Brigham Young University
BYU ScholarsArchive

Faculty Publications

2006

English adjective comparison and analogy

Dirk Elzinga
Brigham Young University, dirk_elzinga@byu.edu

Follow this and additional works at: https://scholarsarchive.byu.edu/facpub

Part of the Linguistics Commons

Original Publication Citation

2006."English Adjective Comparison and Analogy." Lingua. 116.6, pp 757-770.

BYU ScholarsArchive Citation

Elzinga, Dirk, "English adjective comparison and analogy" (2006). *Faculty Publications*. 6565.
https://scholarsarchive.byu.edu/facpub/6565

This Peer-Reviewed Article is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Faculty Publications by an authorized administrator of BYU ScholarsArchive. For more information, please contact ellen_amatangelo@byu.edu.

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Lingua 116 (2006) 757–770

Lingua

www.elsevier.com/locate/lingua

English adjective comparison and analogy

Dirk Elzinga *

Department of Linguistics and English Language, Brigham Young University, 4043 JFSB, Provo, UT 84602, United States

Received 9 August 2004; received in revised form 15 February 2005; accepted 9 March 2005
Available online 19 April 2005

Abstract

There are two strategies for forming the comparative degree of adjectives in English; a synthetic strategy which suffixes *-er* to the adjective stem, and an analytical strategy which uses *more* in composition with the adjective. Many analyses of the choice between analytical and synthetic comparison have been proposed, but all face difficulties. In this paper I show that analogy can not only account for the distribution of analytical and synthetic comparison as well as traditional rule-based approaches, but can also provide a psychologically plausible model for the choice which speakers make.
© 2005 Elsevier B.V. All rights reserved.

Keywords: Analogy; Instance-based; English adjective inflection

1. Introduction

There are two strategies for forming the comparative degree of adjectives in English. The *synthetic* strategy suffixes *-er* to the adjective stem, while the *analytical* strategy uses *more* in composition with the adjective (1).

(1) English adjective comparison
  a. synthetic comparison

great	greater
large	larger
easy	easier

* Tel.: +1 801 422 2117; fax: +1 801 422 0474.
E-mail address: Dirk_Elzinga@byu.edu.

0024-3841/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.lingua.2005.03.003

b. analytical comparison

complex	more complex
serious	more serious
comfortable	more comfortable

While there have been many analysis of the variation between analytical and synthetic comparative constructions, it has defied an account of sufficient predictive power to count as an explanation. In this paper I propose that *analogy* is the guiding principle which English speakers use to form the comparative degree of adjectives. I show that an explicit and rigorous formulation of analogical principles can not only account for the distribution of analytical and synthetic comparison as well as traditional rule-based approaches, but can also provide a psychologically plausible model for the choice which speakers make.

The rest of this paper is organized as follows. In Section 2 I discuss the range of data which an adequate account of adjective comparison must deal with and point out the difficulties which confront traditional, rule-based analysis. In Section 3 I give a brief overview of Analogical Modeling (Skousen, 1989, 1992, 1995; Skousen et al., 2002), an explicit, computational implementation of analogical principles. Section 4 shows how this model is applied to the problem of adjective comparison in English. I show that Analogical Modeling provides an excellent fit to usage information gleaned from the World-Wide Web, lending credence to the idea that analogy can be an explanatory principle in linguistic structure beyond merely accounting for exceptions. Section 5 describes an adjective comparison survey which was conducted among English speakers. Analogical Modeling is shown to fit the range of responses very well, demonstrating that it is psychologically plausible as well as being descriptively adequate. Section 6 provides some concluding remarks.

2. Adjective comparison in Modern English

Standard accounts of the choice between analytical and synthetic comparatives in English rely largely, though not exclusively, on prosodic factors. Monosyllabic adjectives show synthetic comparative marking (Quirk et al., 1985: 461; Huddleston and Pullum, 2002: 1583, 1584) (2).

(2) Monosyllabic adjectives

hard	harder
green	greener

Trisyllabic adjectives show analytical comparative marking (Quirk et al., 1985: 462; Huddleston and Pullum 2002: 1584) (3).

(3) Trisyllabic adjectives

obvious	more obvious
accurate	more accurate

Disyllabic adjectives are split according to their stress pattern and phonological termination: initially stressed disyllabic adjectives show synthetic comparative marking, especially when they end in *-y, -ow,* and *-le* (Quirk et al., 1985: 462; McCarthy and Prince, 1986: 2; Biber et al., 1999: 522, 523; Huddleston and Pullum, 2002: 1583) (4a), while disyllables with final stress show analytical comparative marking (4b).

(4) Disyllabic adjectives

a. initial stress

angry	angrier
mellow	mellower
nimble	nimbler

b. final stress

serene	more serene (*serener)
diffuse	more diffuse (*diffuser)
abrupt	more abrupt (*abrupter)

There are frequent exceptions to these patterns. For example, participles are not generally eligible for synthetic comparative marking (Quirk et al., 1985: 462; Huddleston and Pullum, 2002: 1583), though *tired* is an exception for many speakers (5a); adjectives formed with suffixes *-al*, *-ful*, *-ic*, *-ish*, and *-ous* also seem to resist synthetic comparative marking (Huddleston and Pullum, 2002: 1583) (5b); adjectives formed by prefixing *un-* to a disyllabic stem retain synthetic comparative marking even when they are trisyllabic (Quirk et al., 1985: 462) (5c); and some adjectives are simply exceptional (Quirk et al., 1985: 461; Huddleston and Pullum, 2002: 1583) (5d).

(5) Non-conforming adjectives

a. participles

bored	more bored (*boreder)
pleased	more pleased (*pleaseder)
worn	more worn (*worner)
tired	more tired (?tireder)

b. suffixed adjectives

moral	more moral (*moraler)
venal	more venal (*venaler)
careful	more careful (*carefuler)
useful	more useful (*usefuler)
magic	more magic (*magicker)
tragic	more tragic (*tragicker)
boyish	more boyish (*boyisher)
foolish	more foolish (*foolisher)
famous	more famous (*famouser)
porous	more porous (*porouser)

c. *un-* adjectives

unhappy	unhappier, more unhappy
untidy	untidier, more untidy

d. lexical exceptions

cross	more cross (?crosser)
fake	more fake (?faker)
like	more like (*liker)
real	more real (?realer)

Weakening these generalizations considerably are statements such as ‘‘There is no simple set of rules to indicate which adjectives take which type: in many cases it is a matter of more or less likely rather than possible or impossible.’’ (Huddleston and Pullum, 2002: 1582) and ‘‘Many disyllabic adjectives can also take inflections, though they have the alternative of the periphrastic form’’ (Quirk et al., 1985: 462). In the first case, it is conceded that the generalizations are no more than tendencies; in the second, no prediction is possible since both outcomes are available (at least to a certain group of adjectives). This presents a real challenge to the problem of explaining English adjective comparison, yet English speakers cope with this state of affairs with ease. In the remainder of this paper I will outline an approach to English adjective comparison which avoids rules or similar predictive statements, but which can nevertheless predict with a great deal of accuracy whether a given adjective occurs in an analytical comparative construction or as a synthetic comparative adjective.

3. Analogical Modeling and instance-based approaches

The traditional approach to describing and predicting linguistic behavior has consisted of formulating rules which capture generalizations over a range of data (this includes Optimality Theory-style constraints). It is generally recognized, however, that such rule-based accounts can only provide an approximation to actual linguistic behavior; problems such as the variation seen in English adjective comparison are accounted for only with great difficulty in rule-based approaches.

Instance-based approaches represent a response to the shortcomings of rule systems. Rather than extracting generalizations from data in the form of rules which are then applied to novel forms, instance-based approaches use a database of stored examples which serve as possible models to predict the shape or outcome of a given input. In *nearest neighbor* approaches such as TiMBL (Daelemans et al., 2004), this involves finding the examples which most resemble the input — the nearest neighbors — and then applying the behavior of the nearest neighbors to the input.

Analogical Modeling (Skousen, 1989, 1992, 1995; Skousen et al., 2002) is another instance-based approach to the problem of predicting linguistic behavior. As in all instance-based approaches, items in an AM database are represented as values of a series of features or variables. In determining the set of possible analogical models, AM sifts through the database looking for items similar to the given form; this search is widened by successively disregarding variables until the database has been completely searched. In the course of this search, database items are grouped into sets (called supracontexts) whose members share feature values with the given form. For example, if the given form was represented by three variables whose values were *w e t*, AM would collect database items which share values for these three variables. It would then ignore each variable in succession, collecting forms which share with the given form the features *w e -*, *w - t*, *- e t*, *w - -*, *- e -*, *- - t*, and - - -, until all forms in the database had been examined and grouped into supracontexts.

Examples in the supracontexts are potential analogical models. In this respect, AM differs from nearest neighbor approaches since the models are not simply the *x* number of nearest neighbors, but may extend to more distantly related items. The probability that an example will be used as an analogical model for a given form depends on three properties: (i) *proximity*, (ii) *gang effect*, and (iii) *heterogeneity*.

Proximity

Database items that share more features with the given form will appear in more supracontexts and will therefore have a greater chance of being used as an analogical model.

Gang Effect

When a group of similar examples behave alike, then the probability of selecting one of these examples as an analogical model is substantially increased.

Heterogeneity

An example cannot be selected as the analogical model if there are other examples more similar to the given form which have different behavior.

The set of possible analogical models thus determined is referred to as the *analogical set*.

The outcome for a given form can be predicted in one of two ways: (i) choose an example at random from the analogical set and adopt its behavior, and (ii) adopt the behavior which is favored by a majority of the items in the analogical set (Skousen, 1989: 82). There is good evidence that people use both strategies in selecting outcomes for novel forms (Messick and Solley, 1957); this study will use selection by majority.

4. Applying AM to English adjective comparison

In this section I show how AM can be applied to the problem of adjective comparison in English. I describe the construction of the data set, the coding of the examples, and the results of the AM simulations. I show that there is an excellent fit between the results of the AM simulations and the range of actual occurrences of comparative constructions as found on the World-Wide Web. I show that traditional accounts of adjective comparison are refined by application of analogical methods to the problem.

4.1. Constructing the data set

The data set for the AM simulations of adjective comparison consists of adjectives which were found to occur in comparative constructions in the Brown Corpus (Francis and Kučera, 1964). In constructing the data set I excluded irregular or suppletive comparative forms such as *farther*, *further*, *better*, and *worse*; I also excluded hyphenated forms such as *down-to-earth*, *higher-priced*, *lower-paid*, and *lobster-backed*. Orthographically irregular forms such as *bigga* (*bigger*), were not included as separate data items themselves. The resulting list of adjectives contained 485 forms.

In AM, each example in a data set is represented as specific values for a series of variables; the nature of these variables will differ according to the data under consideration, though most AM studies to date have represented data sets with phonological and morphological variables. In this study, 20 variables were used for each of the adjectives in the list. These variables were assigned values as shown in (6).

(6) Comparative adjective variables

1	the number of syllables
2–4	the stress pattern of the final three syllables, where 1 = primary stress, 2 = secondary stress, and 3 = stressless
5–6	the segments immediately preceding the antepenultimate vowel
7	the antepenultimate vowel; if a word is disyllabic, this variable is ‘0’
8–9	the segments immediately following the antepenultimate vowel
10–11	the segments immediately preceding the penultimate vowel
12	the penultimate vowel; if a word is monosyllabic, this variable is ‘0’
13–14	the segments immediately following the penultimate vowel
15–16	the segments immediately preceding the final vowel
17	the final vowel
18–19	the segments immediately following the final vowel
20	the final segment of the word.

An ASCII transcription scheme was used to represent phonetic segments in variables. Variables whose zero value is unpredictable are filled by ‘0’; variables whose zero value is predictable are filled by ‘=’, the AM non-specification marker. For example, the fact that a word has a single syllable is a property of that word and is not predictable, so the variable representing the vowel of the penultimate syllable will therefore be filled with a ‘0’, given the right-to-left orientation of data items. The fact that the variables representing segments flanking the penultimate vowel will be zeros is predictable from the fact that there is no penultimate vowel, and so those variables are filled with ‘=’s. Also, the fact that a monosyllabic word will not have an antepenultimate syllable is predictable from the fact that it does not have a penultimate syllable. So while the penultimate vowel variable is filled with a ‘0’, the antepenultimate vowel variable is filled with a ‘=’. The variables filled by ‘=’ can be ignored by the AM algorithm (as they are in this study).

I consider two possible outcomes for adjectives in comparative constructions: analytical and synthetic; these outcomes are represented in the data set by ‘a’ and ‘s’, respectively.[1] To determine the outcomes for items in the list, I submitted an analytical comparative construction and a synthetic comparative form for each adjective to the internet search engine Google to obtain occurrence frequencies from the World-Wide Web; these searches were limited to the text of web pages in English. The construction type (analytical or synthetic) which received a majority of hits determined the outcome for a given adjective. For example, the analytical comparative construction *more common* had 1,020,000 hits while the comparative adjective *commoner* had 48,400 hits. Therefore, an analytical outcome was assigned to the adjective *common*.[2]

Examples of variable assignment are given in Table 1 for the adjectives *weak, heavy, distinct, efficient,* and *fundamental*.

Table 1
Examples of variable assignment with outcomes

Out	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Adjective
s	1	=	0	1	=	=	=	=	=	=	=	0	=	=	0	w	i	y	k	k	weak
s	2	0	1	3	=	=	0	=	=	0	h	ɛ	v	i	ɛ	v	i	y	0	y	heavy
a	2	0	3	1	=	=	0	=	=	0	d	ɪ	s	t	s	t	ɪ	ŋ	t	t	distinct
a	3	3	1	3	0	0	ɪ	f	ɪ	ɪ	f	ɪ	š	ə	ɪ	š	ə	n	t	t	efficient
a	4	3	1	3	n	d	ə	m	ɛ	ə	m	ɛ	n	t	n	t	ə	l	0	l	fundamental

4.2. Results of the AM simulations

For the first simulation, I conducted a “leave-one-out” experiment. In an experiment of this type, each item in the data set is momentarily excluded from the data set and used as a test item. The benefit of a “leave-one-out” experiment is that it can determine the degree of *regularity* of the data set. That is, if a given context has the same outcome as its analogical set, it is *regular* (see Skousen, 1989: 66–67 for discussion of this notion). For the current study, we can consider the degree of regularity to be a rough measure of the accuracy of the model in predicting the given outcomes. In this case, AM predicted the correct outcome 92.6% of the time; adjectives with the analytical outcome were correctly predicted at a rate of 94.5%, while adjectives with the synthetic outcome were correctly predicted at a rate of 86.9%.

4.3. Comparing AM with traditional accounts of adjective comparison

Previous accounts of the choice between analytical and synthetic strategies for adjective comparison have relied largely on prosodic and morphological criteria; this is true no less

[1] In this study, I do not treat doubly marked comparatives like *more quicker*. See Kytö and Romaine (1997) for discussion of the history of this construction.

[2] The WWW data used in this study can be obtained at http://humanities.byu.edu/am/ or by contacting the author directly.

for work in the descriptive tradition as for more recent accounts within a generative framework (Browne, n.d., Booij and Rubach, 1984, McCarthy and Prince, 1986). All of these previous descriptions seem to agree on the following points (7, 8):

(7) Phonological predictors
   a. An adjective will have a synthetic comparative form if it is monosyllabic.
   b. An adjective will have a synthetic comparative form if it is disyllabic, and:
      i. it is stressed on the first syllable, and
      ii. it ends in unstressed *y*, *le*, or *ow*.
   c. Otherwise, an adjective will participate in an analytical comparative construction.

(8) Morphological predictors
   An adjective will participate in an analytical comparative construction if:
   a. it is a participle, or
   b. it ends in one of the suffixes *-al, -ful, -ic, -ish,* or *-ous*.

In informal tabulations, the phonological and morphological criteria given in (7) and (8) can correctly account for 93.6% of the comparative constructions in the data set. Given that AM has about the same degree of accuracy as a traditional rule-based approach, what insight does an analogical approach shed on the importance of the phonological and morphological criteria of (7) and (8)? In this section I show that AM confirms the importance of a mix of prosodic and segmental factors, but that the morphological criteria of (8) play little, if any, role in the selection of the correct comparative construction for a given adjective. I also show that AM is much more robust, since changing the variables which are considered in the simulations did not have significant effects on the outcome.

4.3.1. Prosody and segmental information

The restriction of synthetic comparative forms to initially stressed mono- or disyllabic adjectives suggests that the prosodic variables alone from (6) may suffice in predicting the choice between synthetic and analytical comparison. When running an AM simulation using only these variables (variables 1–4 (6)), the rate of correct predictions was 74.8%. However, this was achieved by predicting the analytical outcome for all forms; none of the forms with the synthetic outcome were predicted to have a synthetic outcome. Essentially, all this revealed was that about three quarters of the data set is comprised of analytical comparative constructions. It seems clear from this result that the choice between synthetic and analytical is not determined by the prosody alone; both AM and traditional accounts agree on this point.

Adding the variable encoding the final segment (variable 20 (6)) to the prosodic variables improved the rate of prediction markedly. The overall rate of correct predictions rose to 90.5%, with 96.7% correctly predicted for the analytical outcome and 72.1% correctly predicted for the synthetic outcome. Adding the variables for the final syllable (variables 15–19) to the prosodic variables and the final segment variable brought the overall rate of correct predictions up to 89.9%—the same as for the data set with all variables included; the analytical predictions fell to 93.7%, but the synthetic predictions rose to 78.7%. Adding variables for successive syllables had the greatest effect on the prediction rate for synthetic outcomes. Removing prosodic variables resulted in a slight degradation of performance; the rate of overall correct predictions when all segmental variables were included but prosodic variables were excluded was 91.1%, with 94.8% correctly predicted for the analytical outcome and 80.3% correctly predicted for the synthetic outcome. Table 2 summarizes the results.

Table 2
Results of AM simulations with segmental and prosodic variables

	Prosody alone	Segments alone	Prosody + final segment	Prosody + final syllable	Prosody + final 2 syllables	All variables	Rules
Analytical	100.0	94.8	96.7	93.7	94.8	94.5	92.8
Synthetic	0.0	80.3	72.1	78.7	86.1	86.9	95.9
Overall	74.8	91.1	90.5	89.9	92.6	92.6	93.6

From this series of simulations using AM it can be seen that the variables with the greatest effect on the outcome are the variables encoding the prosody and the segmental content of the final two syllables of an adjective; the additional variables for the antepenultimate syllable have a slight effect for synthetic outcomes. This confirms traditional descriptions of adjective comparison which maintain that prosody and the phonological termination of the adjective are important factors in the choice between analytical and synthetic comparative strategies.

4.3.2. *Morphological effects*

Thus far, AM has been able to reproduce the results of previous accounts of adjective comparison using only phonological variables. However, recall that those descriptions also rely on morphological criteria: the exemption of participles and adjectives which end in suffixes like *-al*, *-ic*, *-ish*, and *-ous* from synthetic comparative marking, even when phonologically suitable (9). To test whether morphological structure has an effect on AM's predictions, I took the data set including the prosodic variables, the final segment, and the variables for the final two syllables and added two additional variables: (i) whether the adjective was morphologically simple (S) or complex (C), and (ii) the final suffix, if any. Alternate spellings of suffixes were normalized; for example, *-ent* and *-ant* appear in the data set as *-ant*, and both *-ible* and *-able* appear as *-able*. Adding the suffix itself as a variable is important, since the fact that there is a suffix present is not recoverable from the segmental variables themselves. For example, the adjectives *able* and *verbal* both end in [əl] but only in the second case is it a suffix. Table 3 provides several examples.

Table 3
Examples of data items including morphological variables

Out	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Adjective
s	1	=	0	1	=	=	0	=	=	0	d	a	r	k	k	S	=	dark
a	2	0	3	1	0	d	ə	r	ɛ	ə	r	ɛ	k	t	t	S	=	direct
s	2	0	2	1	0	0	ʌ	n	r	n	r	i	y	l	l	C	=	unreal
a	3	3	1	3	m	p	ɔ	r	t	r	t	ə	n	t	t	C	ant	important
a	4	1	3	3	p	t	ə	b	ə	ə	b	ə	l	0	l	C	able	acceptable

Running the simulation with morphological variables included did not improve overall performance as might be expected if the morphology were playing a role in the selection of the outcome. The percentage of overall correct predictions was 92.8%; the percentage of correct analytical predictions was 95.0%, and the percentage of correct synthetic predictions was 86.1%.

In this encoding of the data set there are 15 phonological variable positions but only 2 for morphology; it seems likely that the reason morphology is not seen as having any effect in this simulation is that the phonological variables are overwhelming the morphological variables. The question is actually a more general one in AM: Should variables of different types (e.g., phonological, morphological, syntactic, semantic, etc.) be weighted equally in variable vectors? It seems unlikely that they should be, but at present there is no principled way to compare the weight of variables of different types (see Skousen, 2002: 39–42 for discussion). Eddington (2002) resorts to the expedient of repeating a morphological variable a number of times to ascertain whether it has any effects in the predictions made by AM. In his investigation of Spanish stress assignment (Eddington, 2002: 148–151) he shows that including a variable indicating verb tense and repeating it three times reduced the number of stress assignment errors which occurred on preterit verbs by nearly one half (from 50 errors to 27). This demonstrates that variable weighting by repetition can work for AM. For the present study, I conducted simulations which repeated the final variable (the suffix) twice, three times and four times. The repetition of the suffix variable had slight, but insignificant effects on the rate of correct predictions; Table 4 shows the results.

Table 4
Results of AM simulations with morphological variables

	Morphological variables	Suffix repeated 2×	Suffix repeated 3×	Suffix repeated 4×
Analytical	95.0	95.0	94.8	94.8
Synthetic	86.1	86.9	86.9	86.9
Overall	92.8	93.0	92.8	92.8

Although the overall percentage of correct predictions did not improve significantly, it might be expected that adding morphological variables nevertheless had effects for individual lexical items, as in Eddington's study. However, the set of wrong predictions made by AM under both phonological and mixed phono-/morphological encodings overlapped considerably; no trends could be found which might indicate that morphological variables played a role in the selection of the outcome of any given form. What this means is that the terminations which control the choice of comparative constructions are handled by the phonological representations and not by (knowledge of) specific suffixes; from this I conclude that morphological structure is not a significant factor and that the morphological predictors listed in (9) do not play a role in determining the outcome for morphologically complex adjectives.

4.4. Summary

In this section I have shown that predictions made by AM concerning the choice between analytical and synthetic comparatives are a good fit with the range of attested data from the Brown Corpus and Google. Phonological variables, including the number of syllables, stress pattern, and segmental makeup of the final two syllables were shown to be significant in predicting the correct comparative outcome. On the other hand, morphological variables were seen to have no effect on the predictions made by AM.

5. The survey

In the preceding section I showed that AM can predict with reasonable accuracy whether an adjective will take an analytical or synthetic comparative. A more interesting question is whether these results are psychologically valid; that is, whether AM can be taken as a model of how people make the choice between analytical and synthetic comparative constructions for a given adjective. To answer this question, I conducted a survey of English speakers and compared their responses to the predictions made in an AM test of the same survey items.

5.1. Stimulus materials

I collected a list of adjectives which appeared in comparative constructions in the British National Corpus (Fletcher, 2004) but which did not appear in comparative constructions in the Brown Corpus. Irregular, suppletive and hyphenated forms were excluded, as in the ‘‘leave-one-out’’ simulations described above. I used the 300 most frequent adjectives from the resulting list as survey items.

5.2. Subjects

Ninrty-seven undergraduate students participated in the survey, 47 men and 50 women. All were native speakers of American English.

5.3. Procedure

The 300 test items were presented to the survey participants in the form of written questionnaires. On each line were printed the adjective with the degree word ‘more’ preceding and the suffix ‘-er’ following. Participants were asked to circle the comparative construction which sounded best. After conducting the survey, two items were excluded from the list—*best* and *worse*. The percentages of analytical and synthetic responses for each of the remaining 298 forms were then tabulated. Overall, 239 adjectives (80.2%) had a majority of responses preferring analytical constructions, while 59 adjectives (19.8%) had a majority of responses preferring synthetic forms.

5.4. AM comparison

I took the same 298 survey items as the test set for an AM simulation using the encoding described in (6); the data set was the same as the one used in the ‘‘leave-one-out’’ experiments described in Section 4. Since the adjectives in the survey did not occur in the Brown Corpus, they also did not appear in the data set. Outcomes were assigned to the test set based on the survey results; if a form received more than 50% analytical responses in the survey, it was assigned an analytical outcome; otherwise, it was assigned a synthetic outcome. This was done to calculate overall percentages and compare the AM simulation with the survey results directly.

5.5. *Results*

When running the AM simulation, overall correct prediction occurred 87.9% (262/298) of the time. The percentage of correct analytical predictions was 88.3%; the percentage of correct synthetic predictions was 86.4%. This compares favorably with the results of the simulations described in Section 4. To compare the AM result with the survey findings directly, a Pearson Product Moment Correlation was performed with highly significant results ($r(296) = 0.726$, $p < 0.0005$, two-tailed). This suggests that the results of the leave-one-out study are valid not only within their computational domain, but that speakers may use some form of analogy in forming comparatives.

5.6. *Discussion*

The degree of overlap between the survey results and the AM simulation is impressive, but it is instructive to examine the mismatches. There were 36 forms of the 298 whose AM predictions differed from the survey results. These can be divided into two groups: the first group contains those forms which AM predicted as synthetic but which the survey showed as analytical; the second group contain those forms which AM predicted as analytical but which the survey showed as synthetic.

5.6.1. *False analytical predictions*

AM predicted eight synthetic forms as analytical; four of these forms have the phonological termination *le*, which is conventionally associated with synthetic comparatives: *little*, *noble*, *stable*, and *subtle*. In traditional descriptions of English adjective comparison, the presence of the suffix *-al* is taken to be a strong predictor for an analytical comparative construction. However, since the variables used to compare AM predictions with the survey results did not include morphological criteria, the phonological variables which capture the orthographic *le* termination are indistinguishable from the phonological variables which encode the suffix *-al*, both being pronounced [əl]. It should thus come as no surprise that AM confounds the two types of adjectives and fails to make accurate predictions for some of the *le* forms. However, including morphological variables did not substantially improve matters. The only change was to the adjective *subtle*, which was correctly predicted when morphological variables were included; the others were still incorrectly assigned analytical outcomes.

Of the other adjectives, *remote* and *common* had survey scores of 42.4% and 41.9%, respectively (100% indicates a unanimous analytical preference); this indicates a great deal of variability on the part of the survey participants. So the fact that AM incorrectly predicts these adjectives is also not surprising.

5.6.2. False synthetic predictions

Twenty-eight of the 36 incorrectly predicted adjectives were assigned synthetic outcomes by AM, while the survey results indicated a preference for analytical outcomes. Of these adjectives, none are longer than two syllables, which, under traditional accounts of adjective comparison, is one of the properties of adjectives which prefer a synthetic comparative. That AM predicts a synthetic outcome for them is thus not surprising.

Five of the incorrectly predicted forms are participles, which is a predictor for an analytical comparative under traditional accounts of adjective comparison (see (9)). Since morphological variables were not included in the simulation, AM's prediction of a synthetic outcome is also not surprising given their prosodic shape. Including morphological variables improved matters for the participles, though; three of the five participles were assigned the expected analytical outcome when morphological variables were taken into account.

5.7. Summary

The results of the survey confirm AM's suitability as a model for predicting the choice between analytical and synthetic comparatives. Not only was there impressive overlap between the survey results and the AM simulation, but the kinds of errors made by AM were found to be perfectly natural given the way the variables were set up. Including morphological variables had a significant effect only for participles.

6. Conclusion

AM can predict with an impressive degree of accuracy the distribution of comparative adjective construction types. In these simulations it was found that information about the prosody and segmental content of the final two syllables of an adjective was crucial for determining the outcome; this accords with traditional accounts of adjective comparison. However, unlike traditional accounts, morphological variables were found to be significant only for participles; otherwise, morphology played little role in predicting the correct comparative outcome. AM also closely matched survey results of adjective comparison. Therefore, speakers of English appear to use the same phonological information in making the choice between analytical comparative constructions and synthetic comparative adjectives.

References

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., 1999. The Longman Grammar of Spoken and Written English. Longman, New York.

Booij, G., Rubach, J., 1984. Morphological and prosodic domains in Lexical phonology. Phonology Yearbook 1, 1–27.

Browne, W., n.d. On adjectival comparison and reduplication in English. Unpublished manuscript.

Daelemans, W., Zavrel, J., van der Sloot, K., van den Bosch, A., 2004. TiMBL: Tilburg Memory Based Learner, Version 5.1, Reference Guide. ILK Technical Report 04-02. Induction of Linguistic Knowledge Research Group, Tilburg.

Eddington, D., 2002. A comparison of two analogical models: Tilburg memory-based learner versus analogical modeling. In: Skousen, R., Lonsdale, D., Parkinson, D.B. (Eds.), Analogical Modeling: An Exemplar-Based Approach to Language. John Benjamins Publishing Company, Amsterdam.

Fletcher, W.H., 2004. Phrases in Englishhttp://pie.usna.edu.

Francis, W.N., Kučera, H., 1964. A Standard Corpus of Present-Day Edited American English, for Use with Digital Computers. Department of Linguistics, Brown University, Providence, Rhode Island.

Huddleston, R., Pullum, G., 2002. The Cambridge Grammar of the English Language. Cambridge University Press, New York.

Kytö, M., Romaine, S., 1997. Competing forms of adjective comparison in modern English: what could be more quicker and easier and more effective? In: Terttu Nevalainen, Leena Kahlas-Tarkka (Eds.), To Explain the Present. Studies in the Changing English Language in Honour of Matti Rissanen (Mémoires de la Société Néophilologique 52). Société Néophilologique, Helsinki, pp. 329–352.

McCarthy, J., Prince, A., 1986. Prosodic Morphologyhttp://roa.rutgers.edu.

Messick, S.J., Solley, C.M., 1957. Probability learning in children: some exploratory studies. Journal of Genetic Psychology 90, 23–32.

Quirk, R., Greenbaum, S., Leech, G., Svartvik, J., 1985. A Comprehensive Grammar of the English Language. Longman, New York.

Skousen, R., 1989. Analogical Modeling of Language. Kluwer Academic Publishers, Dordrecht.

Skousen, R., 1992. Analogy and Structure. Kluwer Academic Publishers, Dordrecht.

Skousen, R., 1995. Analogy: a non-rule alternative to neural networks. Rivista di Linguistica 7 (2), 213–231.

Skousen, R., 2002. Issues in analogical modeling. In: Skousen, R., Lonsdale, D., Parkinson, D.B. (Eds.), Analogical Modeling: An Exemplar-Based Approach to Language. John Benjamins Publishing Company, Amsterdam.