Listening and clarity of syntactic structure

Virginia Valian*

CUNY Graduate Center, New York, N.Y. 10036, U.S.A.

Abstract: Listeners repeated fully grammatical sentences, exemplifying 12 linguistic constructions, heard through noise. In half the sentences the basic grammatical relations or constituent structure were more clearly displayed than in the matching versions. Although the differences in structure between the two versions were minimal (often just the presence or absence of a function word), the "clear" sentences were correctly repeated on the average 19% more often than the "distorted" sentences were. The results suggest that minor structural cues are important in listening to speech, at least under adverse conditions.

No sooner do we hear the words of a familiar language pronounced in our ears but the ideas corresponding thereto present themselves to our minds: in the very same instant the sound and the meaning enter the understanding: so closely are they united that it is not in our power to keep out the one except we exclude the other also. We even act in all respects as if we heard the very thoughts themselves. (Berkeley, 1901, 151, rubric 51).

   Berkeley's comments on the transparency of speech in much of everyday listening have a methodological implication for psycholinguistics: the individual contributions to understanding of syntax, vocabulary items, pronunciations, and so on will be hard to measure. When listening is easy, small hindrances such as a missing syntactic cue will not be noticeable, because there is so much compensating material. Thus, it will be hard to test Fodor and Garrett's (1967) suggestion that the presence of surface structure cues to underlying sentential relations aids understanding.

   In a study of one kind of surface structure cue, Fodor & Garrett (1967) found that people could paraphrase doubly self-embedded sentences with relative pronouns (e.g. "the man *that* the girl knew got sick") more accurately than they could the same sentences with the relative pronoun deleted. Several later studies by Hakes (Hakes & Cairns, 1970; Hakes & Foss, 1970; Hakes, Evans & Brannon, 1976) showed that, in general, the absence of relative pronouns slowed performance in a phoneme-monitor task, but did not consistently improve paraphrase accuracy. A suggestive side note is that Hakes, Evans & Brannon (1976) found that relative pronouns aided phoneme monitoring speed for both subject-object and object-object relatives but not for either subject-subject or object-subject relatives, which independent evidence indicated were easier to process. On the other hand, Hakes *et al.* found that relative pronouns aided phoneme monitoring speed for both subject–object and object–object relatives but not for either subject–subject or object–subject relatives, which Fodor and Garrett's argument is that relative pronouns and other function words that provide syntactic cues are the words hardest for listeners to identify out of context and the words most likely to be slurred over in speech (cf. Clark & Clark, 1977).

*Present address: Department of Psychology, Schevmerhorn Hall, Columbia University, NYC 10027, U.S.A.

In order to determine what role syntactic cues play in comprehension distinguish the ordinary listening situation Berkeley describes from an attenuated version of the one he alludes to when he says "so closely are [sound and meaning] united that it is not in our power to keep out the one except we exclude the other also". In a noisy situation speech is not transparent. The listener is aware of straining to hear each word and of attempting to reconstruct the utterance once it is over. A noisy background makes listening more difficult; the contribution of syntactic cues to understanding should be correspondingly more noticeable.

Two studies provide indirect confirmation. Valian & Wales (1976) and Valian & Caplan (1979) investigated how talkers behaved when a listener (experimenter) asked "What?" after a sentence they had read. Across the constructions tested, adults and children changed very slightly "distorted" (but fully grammatical) sentence versions like those presented in the second column of Table 1 to "clear" versions like those in the first column significantly more often than the reverse, but not for constructions 3, 6, 7 and 10. In the sentences of Table 1 the differences between the two sentence versions are minimal: in the first two constructions a relative pronoun or relative pronoun plus *be* is present or absent; in the next three constructions the complementizer *that* is present or absent. For the first five constructions, then, word order is kept constant, but a cue that makes a constituent boundary and the introduction of a new clause explicit is deleted.

In the remaining constructions word order is used as a cue, as well as word presence or absence. The closer two or more elements are in surface structure, the better they display their joint memberhsip in a grammatical category or the modifer-head relation. In permuted relatives, 8, the unpermuted version does not interrupt the subject noun phrase and hence contributes to clarity of constituent structure and the basic grammatical relation subject

Table I: Examples of clear and distorted sentence versions
for each linguistic construction type

		Clear	Distorted
(1a)	Subject—object relative	The treasure that she found was valuable.	The treasure she found was valuable
(1b)	Object—object relative	Tom watered the plant that the florist had sold him.	Tom watered the plant the florist had sold him.
(2)	Subject—subject relative with aux	The people who were criticizing the politician were angry.	The people criticizing the politician were angry.
(3)	Object NP complement	Roger insisted that he had the right of way.	Roger insisted he had the right of way.
(4)	Subject NP complement-	It gratified Marcy that her thesis was a success.	It gratified Marcy her thesis was a success.
(5)	Subject NP complement-intrans verb	It appears that William is going to Chicago.	It appears William is going to Chicago.
(6)	Tag questions	The chef hasn't started our order yet, has he?	Has the chef started our order yet?
(7)	Manner adverbials	Ginny persuasively argued her case.	Ginny argued her case persuasively.

(8)	Permuted relatives	Somebody who loves me called me.	Somebody called me who loves me.
(9)	Verb plus particle	Jesse put on his shirt.	Jesse put his shirt on.
(10)	To-dative	The salesman sold a watch to Jerry.	The salesman sold Jerry a watch.
(11)	Regular passive	The spy divulged the secret to Emma.	The secret was divulged to Emma by he spy.
(12)	Double-agent passive	Tom took advantage of Lou. (4 only)	Lou was taken advantage of by Tom. Advantage of Lou was taken by Tom.

of the sentence. In verb + particle, 9, the fact that they jointly make up the constituent *verb* is clearly displayed in the uninterrupted version. In manner adverbials, 7, on the assumption that the adverb modifies the verb, that relation is more perspicuous if the modifier and head are adjacent.

Position and presence of the word *to* make the indirect object explicit in 10. If, as Katz & Postal (1964) claim, yes—no questions, 6, have a disjunctive deep structure, it is better preserved in the tag form than the non-tag form. Finally, the passives, 11 and 12, are hard to categorize. On the one hand, in the clear version the positional cues allow a match between the surface 'subject' and deep subject, and surface 'object' and deep object. On the other hand, the distorted version has considerably more syntactic baggage in it: *be+en* and *by*.

Valian & Wales (1976) speculated that talkers in a formalized What? situation changed distorted versions to clear ones because they thought such changes would make the sentence easier for the listener to hear and understand. Delis & Slater (1977) have also investigated the contribution of syntactic cues, but manipulated a different variable. They asked subjects to compose a passage from elementary sentences about cell biology for naive or sophisticated audiences; some subjects spoke their passage, other subjects wrote. Delis & Slater found that subjects talking or writing for sophisticated audiences used the most reduction "transformations" (e.g. changing "The molecule contains energy" to "It contains energy"), while subjects talking to naive audiences used the fewest.

The Valian and Wales, and Delis and Slater results indirectly suggest that minimal structural cues are most important for listening when there is difficulty in understanding, either because the background noise is high, or because the conceptual level of the material is difficult for the listener. Or so talkers think. The only direct evidence about the relevance of syntactic information in a noisy background comes from a study by Miller & Isard (1963), who found that fully grammatical and meaningful sentences were better shadowed through noise than were syntactically well-formed anomalous strings, which in turn were easier than word salad. Miller and Isard's syntactic manipulations were relatively crude, however, compared to the changes subjects make in the Valian experiments and the Delis and Slater study.

A test of the hypothesis that minimal structural cues will be highlighted in importance for a listener under difficult hearing and understanding conditions is to play people the sentences of Table I through noise, ask them to shadow them or repeat them immediately after the sentence, and see if more accurate retrieval occurs in sentences with clear (as defined above) rather than distorted structure. Listening through noise is a laboratory

analogue to the listener's plight in a noisy background. The two different tasks, shadowing and repeating immediately after, are also analogous to the listener's efforts both to get each word and to reconstruct the utterance once it is finished. Syntactic cues may be important only in reconstruction, rather than in direct perception.

It is important to emphasize how minor the syntactic cues in the present experiment are that differentiate the clear and distorted versions, often just one word that is present or absent, or that appears in one position or another. There is little empirical evidence to back up the importance of such minor syntactic cues and, except within a theoretical framework in which such cues crucially contribute to comprehension, it would seem counter-intuitive to predict that minimal structural cues will be important in listening to speech through noise.

Method

Subjects

Subjects were 36 paid volunteers who responded to signs in the CUNY Graduate Center requesting native English speakers of normal hearing. There were 18 subjects in each of the two response conditions.

Materials

Experimental and filler sentences were the same as those used by Valian & Wales (1976). Data are reported for the 12 constructions listed and exemplified in Table I. For constructions 1–11 there were eight sentence pairs; the syntactic relations were clear in one version and distorted in the matching version. For construction 12 there were four clear forms and eight distorted forms; an active here had two corresponding passive forms. Most experimental sentences were between six and nine words long.

Subjects heard a total of 196 sentences: 20 practice sentences, 94 experimental sentences representing the 12 constructions discussed here, 52 sentences representing an additional 7 constructions and 30 filler sentences. Of the seven constructions for which data are not reported, one construction (reported by Valian & Wales) did not present sufficient data: two sentences were only one word long in their distorted version, and were thus frequently missed altogether. Their inclusion would have biased the results in favor of the hypothesis. Of the remaining six constructions, five explored variations not encompassed by the clear-distorted specification, such as nouns vs gerunds, and one contained strings in putative deep structure form.

Design

Within each response condition (shadowing or immediately after) subjects were divided into two groups of nine each. One group heard four clear and four distorted sentences from each of constructions 1–11 and the other group heard the eight complementary versions. For construction 12 (double-agent passives) one group heard two active and four passive sentences and the other group read the six complementary sentences. Each subject received the same random order of sentences.

Assignment of subjects to reponse condition was not random, because many subjects are unable to shadow well. Subjects were assigned initially to the shadowing condition; if their performance during the first 15 practice sentences indicated that shadowing was too hard they were switched to the immediately after condition for at least the final five practice sentences (noise level = +2) and the test sentences. Once 18 subjects had been obtained for the shadowing condition, all additional subjects were assigned to the immediately after condition. This subject assignment procedure placed good shadowers in the shadowing groups and bad shadowers plus other subjects in the immediately after group. Hence, any absolute differences between the groups must be interpreted with caution.

Apparatus and procedure

A female speaker (VVV) recorded two sentence sets, one each per tape, in the same random order on both tapes. The sentences were read with flat intonation through a speech compressor that clipped amplitude peaks. A brief signal tone was recorded before each sentence, and a calibration tone was placed at the beginning of each tape.

Subjects were tested individually. They were told they would listen to sentences through noise, so that they would not be able to hear the sentences clearly. They were asked to repeat what they heard, and to guess when they were unsure. Subjects in the shadowing conditions were asked to begin repeating the sentences as soon as they could without waiting for the end of the sentence. Subjects in the immediately after condition were asked to begin repeating the sentence immediately after it was over. All subjects were given 20 practice sentences, five with no noise, five with noise at -3 on a Grason-Stadler Model 901B noise generator decibel meter, and the remaining ten at 0 or +2.

Subjects binaurally listened to the sentences and noise through Koss Pro-4A headphones. For each session the calibration tone measured +3 on the noise generator decibel meter; the noise measured 0 in the shadowing condition and +2 in the immediately after condition. (The noise level was lower in the shadowing condition because of the additional noise created by bone conduction of the subject's own voice.) Subjects' responses were manually transcribed and tape-recorded during the experiment. All doubtful transcriptions were later checked against the recordings.

Control measures

The duration in s and average db level of each sentence was measured at the Speech Communications Group of the Research Laboratory of Electronics at the Massachusetts Institute of Technology. The computer displayed each sentence visually and auditorily, so that beginning and end points could be manually marked. The computer measured the duration of a marked sequence, measured the total energy in the signal, and divided energy by duration to compute average energy. The mean level of clear sentences was 84·64 db *re* an arbitrary standard, of distorted sentences 84·69 db. A repeated measures analysis of variance using items as the random effect, with syntactic form (clear *vs* distorted) the within-items variable and construction (1—12) the between items variable showed no difference in db as a function either of form or construction, and no interaction.

The mean duration of clear sentences was 2·67 s, of distorted sentences 2·58 s. A similar repeated measures analysis of variance showed, as expected, that the duration of clear sentences was reliably longer than the duration of distorted sentences, $F(1,84) = 16·82$ $P<0·001$, that some constructions were longer than others, $F(11,84) = 1·94$, $P = 0·045$, and that there was an interaction between construction and syntactic form, $F(11,84) = 15·15$, $P<0·001$. Thus, it was important to see if any observed retrieval advantage of clear sentences could be attributed to their greater length (see Results).

Results

The main result of interest is whether more clear sentences than distorted sentences are repeated verbatim. The average percentage of verbatim repetitions is shown in Table II. The mean percentage of clear sentences repeated verbatim was 21·4, for distorted sentences it was 18·0. Both the subjects and items analyses were significant at an acceptable level, $F_1(1,34) = 3·998, P = 0·054; F_2(1,84) = 3·787, P = 0·056$, but *min F'* was not.

Table II: Percent verbatim repetitions of clear and distorted sentence versions

Construction[a]	Shadowing		Immediately after		$\overline{X}$	
	clear	distorted	clear	distorted	clear	distorted
(1) relative ± marker	18·1	19·4	13·9	20·8	16·0	20·1
(2) relative ± mrk + aux	30·6	22·2	29·2	13·9	29·9	18·1
(3) obj NP complement	33·3	43·1	31·9	20·8	32·7	31·9
(4) subj NP comp-trans v	16·7	8·3	22·2	8·3	19·4	8·3
(5) subj NP comp-intr v	26·4	22·2	30·6	15·3	28·4	18·8
(6) yes-no (tag) questions	27·8	15·3	23·6	29·2	25·7	22·2
(7) manner adverbials	13·9	13·9	4·2	9·7	9·0	11·8
(8) permuted relatives	22·2	15·3	23·6	11·1	22·9	13·2
(9) verb + particle	36·1	19·4	29·2	29·2	32·7	24·3
(10) to-dative	20·8	22·2	16·7	13·9	18·8	18·1
(11) passive	15·3	12·6	16·7	16·7	16·0	14·6
(12) double-agent passive	5·6	19·4	5·6	9·7	5·6	14·6
X	22·2	19·4	20·6	16·6	21·4	18·0

[a]sentences/form/construction = 8, except for (12), where clear version has four sentences.

The main effect of construction (1–12) was also significant, $F_1(11,374) = 10·79$, $P < 0·001$; $F_2(11,384) = 1·89, P = 0·052$, but the interaction between construction and syntactic form was not reliable, $F_1(11,374) = 2·68, P = 0·003; F_2(11,84) = 1·24, P = 0·276$. Some constructions are easier to repeat correctly than others, but there is not a reliable difference in the advantage of clear over distorted as a function of construction.

The main effect of response condition (shadowing *vs* immediately after) was not reliable, $F_1(1,34) = 0·365, P > 0·50; F_2(1,84) = 0·289, P = 0·093$, although 20·8% of sentences were correctly shadowed and 18·6% were correctly repeated immediately after. Response condition did not interact with syntactic form or construction type, nor was the three-way interaction significant. That is, the likelihood of complete retrieval does not differ as a function of on-line vs slightly delayed report.

Since the clear sentences were consistently longer than the distorted sentences, the possibility existed that their greater length was responsible for their better retrieval. One of the two conditions was arbitrarily chosen to test whether greater length of clear sentences correlated with greater retrieval. For the immediately-after results, a Pearson product moment correlation was computed between clear length minus distorted length of each sentence and number of correct clear repetitions minus number of correct distorted repetitions. It was not significant, $r(96) = 0·10$, $P = 0·16$. Thus, although clear sentences were 90 ms longer than distorted sentences, their extra length did not correlate with the extent of the superiority of their retrieval in the immediately after condition.

Discussion

Two aspects of the results are problematic. First, although there was a significant facilitating effect of clear syntactic form on number of completely correct repetitions whether subjects or items were used as the random effect, *min F'* was not significant. Second, although there is an overall advantage for clear forms in each response group, and although there are only three reversals when the two response groups are pooled (which the lack of a main effect for response group and of an interaction between response group and syntactic form shows is appropriate), the shadowing group considered alone shows four reversals and one case of no difference, and the immediately after group considered alone shows four reversals and two cases of no difference.

Both problems indicate large variability, for which there may be two contributing causes. The kinds of differences under experimental control, namely the presence or absence of minimal syntactic cues, are much smaller than the kinds of differences not under experimental control, namely length, subject matter, and vocabulary. Although subject matter and vocabulary were held constant within matching sentence versions, they varied widely across sentences, as did length. The importance of minimal syntactic cues is obviously greater for some sentence materials than others, but a *post hoc* sentence by sentence inspection of the data did not suggest any properties held in common by sentences that show improved retrieval for clear versions.

Another cause for variability is that the phonetic environments which differed from sentence to sentence may have conditioned the pronunciation of some cues in such a way that they were easily masked by the noise. One problem in using low signal to noise ratios as the technique for inducing difficult listening is that there is no guarantee that the syntactic cues which should aid listening will be heard.

Even with the caveats entered, however, the results do support the prediction that minimal syntactic cues will be of measurable importance under difficult listening conditions. The result is surprising, considering how small the cues are in relation to the other sources of variability within the experiment, and how crude a measure percentage of correct repetitions is. The lack of an interaction between syntactic form and response condition indicates that syntactic cues are made use of in direct perception as well as in reconstruction afterwards, supporting Marslen-Wilson's (1975) observations of shadowers.

Although the lack of a consistent interaction between syntactic form and construction does not warrant looking at constructions separately, there are two reasons for doing so informally. The first is to compare the present results on constructions 1 and 2 with those of Hakes *et al.* (1976). The second is to compare listeners' behavior with talkers' behavior on the same constructions.

First, the comparison with Hakes *et al.* (1976) shows agreement with their paraphrase accuracy measure, but not their phoneme-monitor speed measure. Since paraphrase accuracy requires understanding at a deeper level than does phoneme detection, it is perhaps not surprising that our correct retrieval results correspond with Hakes *et al's* paraphrase results. For relatives like those in contruction 1 Hakes *et al* found better paraphrase for relatives *lacking* the relative pronoun; in the present experiment, retrieval was also better for relatives lacking the pronoun (distorted versions). Similarly, for relatives like those in construction 2 Hakes *et al.* found marginally better paraphrase accuracy for relatives *containing* the relative pronoun + auxiliary; in the present experiment retrieval was also better for versions containing the pronoun + auxiliary (clear versions). Unexplained in both experiments is why subject and object relatives behave differently.

Second, the comparison of the present results for listeners with those of Valian & Wales for talkers shows several cases of non-correspondence. [The two sets of results cannot be directly compared, because talkers make many kinds of changes (phonological, syntactic and semantic) simultaneously, whereas only one kind of change per construction was manipulated in the present experiment. There are however, three constructions where talkers' and listeners behaviour are very different]. Talkers overwhelmingly change tag-questions to yes-no questions (construction 6), while listeners are variable in which form they more often repeat correctly. Talkers overwhelmingly change both kinds of passives (constructions 11 and 12) to actives, while listeners show either only a slight advantage for actives (construction 11) or a great advantage for passives (construction 12).

As mentioned in the introduction, passives have much more syntactic baggage in them than actives do, which may account for their good performance. In listening to the passives through noise, one easily perceived the *was* and *by*. Compton (1967) also found passives to be quite intelligble, more so than any other sentence type except simple active affirmative declaratives. Since Compton's actives were not matching versions of his passives, his actives

may have been simpler in other ways as well. The tag-question form (clear version) also has more syntactic cues than the yes-no form, which may account for their slight advantage overall (but does not explain the variability as a function of response condition).

The existence of non-correspondences between talkers' and listeners' preferences suggests that, all other things being equal (which they seldom are in a What? situation, because by raising his or her voice a talker can compensate for changes which would otherwise interfere with comprehension), some changes talkers make are not useful to listeners. By and large, the results from the present experiment indicate that the cooperative talker will help the listener most by maximizing the presence of structural cues.

The studies reviewed in the introduction were limited in their use of syntactic cues to the presence or absence of relative pronouns. The present study uses a much broader range of syntactic cues, including relative pronouns, complementizers, particle placement, adverb placement, and so on. The results indicate that many kinds of minimal structural cues are important in understanding, and that their importance can be seen in difficult listening situations. Only one kind of difficulty was used here; future experiments should confirm these results with other kinds of difficult situations and with other measures of understanding.

References

Berkeley, G. (1901). An essay towards a new theory of vision. In *Berkeley's Complete Words.* (Frazer A. C., Ed.) Oxford, England: Clarendon Press.

Clark, H. H. & Clark, E. V. (1977). *Psychology and Language.* New York: Harcourt Brace Jovanovich.

Compton, A. (1967). Perception of sentences of different structures and lengths. *Language and Speech* 10, 81–87.

Delis, D. & Slater, A. S. (1977). Toward a functional theory of reduction transformations. *Cognition* 5, 119–132.

Fodor, J. A. & Garrett, M. (1967). Some syntactic determinants of sentential complexity. *Perception & Psychophysics* 2, 289–296.

Hakes, D. T. & Cairns, H. S. (1970). Sentence comprehension and relative pronouns. *Perception & Psychophysics* 8, 5–8.

Hakes, D. T. & Foss, D. J. (1970). Decision processes during sentence comprehension: effects of surface structure reconsidered. *Perception & Psychophysics,* 8, 413–416.

Hakes, D. T. Evans, J. S. & Brannon, L. L. (1976). Understanding sentences with relative clauses. *Memory and Cognition* 4, 283–290.

Katz, J. J. & Postal, P. M. (1964). *An Integrated Theory of Linguistic Descriptions.* Cambridge: MIT Press.

Marslen-Wilson, W. (1975). Sentence perception as an interactive parallel process. *Science* 189, 226–228.

Miller, G. A. & Isard, S. (1963). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior* 2, 217–228.

Thorne, J. P. (1966). On hearing sentences. In *Psycholinguistic Papers.* (Lyons, J. & Wales, R. J. Eds). Edinburgh: Edinburgh University Press.

Valian, V. & Caplan, J. (1979). What children say when asked "What?": a study of the use of syntactic knowledge. *Journal of Experimental Child Psychology* 28, 424–444.

Valian, V. & Wales, R. (1976). What's what: talkers help listeners hear and understand by clarifying sentential relations. *Cognition* 4, 155–176.