list. One was in fact an antonym pair, not pertinent to examination of nonantonymic usage. In 16 of these 250 sentences, the two adjectives selected were premodifiers within the same noun phrase; usually this structure is semantically incongruous for antonyms, and it was not observed among our antonym co-occurrences, so these 16 were also excluded. This left 233 sentences with random adjective co-occurrences to compare with the antomymic cases. The result: of the 233 adjective pairs selected from these

233 sentences, only 7% (16/233) occurred in lexically identical constituents, only 42% (97/233) occur in structures joined by a conjunction. These percentages are drastically lower than observed for any of the antonym groups.

In summary, we have shown that antonyms co-occur sententially mainly by substituting for one another in otherwise identical or near-identical phrases. Repeated phrases, which are bound to be linked to one another during processing, yield a word-for-word alignment — a pairing of their repeated or anaphorically related words that induces a direct pairing of the substituting antonyms with one another. Most words co-occurring in a sentence are not directly paired by any mechanism highlighting their co-occurrence; the direct pairing of antonyms gives them more salience as potential associates, amplifying the effect of co-occurrence in the formation of a lexical association between the antonyms. Arguably, the mechanism of antonym pairing via word alignment produces an immediate, short-term association that with repeated training may stabilize into a long-term association.

4. Discussion

4.1 A Co-occurrence Theory of Antonym Association

Charles and Miller (1989) demonstrated that antonymic adjectives are generally not substitutable for one another in sentences containing only one member of the pair; and they pointed out that a lexical association between antonyms would result from frequent co-occurrence of antonymic adjectives in the same sentence, according to general association theory. Grammatical class associations that may form on the basis of the usual sentential contexts of adjective occurrences evidently do not account for the formation of the lexical association between antonyms specifically.

We have verified the conjecture that lexical antonyms co-occur sententially far more than would be expected by chance. We have also shown that this is a very general phenomenon: it characterizes all antonym pairs involving a frequent adjective, whether or not they show strongly reciprocal responses in word-association tests. These results support the co-occurrence hypothesis for the formation of the lexical association between nonmorphological antonyms, in that both the existence of the phenomenon and its generality are crucial to this acquisitional hypothesis.

Accordingly, sentences with antonym co-occurrences do seem to be the crucial ones for understanding the formation of antonymic associations. Based on our analysis of these sentences, we are able to elaborate the original co-occurrence hypothesis: we now characterize the co-occurrence phenomenon not simply in terms of its excess over chance expectation, but in terms of regular syntactic patterns that they exhibit; and we have identified a mechanism for association formation, antonym alignment via phrasal substitution.

This is the essence of the theory we propose: co-occurrence takes place via substitution, substitution yields antonym alignment, and alignment leads to association. It is crucially a *co-occurrence* theory — the improbably high rates of co-occurrence of antonyms result in the formation of associations between them; furthermore, co-occurrence takes place mainly by substitution in repeated phrases, and phrasal repetition with a substitution of antonyms evidently occurs mainly when these phrases occur very near one another, particularly in the same sentence. Perhaps less crucially, it is also a *substitution* (not substitutability) theory: phrasal substitution provides a mechanism, antonym alignment, that yields an explicit pairing of the antonyms and enhances the efficacy of training on the association between them.

Finally, these investigations have changed our understanding of what antonymy is. To the semantic criterion for antonymy, opposition in meaning, we now add a lexical

criterion: improbably frequent substitution in nearby, otherwise essentially identical phrases. Together, the semantic and lexical criteria define antonymy.[13]

4.2 Acquisitional Implications of Textual Co-occurrence

Additional results of this research provide further support for the co-occurrence hypothesis, while suggesting three factors connected with associative pairing as having a primary influence on the formation of word associations generally and of antonymous pairing in particular. The suggested factors are the overall rate of occurrence of a word with the stimulus word; the improbability of the extent of pairing under a hypothesis of chance; and the inherent *associability* of the word pairs when they do co-occur, a pattern effect.[14]

1. One implicit assumption of a co-occurrence hypothesis for acquisition of a lexical association is a high enough frequency of co-occurrence to provide adequate training for the association. We find mostly quite high rates of occurrence of adjectives with their antonyms, averaging about once per 15 sentences having the less frequent member of an antonym pair; and overall, about 1 sentence in 150 includes the co-occurrence of an antonym pair from our study.

2. Such high rates should normally suffice to yield associations, but this is not the case when they can occur by chance. For example, most sentences containing a given adjective also contain the word *the*, so that the article occurs with that adjective at a higher rate than does any genuinely associated word, including an antonym. Conversely, lower rates of co-occurrence may be equally effective (provided they achieve some threshold assuring adequate training) if these co-occurrences are substantially more surprising (i.e., improbable under a hypothesis of chance co-occurrence).

3. Finally, some contexts in which words co-occur highlight the fact of their co-occurrence, increasing their inherent potential for forming associations; such contextual enhancement of associability may be required for forming associations between infrequently occurring (or co-occurring) words. We have found that antonyms habitually substitute for one another in otherwise essentially identical phrases within the same sentence, and that this powerfully supports their association via antonym alignment. This is probably the single most important pattern increasing the efficacy of co-occurrence in fostering the development of a lexical association between antonyms, and probably between members of any one grammatical class. Other patterns raise associability as well; adjacency, for example, is especially effective in raising the associability of members of different grammatical classes, such as nouns with their modifying adjectives.

13 Judgments of the antonymy of specific word pairs are sometimes subjective and uncertain; often, responses are that a word pair is somewhat antonymous or that the words are fairly good or not so good as antonyms. The lexical criterion accommodates such uncertainty and gradability, since improbability and co-occurrence rate are graded factors.

14 An unusual number of textual co-occurrences, as measured by their observed to expected ratio, is coming to be widely used by computational linguists as evidence for word associations. For example, Wilks et al. (1989) use this ratio as a criterion for establishing links between words in a semantic network; Church and Hanks (1989) use the logarithm of this ratio as a measure for word association.

Under this formulation of the co-occurrence theory, acquiring the lexical relation of antonymy requires a certain amount of training for the association, and as the frequency of adjectives declines, so must the frequency of training for its associations. On the whole, then, very infrequent training should result in weaker associations; more generally, adjective frequency should correlate with the strength of lexical associations. One consequence of this correlation for the formation of a lexical association between semantically opposed adjectives is that, as the frequency of adjectives declines, so does the proportion of those adjectives that have good antonyms (apart from morphological antonyms, which can be derived by rule with no associative training); we have verified this gradient for adjectives in the Brown Corpus. Another consequence is that semantic (especially synonymic) associations should be stronger relative to the lexical (especially antonymic) associations for lower-frequency adjectives. This may account for Deese's (1962:82) finding that the frequency of an adjective stimulus on a word-association test correlates very strongly with the proportion of adjective responses that are antonymic.

4.3 Word-Association Tests and Textual Co-occurrence

We believe that frequencies of response to a stimulus on word-association tests can be accounted for in large part on the basis of patterns of textual co-occurrence of the associates of the stimulus word. In particular, these frequencies depend not only on the strength of antonym associations but on those of the entire range of words with which the stimulus has formed associations. We illustrate the phenomenon using the antonym pair *good-bad*; they co-occur 16 times in the corpus, which is highly improbable as a chance effect (having probability 5×10^{-12}), and all 16 co-occurrences are in highly associable syntactic patterns. Finally, the co-occurrence rate as defined in Section 2, which is the conditional probability of encountering *good* in a sentence containing *bad*, is quite high (16/125), about once per 8 sentences. All three of the conditions discussed in Section 4.2 are satisfied, promoting the formation of a strong association between these words. No other open-class word has so high a probability of being encountered in a sentence containing *bad*. In contrast, with five times as many sentences containing *good* as *bad*, the conditional probability of encountering *bad* in a sentence containing *good* is only one-fifth as high (16/682). Furthermore, several other open-class words occur more frequently than *bad* in sentences containing *good*. Among adjectives, none has a significantly greater than chance number of co-occurrences, and none occurs in the high associability pattern of substituting adjectives in conjoined and otherwise identical phrases, so no other adjective competes strongly with *bad* as an associate of *good*. Among nouns, however, a few (*man, people, time, day*) co-occur with *good* more often than *bad* does, and do so with greater than chance numbers of co-occurrences. While they cannot substitute phrasally for *good* in these sentences, as adjectives can, all of them often occur immediately following the adjective; we suppose that this close association facilitates association formation, though probably not as dramatically as direct substitution in context. This leaves *bad* as the single strongest associate for *good*, but with more competition from other candidates than *good* had from other associates of *bad*. In Deese's word-association tests, *bad* was indeed a much weaker response (13/100) to *good* than *good* was to *bad* (43/100), though each was the other's most common response. This pattern is typical of Deese's antonyms: it is usual, though not exceptionless, for the more frequent adjective in an antonym pair to exhibit the lower response frequency to its antonym; a substantially higher frequency normally entails a larger number of significant textual associates.

We presume that this competition among a set of unequal lexical associates, and of these associates with semantically but not lexically associated words, can account not only for the high proportion of antonymic responses to high-frequency adjectives,

but also for the low proportion of antonymic responses to low-frequency adjectives. In particular, we have already suggested that nonmorphological antonymic association must be quite weak in the case of very infrequent adjectives, due to insufficient training opportunities; and morphologically negative antonyms of low-frequency adjectives are almost always much lower in frequency than the base form. Assuming that such infrequent adjectives typically have quite limited semantic domains of relevance, they are apt to show repeated co-occurrence with nouns in noun phrases, especially in a particular type of corpus or a particular person's sphere of activity and experience. Accordingly, it will usually happen that a few nouns premodified by a given rare adjective have the highest co-occurrence rate and improbability factor of any open-class word, and in corpora whose size reflects the language learner's early experience with language, the antonym (which can hardly be other than a morphological antonym) will rarely or never have appeared. In contrast, the highest-frequency adjectives have extensive training opportunities, resulting in a great deal of experience with antonym co-occurrence, whose associative strength has the support of phrasal substitution; and competition from nouns and verbs is much more diffuse than in lower-frequency adjectives due to a usually broad range of applicability. Accordingly, the three factors governing associative strength, together with competition among associates of varying strengths, induce a correlation between the frequency of a stimulus adjective and the proportions of adjective vs. noun responses on word-association tests. Deese interpreted this correlation as reflecting a greater effectiveness of paradigmatic (same part of speech) associations at higher frequencies, and of syntagmatic (typical or prototypical accompaniments) at lower frequencies. It now appears that the opposition between paradigmatic and syntagmatic is not directly relevant to the process of association formation. Rather, these are descriptively different effects of competition among all associated words; apart from possible differences in the efficacy of substitution vs. adjacency in raising the associability of co-occurring words, and in the relative strengths of semantic vs. lexical associations in the paradigmatic and syntagmatic cases, the outcome of the competition is based on the same kinds of textual co-occurrence preferences for both paradigmatic and syntagmatic associates.

4.4 Co-occurrence and Substitution

The factors discussed here, and their relations to word-association response frequencies, are quite general and thus likely to apply to other major word classes. In fact, the two types of syntactic patterns characterizing most antonym co-occurrences ought to produce associations of frequent words with other words having the same part of speech, because these contexts of word co-occurrence are specific to words agreeing in part of speech.[15] (1) Syntactically, conjunctions are explicit pairing devices, and the words or constructions they join necessarily agree in grammatical class; the conjoining of two words within a phrase thus provides training for a lexical association between words of a single class. (2) Similarly, when a given environment is literally repeated in a sentence, except that one member of a word pair occurs in one of them and another occurs in the same slot in the other, this usually guarantees grammatical class agreement between the substituting words; and the repetition with word substitution results in direct pairing of the substituting words and provides training for a lexical association between them. Both contexts also substantially raise the associability of the co-occurring words, so the training they receive should be effective.

15 These syntactic contexts may also aid in the acquisition of word morphology, through part-of-speech substitutions that exhibit the morphological variations. We have shown that morphological antonymy as a relation is amply trained via co-occurrence, though most individual pairs receive no training.

We conclude that the acquisition of antonymy in adjectives takes place by a process that may contribute to the acquisition of associations between other words agreeing in part of speech: co-occurrences of words of a single grammatical class in two special types of syntactic context that are characteristic of antonym co-occurrence. Such a unified account for antonymic and part-of-speech associations was the motivation behind the substitutability hypothesis. The syntactic contexts that typify the co-occurrences of antonyms introduce a restricted and especially noticeable form of substitutability into the co-occurrence hypothesis — actual substitution of associated words in a repeated context.

Appendix: Statistical Calculations

We model the co-occurrence problem using the hypergeometric distribution for computation of expected values and probabilities of occurrence of as many co-occurrences as observed. Suppose there are N sentences in the corpus, n_1 of which contain adjective 1 and n_2 of which contain adjective 2, and that k sentences contain both adjectives. To model this situation, we suppose that we have selected at random the n_1 sentences to contain adjective 1 and are now selecting the sentences to contain adjective 2. Since we have k co-occurrences, we must select k sentences from among those containing adjective 1, and the remaining $n_2 - k$ sentences from among those not containing adjective 1. There are

$$\begin{pmatrix} n_1 \\ k \end{pmatrix}$$

ways of choosing k sentences from among n_1, and

$$\begin{pmatrix} N - n_1 \\ n_2 - k \end{pmatrix}$$

ways of choosing $n_2 - k$ sentences from among $N - n_1$; since the choices of sentences from those containing adjective 1 are independent of the choices from those not containing adjective 1, the probability $p(k)$ of choosing at random k sentences to contain both adjective 1 and adjective 2 is

$$\begin{pmatrix} n_1 \\ k \end{pmatrix} \begin{pmatrix} N - n_1 \\ n_2 - k \end{pmatrix} \Big/ \begin{pmatrix} N \\ n_2 \end{pmatrix}.$$

These probabilities form the hypergeometric distribution. The expected value of k is $n_1 n_2 / N$; this formula is used to compute the expected number of co-occurrences in Tables 1 and 2.

The probability $P(k)$ of observing *at least* k co-occurrences is

$$\sum_{i=k}^{\min(n_1,n_2)} p(k) = 1 - \sum_{i=0}^{k-1} p(k).$$

The values in the *probability* column of Tables 1 and 2 are computed using this formula for $P(k)$; this allows us to determine whether the observed number of co-occurrences is unlikely to have been produced by chance.

Using the above formula, the probability that the antonym pair in question has at least one co-occurrence is $P(1) = 1 - p(0)$. Let j refer to the jth antonym pair, and $p_j(k)$ and $P_j(k)$ the probabilities of k co-occurrences and at least k co-occurrences

respectively, and r be the total number of antonym pairs being considered. Now, select s pairs, $\{j_1, \ldots, j_s\}$, out of the total of r pairs to have at least one co-occurrence each, and the remaining $r - s$ pairs $\{j_{s+1}, \ldots, j_r\}$ as having no co-occurrences. Call this particular combination c, an element of the set C_s of combinations of s out of the r pairs. The probability q_c that this particular combination would be found by chance is

$$\left(\prod_{i=1}^{s} p_{j_i}(0)\right)\left(\prod_{i=s+1}^{r} P_{j_i}(1)\right).$$

The overall probability Q_s that exactly s pairs will co-occur by chance is the sum $\sum_{c \in C_s} q_c$; it consists of

$$\binom{r}{s}$$

such terms, one for each combination c in C_s. Finally, the probability that at least t pairs have at least one co-occurrence is $\sum_{s=t}^{r} Q_s$.

This is the formula used to compute the probabilities of getting at least as many pairs co-occurring as we observe, in Tables 1 and 2. However, the number of terms required for computing the probabilities Q_s become astronomical when r is large and s is not close to 0 or to r. This was the situation in our analysis of morphological antonyms ($r = 346$, $s = 48$) in Section 2.4. In such cases, simulation must be used to estimate $\sum_{s=t}^{r} Q_s$. The simulation proceeds as follows. In each trial, r random numbers are generated, uniformly distributed between 0 and 1. If the ith number is greater than $p_i(0)$, then the trial produces a co-occurrence for pair i. We keep track of the number of such co-occurrences generated for each trial. After 1,000,000 trials, we have an approximation to the probability distribution for the number of co-occurrences expected to be produced by chance. If k or more pairs never co-occur in 1,000,000 trials, the probability of so many pairs co-occurring is very likely to be smaller than 10^{-5}; otherwise, the expected number of observations would be at least 10, the probability of having no trial with k or more pairs co-occurring would be less than 0.003.

Acknowledgments

This paper owes its existence to productive interaction with George Miller, who brought the phenomenon of antonym co-occurrence to our attention and provided very useful critical comments on the paper. We have also received helpful comments from Ted Briscoe, Roy Byrd, Martin Chodorow, Judith Klavans, and Yael Ravin. We thank Sue Atkins, Claudia Justeson, and Yael Ravin for evaluating candidate antonyms for high-frequency adjectives.

References

Charles, Walter G. and Miller, George A. (1989). "Contexts of antonymous adjectives." *Applied Psycholinguistics*, 10(3):357–75.

Church, Kenneth W. and Hanks, Patrick (1989). "Word association norms, mutual information, and lexicography." *Proceedings, 27th Annual Meeting of the Association for Computational Linguistics*, Vancouver, 76–83.

Deese, James E. (1962). "Form class and the determinants of association." *Journal of Verbal Learning and Verbal Behavior*, 1(3):79–84.

Deese, James E. (1964). "The associative structure of some common English adjectives." *Journal of Verbal Learning and Verbal Behavior*, 3(5):347–57.

Deese, James E. (1965). *The Structure of Associations in Language and Thought*. The Johns Hopkins Press.

Ervin, Susan (1961). "Changes with age in the verbal determinants of word-association." *American Journal of Psychology* 74:361–72.

Ervin, Susan (1963). "Correlates of associative frequency." *Journal of Verbal Learning and Verbal Behavior*, 1(6):422–31.

Fenk-Oczlon, Gertraud (1989). "Word frequency and word order in freezes." *Linguistics*, 27(3):517–56.

Francis, Winthrop N. and Kučera, Henry (1982). *Frequency Analysis of English Usage: Lexicon and Grammar*. Houghton Mifflin.

Gross, Derek; Fischer, Ute; and Miller, George A. (1989). "Antonymy and the representation of adjectival meanings." *Journal of Memory and Language*, 28(1):92–106.

Shanks, David R. and Dickinson, A. (1987). "Associative accounts of causality judgment." In *The Psychology of Learning and Motivation*, edited by Gordon H. Bower, 21:229–61, Academic Press.

Thorndike, Edward L. and Lorge, Irving (1944). *The Teacher's Word-Book of 30,000 Words*. Teacher's College, Columbia University.

Wilks, Yorick; Fass, Dan; Guo, Cheng-Ming; McDonald, James E.; Plate, Tony; and Slater, Brian M. (1989). "Machine tractable dictionaries as tools and resources for natural language processing." In *Computational Lexicography for Natural Language Processing*, edited by Branimir Boguraev and Edward J. Briscoe, 193–228, Longman.