Chapter 40

Variation: from dialect to pragmatics, a progress report

Annie Zaenen
Stanford University

Brian Hicks
Stanford University

Cleo Condoravdi
Stanford University

Lauri Karttunen
Stanford University

Stanley Peters
Stanford University

The paper reports on experimental studies investigating the way native speakers of English use the construction *NP be ADJ to VP* with evaluative adjectives. We show that, contrary to established linguistic theory, this construction is not always interpreted as factive. We isolated one major context in which an important minority and, for some adjectives, a majority of native speakers prefers an implicative use. We investigate whether there might be a dialect split between factive and implicative users and conclude that, contrary to what we suggested in a previous paper, this is not the case. We discuss different subclasses and end with the tentative hypothesis that the variability that is found among speakers has to do with the difference in importance that different language users attach to the non-linguistic context against which they interpret utterances.

1 Introduction

When we think about language variation we tend to think about dialect variation and about variation that is conditioned by sociological variables such as gender, age and

social status. But recent technological advances allow us to collect data from many more speakers than before and studies involving more subjects show unexpected areas of variation that do not fall within these traditional categories. Here we report on one such case: variation in interpretation of the infinitival complements of evaluative adjectives illustrated in (1).

(1) John was smart to go to Groningen.

These adjectives are supposed to be factive, hence their infinitival complement is presupposed to be true even when the matrix clause is negative. Some speakers, however, will, in some circumstances, interpret the event referred to in the infinitival clause as not having happened when the matrix clause is in the negative. In other cases they interpreted it as having happened in accordance with linguistic theory. What are the reasons for this difference in interpretation? Is it a dialect difference or is something else going on?

In this short paper we give a characterization of the adjective class that exhibits the unexpected behavior and summarize the first experiment that led us to the hypothesis of two different dialects. We then describe some further experiments that show that the variation is most likely of a different nature. In conclusion, we formulate a new tentative hypothesis to account for our findings.

2 Two types of NP *be* ADJ *to* VP adjectives

The syntactic frame *NP be ADJ to VP* can be used with adjectives that semantically belong to different classes. Here we focus on the subclasses that have described as factive (see Norrick (1978), for more references see Karttunen et al. (2014)). As first reported in Karttunen & Zaenen (2013), we found that emotive and evaluative adjectives behave very differently under negation. Emotive adjectives such as *outraged, dumbfounded, ecstatic, furious, ...* behave indeed like factives, evaluative adjectives, such as *stupid, smart, lucky, mean, nice, brave, ...* exhibit more complex behavior. Our initial findings, which were the by-product of a study on *lucky*, are given in percentages in Table 1.

Table 1: Emotive and evaluative adjectives.

adjective	matrix polarity					
	affirmative			negative		
	Yes	No	undecided	Yes	No	undecided
emotive	100	0	0	95.7	4.3	0
evaluative	98.9	0.9	0.3	25	64.2	10.7

The subjects were presented with sentences such as (1), with either a positive or a negative matrix clause and had to decide whether the eventuality described in the in-

finitival clause had happened or not. Some speakers/hearers interpreted the negated evaluative adjectives sometimes as implicatives, i.e. under matrix negation, the complement is also interpreted as negated (*No* answer in the table). A first hypothesis that comes to mind is that speakers simply misread the stimulus and think that in fact the sentence was (2):

(2) John was not stupid *enough* to waste money.[1]

That hypothesis might be plausible for oral presentation but, in our case, the stimuli were always presented in written form. Moreover, a study of web data and of the enTenTen-2.0 corpus[2] revealed that, for some of these adjectives, the implicative use seems to be the most prevalent one: e.g. for *lucky*, 9 out of 11 occurrences are implicative, for *fortunate*, 16 out of 18 are, for *stupid*, two fifths of the uses are implicative, two fifths, factive and one fifth could not be determined in the enTenTen-2.0 corpus. To give just a couple of examples; (3) from the enTenTen-2.0 corpus and (4) from the web:

(3) I have a family to support and I'm not stupid to put that in jeopardy, maybe you are.

(4) This is my first trip to Italy, so I was not brave to venture out alone.

Moreover, in the first large experiment that we report on in Section 3, we asked the subjects that gave implicative answers whether they would say the same thing to convey the interpretation that they had assigned to the stimulus sentence. 79% answered this question positively. A further small investigation, exemplified in Table 2, quickly led to one variable that seemed to influence the judgments. (In what follows we indicate the factive with F and the implicative reading with I.)

The second column shows that the relation between the adjective and the VP complement is not the same in all cases. In some cases, *for NP to VP* would be ADJ (to choose the best piece would be clever, to waste money would be stupid) the relation is the socially expected one. We call this the CONSONANT relation. These combinations often get an implicative reading. In others, *for NP to VP* would **not** be *ADJ* (to choose the worst piece would not be clever, to save money would not be stupid), the relation is not the socially expected one. We call these DISSONANT. These cases have a nearly unambiguous factive reading.

To spell out the hypothesis explicitly: *for John to save money would not be stupid* is a dissonant combination of adjective and VP hence a factive interpretation is likely for (5):

(5) John was not stupid to save money.

For John to waste money would be stupid is a consonant combination, hence an implicative reading is likely for (6):

[1] This example seems to entail that John did not waste money. The *be ADJ enough to VP* construction is in general systematically ambiguous between an implicative and non-committal reading. (See Karttunen (1971) and Meier (2003) for discussion.)

[2] `https://www.sketchengine.co.uk/xdocumentation/wiki/Corpora/enTenTen`.

Table 2: Emotive and evaluative adjectives.

STIMULUS	ADJECTIVE-COMPLEMENT RELATION	ANSWERS	CHOICE	%
R. was not clever to choose the best piece	to choose the best piece is clever CONSONANT	R. chose the best piece	F	25
		R. did not choose the best piece	**I**	**64.2**
		undecided		10.7
R. was not clever to choose the worst piece	to choose the worst piece is not clever DISSONANT	**R. chose the worst piece**	**F**	**80**
		R. did not choose the worst piece	I	10
		undecided		10
K. was not stupid to save money	to save money is not stupid DISSONANT	**K. saved money**	**F**	**78.6**
		K. did not save money	I	14.2
		undecided		7.1
K. was not stupid to waste money	to waste money is stupid CONSONANT	K. wasted money	F	28.6
		K. did not waste money	**I**	**66.7**
		undecided		4.8

(6) John was not stupid to waste money.

(5) is assumed to cohere with people's social expectations, whereas (6) is assumed to be surprising under the factive reading, whereas an implicative interpretation would render it coherent.

3 Experiment 1 and the Implicative Dialect Hypothesis

Our 2014 paper (Karttunen et al. 2014) details a first large scale experiment in which we showed that the consonance/dissonance distinction is indeed involved. The experiment was run on Amazon Mechanical Turk with 206 native speakers of English who each responded to 30 test sentences which exemplified 19 adjectives (*arrogant, brave, careless, cruel, evil, foolish, fortunate, heroic, humble, lucky, mean, nice, polite, rude, sensible, smart, stupid, sweet* and *wise*). The test sentences presented the adjectives in a VP context that we judged either consonant, dissonant or neutral. The subjects had the choice between an implicative interpretation (according to the author of the sentence the event referred to in the embedded infinitival did not happen), a factive interpretation (the event happened) and *Either.* (For a detailed description of the experiment, see Karttunen et al. (2014).)

The results confirm the importance of the consonant/dissonant distinction as shown in the summary in Figure 1.

In the 2014 paper, we put forward the hypothesis there might be a dialect split between speakers of English. One group for which the factive use is the norm, another group for which the implicative use is. Both groups would adapt to the other dialect if the consonance or dissonance of the infinitival complement nudged them to do so. Our reasoning was that, if the only factor at issue was the consonance/dissonance nature of the complements, speakers should give factive judgments for neutral stimuli and only be swayed by discourse coherence pressures in case of negated consonant

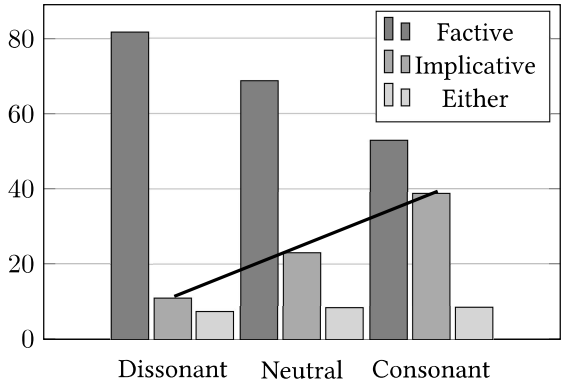

Figure 1: Results: Percentage of Factive, Implicative, and Either choices for *NP was not ADJ to VP* .

readings.[3] But if we have two dialects we expect there to be implicative judgments even in the neutral contexts. In fact, the neutral contexts would reveal the relative strength of the two dialects. Given the data in Figure 1, around one third of the speakers could be hypothesized to be implicative speakers.

But the results did not establish this dialect split conclusively as we had too few judgments per subject. Moreover we did not have any sociological data about our subjects. An other problem with the experiment was that we had labeled the infinitival complements as consonant or dissonant on an intuitive basis without checking whether our own judgments coincided with those of the community.

4 Experiment 2: consonance/dissonance judgments

To address the latter criticism, we ran a norming experiment asking our subjects to judge *How ADJ it would be to VP* (e.g. *How brave would it be to fight injustice in the face of danger?*), on a sliding scale (extremely = 1, not at all = .50, completely the opposite = 0). This experiment confirmed most of our intuitive judgments (with some exceptions: e.g. contrary to us, most of our subjects thought that living in Europe was an unlucky experience). We then combined these results with the results of a new consonance/dissonance experiment of the same design as the previous one but adding the adjectives *cowardly, kind, prudent, right* and *wrong* and leaving out *sweet*, ending up with 23 adjectives. The results showed that the the consonance/dissonance distinction was highly significant for the choice between the two interpretations:[4]

[3] Our results focus on negated sentences: in positive sentences, there is no difference in the interpretation of the embedded complement as having happened or not between implicatives and factives. There might be more subtle differences but these our experiments do not address.

[4] The p-value is a measure of the likelihood that the result of an experiment is due to chance. An experiment with a p-value lower than 0.001 ($p < 10^{-3}$) is commonly accepted as a statistically highly

Factive: p-value < $2.2 \cdot 10^{-16}$
Implicative: p-value < $2.2 \cdot 10^{-16}$
Cannot Decide: p-value = $3.261 \cdot 10^{-5}$

But the correlation between the consonance/dissonance judgments and the factive/implicative readings also showed that consonant/dissonant variable explained less than fifty percent of the variation:[5]

Factive: Adjusted R-Squared = 0.433
Implicative: Adjusted R-Squared = 0.4344
Cannot Decide: Adjusted R-Squared = 0.07718

In the 2014 paper, we already observed that not all adjectives in our sample were equally sensitive to the consonant/dissonant distinction. The findings reported there confirmed what we had learned from the corpus data for *stupid, fortunate* and *lucky* (see Karttunen et al. (2014) for details). A further inspection of our list of adjectives suggested that they fall in a few broad classes. *Arrogant, cruel, evil, humble, mean, nice, polite, rude, kind* are about character, *fortunate* and *lucky* are about good or bad luck, *brave, heroic, cowardly* are about courage, *foolish, prudent, sensible, smart, stupid, wise* are about judgment whereas *right* and *wrong* give an overall moral appreciation. With respect to factive and implicative readings the classes are ordered as shown in Figure 2.

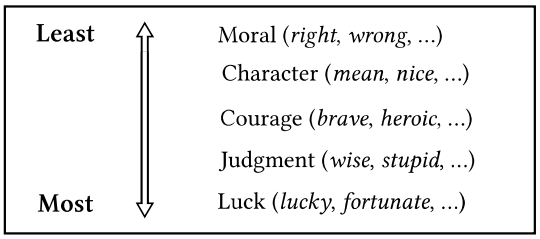

Figure 2: Adjective classes.

Right and *wrong* are least affected by the consonant/dissonant effect, *lucky* and *fortunate* are at the opposite end of the scale. When we add these classes to our calculations, the measure of variation accounted for improves dramatically to an adjusted R-Squared of 0.7929. Of course, this only tells us that the distinction among the classes is important but not what the distinction exactly is and certainly not how it should be explained. We are currently trying to find an operational characterization of this distinction, hypothesizing that factors like the control of the protagonist

significant result.

[5] R-Squared is the percentage of the response variable variation that is explained by a linear model. The adjusted R-squared is a modified version of R-squared that has been adjusted for the number of predictors in the model. The adjusted R-squared increases only if the new term improves the model more than would be expected by chance.

over the characteristic given in the adjective (one can be *mean* on purpose but one cannot be *lucky* on purpose) play a role.

5 Experiment 3: the Implicative Dialect Hypothesis rejected

In our third experiment, we concentrated on 39 native speakers of English. We recorded data about the age, the gender and the regional provenance but did not balance the sample with respect to these variables. We chose 6 adjectives (*lucky, fortunate, mean, nice, stupid, foolish*) from our previous set and added *right* and *wrong* to them and presented 9 sentences (three consonant, three dissonant and three neutral ones) per adjective to each subject. A logistic model of the probability of the answer as a function of the consonance score correlated, for each subject, his/her judgments with the judgments about consonance and dissonance that we had obtained in our previous experiment. The results for the negative examples confirm that there are subjects that interpret all these adjectives unambiguously as factive but they are a minority (8/39 allowing one 'mistake' per speaker, 5/39 counting very strictly). There are, however, no subjects for which these adjectives are unambiguously implicative. For most subjects, we find a mixed pattern that, as expected, contains more implicative judgments for the consonant examples but also a substantial number of implicative judgments for neutral cases. For about half of the subjects the susceptibility to the consonance/dissonance dimension is quasi-linear, for the other half the susceptibility increases the more we get to the consonant end of the scale. As a rough measure of the degree of susceptibility we calculated the difference between the intercept at consonance 0 and that at consonance 100 on the normed scale. According to this calculation (leaving out the unambiguous factive interpreters for whom the difference is of course zero), we get differences between 7.95 and 83.43 (on a scale for 1 to 100), meaning that the difference in the probability that a subject will give an implicative answer for a consonant sentence compared to the probability that (s)he will give an implicative answer for a dissonant sentence is between 7.95% and 83.43%! This is very sizable individual variation among speakers who are not reliably factive. For the speakers at the very high end, the consonant/dissonant factor explains the difference between factive and implicative judgments nearly completely but for other speakers, especially those that do not give consistent factive judgments at the dissonant end of the scale, other factors must play a role (one of these factors being the difference in susceptibility for different adjective classes).

The aim of this experiment was to see whether there are any reliable implicative interpreters. The answer to that question is *no*. That doesn't mean that the differences among speakers that we found might not correlate with sociological variables. We recorded the regional provenance, the age and the gender of the participants. 39 subjects, however, is a very small number to study these variables. For age, we put the subjects in four buckets (see Table 3).

One can argue that unambiguous factive interpreters are less common under

Table 3: Age related differences.

	number of subjects	number of unambiguous factive interpreters
19-30 years	23	4
31-40 years	11	2
41-50 years	2	1
51-60 years	3	1

younger subjects but it seems rather hazardous to draw conclusion from such a small sample. The subjects came from four regions in the U.S. The distribution is given in Table 4.

Table 4: Regional differences.

	number of subjects	number of unambiguous factive interpreters
Northeast	10	4
Midwest	10	2
South	12	1
West	5	0

Again, there is a possible trend but not enough data. As far as gender goes, we had 21 males among them 4 unambiguous factive interpreters and 18 females again with 4 unambiguous factive interpreters among them.

In all the categories, the unambiguous factive interpreters were a minority. However, one possibly important factor we do not have information for, is level of formal education.

As before, there was a difference in the adjective classes. Given we have only 2 adjectives per class, it is not possible to conclude much from the data here but overall the data confirmed the ordering of the classes that we gave before (except for the courage class that was not represented). The trends are clear from Table 5 (in percentages).

Right and *wrong* are nearly always factive in all conditions. The *luck* class is interpreted as implicative in most cases when it is in a consonant context, and even in a neuter one but that result might not be significant. The two other classes give rise to a majority of factive interpretations but there is a big difference, especially in the consonant class, between the judgment class and the character class. We have yet to do a formal susceptibility calculation per class but the data is very limited.

Table 5: Judgments per adjective class.

	dissonant			neuter			consonant		
class	yes	no	?	yes	no	?	yes	no	?
Moral	33	0	0	27	0	0	30	2	1
Character	31	0	2	27	3	3	24	7	2.5
Judgment	27	4	2	23	8	2	17	14	2
Luck	27	5	1.5	15	16	2	11	21	1

6 Conclusion

The experiments described above show that native speakers of English differ in their interpretation of negative sentences with evaluative adjectives and infinitival complements. They also indicate that this is most likely not a dialect variation although it is possible that age plays a role. We have at this point no good explanation for the difference but the importance of the consonant/dissonant factor suggests that it might be the case that it is not so much a linguistic difference than a difference in the way different language users evaluate the importance of linguistic structure versus the importance of arriving at interpretations that are coherent with their overall beliefs about the state of the world. A similar result was found in a study by Wason & Reich (1979) (See also Cook & Stevenson (2010)). The preference for an interpretation coherent with the previous state of beliefs is most likely more common when there is an additional load on sentence processing such as negation. But it remains astonishing that one finds these effects even in a setting where there is no time pressure or any other factor that would explain degraded performance.

It remains also to be explained why we find the effect not only in understanding but also in production as shown by the corpus data. Here the fact that the construction under investigation is very close to one that can have the intended implicative meaning (exemplified in (2)) may play a role.

References

Cook, Paul & Suzanne Stevenson. 2010. No sentence is too confusing to ignore. In *Proceedings of the 2010 Workshop on NLP and Linguistics: Finding the Common Ground*, 61–69. Uppsala, Sweden.

Karttunen, Lauri. 1971. Implicative verbs. *Language* 47(2). 340–358.

Karttunen, Lauri, Stanley Peters, Annie Zaenen & Cleo Condoravdi. 2014. The chameleon-like nature of evaluative adjectives. In Chris Piñòn (ed.), *Empirical Issues in Syntax and Semantics*, vol. 10, 233–250. CSSP-CNRS.

Karttunen, Lauri & Annie Zaenen. 2013. Factive investigations. In *Structure and Evidence in Linguistics*. Stanford.

Meier, Cécile. 2003. The meaning of *Too*, *enough*, and *So…That*. *Natural Language Semantics* 11(1). 69–107.
Norrick, Neal. 1978. *Factive Adjectives and the Thoery of Factivity*. Tübingen: Niemeyer.
Wason, Peter & Shuli Reich. 1979. A verbal illusion. *Quaterly Journal of Experimental Psychology* 31. 591–597.