Comparative studies of comprehension:
An investigation of Chinese, Norwegian, and English

MARCEL ADAM JUST and PATRICIA A. CARPENTER

Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213

The present paper compared the processing of complex embedded sentences in Chinese, Norwegian, and English to determine if the underlying mental operations are similar in these diverse languages. The task involved reading a sentence like *It's true that the dots aren't red* and deciding whether it was true or false according to an accompanying picture. The verification latencies were analyzed in terms of a model based on one central mental operation. The results for all three languages conformed to the predictions of the model and showed similar processing rates. Moreover, the Norwegian study expanded the analysis to include quantification in sentences like *It's true that many of the dots aren't red*. These studies suggest that certain fundamental operations may be universally implicated in language processing.

One of the strongest links that people have with one another is the ability to communicate through language. The pervasiveness of this linguistic ability has aroused considerable interest in the cause and nature of its universality. For example, Lenneberg (1967) has documented the search for the biological basis of language, while Greenberg (1966) has explored some of the structural properties of language that are universal. These research efforts in the disciplines of biology and linguistics invite a parallel effort in psychology to discover universals of language processing. To this end, the present paper examines certain aspects of sentence comprehension in Chinese and Norwegian, with a view to comparing the processing in these two languages, as well as in English.

A comparative study of language processing requires a well defined experimental paradigm in which to examine performance, as well as a precise model of the underlying mental processes. These prerequisites have been provided by some of our own recent research (Carpenter & Just, 1975), in which we have developed a model of the mental processes involved in deciding whether a sentence is true or false according to an accompanying picture. The affirmative sentences in these studies were embedded sentences such as *It's true that the dots are red*, and could be negated in two ways. With one type of negation, the negative has a small scope, namely the inner predication: *It is true that the dots aren't red*. This will be called predicate negation. (The scope of the negation is simply the range of constituents to which it applies, cf. Klima, 1964; Jackendoff, 1969). The second type of negation, *It isn't true that the dots are red*, has a larger scope, since the negative in the superordinate clause applies to the entire inner proposition. This type of negation will be called denial.

The assumptions of the model and their justification are detailed elsewhere (Carpenter & Just, 1975), so we will only outline the representations and basic operations. The internal representation of sentences in this task is assumed to be propositional, which we will denote with a *(PREDICATE, ARGUMENT)* notation. The basic affirmative may be represented fairly simply as *(AFF,(RED,DOTS))*, meaning that redness is affirmatively predicated of dots. Similarly, the representation of a predicate negative may only differ by the fact that the predication is now negated, *(NEG, (RED, DOTS))*. However, the representation of a denial like *It isn't true that the dots are red* must also reflect the negative embedding clause, since it alters the truth value of the embedded clause. Thus, denial sentences may be represented as *(NEG, (AFF,(RED, DOTS)))*. Pictures are represented as simple propositions *(RED, DOTS)* or *(BLACK, DOTS)*, where the absence of an explicit polarity marker denotes affirmation. Table 1 shows the representations in the six conditions.

The main focus of the model is on the operations that compare the sentence and picture representations. The model postulates that the corresponding constituents from the two representations are retrieved and compared, pair by pair. Moreover, the number of these retrieve and compare operations is assumed to be the primary determinant of the pattern of verification latencies. Figure 1 shows the proposed process in flowchart form. The representations' propositional

Table 1
Representations and Predictions for the Six Information Conditions*

	True Affirmative	False Affirmative
Sentence	It's true that the dots are red.	It's true that the dots are red.
Picture	Red Dots	Black Dots
Sentence Representation	(AFF, (RED, DOTS))	(AFF, (RED, DOTS))
Picture Representation	(RED, DOTS)	(BLACK, DOTS)
	+ +	− Index = False
	Response = True	+ +
	k Comparisons	Response = False
		k + 1 Comparisons

	False Predicate Negative	True Predicate Negative
Sentence	It's true that the dots aren't red.	It's true that the dots aren't red.
Picture	Red Dots	Black Dots
Sentence Representation	(NEG, (RED, DOTS))	(NEG, (RED, DOTS))
Picture Representation	(RED, DOTS)	(BLACK, DOTS)
	− + Index = False	− Index = False
	+ +	− + Index = True
	Response = False	+ +
	k + 2 Comparisons	Response = True
		k + 3 Comparisons

	False Denial	True Denial
Sentence	It isn't true that the dots are red.	It isn't true that the dots are red.
Picture	Red Dots	Black Dots
Sentence Representation	(NEG, (AFF, (RED, DOTS)))	(NEG, (AFF, (RED, DOTS)))
Picture Representation	(RED, DOTS)	(BLACK, DOTS)
	− + + Index = False	− Index = False
	+ + +	− + + Index = True
	Response = False	+ + +
	k + 4 Comparisons	Response = True
		k + 5 Comparisons

*Plus and minus signs denote matches and mismatches of the corresponding constituents. Each horizontal line of plus and minus signs indicates a re-initiation of the comparison process.

structure and embeddings determine the sequence in which constituents are compared. Inner propositions are compared before polarity markers. The "find and compare" process is a serial, iterative operation. For example, in the true affirmative condition, the first comparison is between the inner propositions. Since they match, the next constituent to be retrieved and compared is the polarity marker. This is also a match because the absence of an explicit marker in the picture representation denotes affirmation. The number of "find and compare" operations in the true affirmative condition serves as a baseline for other conditions, and will be denoted as k. Here, k equals 2.

A central assumption is that whenever two corresponding constituents from the sentence and picture representations mismatch, then the entire comparison process is reinitiated. To prevent the process from looping forever on mismatching constituents, we assume that the first time a mismatch is discovered, the two constituents involved are tagged, so that on subsequent recomparisons the two will be treated as a match.

Since mismatches cause the comparison process to be reinitiated, the total number of comparison operations, and consequently the total latency, increases with the number of mismatches. Moreover, a mismatch that occurs later in the comparison process results in more recomparisons than a mismatch on earlier constituents. So the total latency is a function of both the number of mismatches and their locus in their respective representations.

A response index monitors the matches and mismatches between constituents. The index has two possible states, true and false. At the beginning of each trial, its initial state is true, but each mismatch causes it to change from its current state to its other state. The time spent in changing the response index (and, for that matter, tagging mismatching constituents) is assumed to be negligible relative to the time needed to perform the find and compare operation.

The effects of a mismatch between constituents can be observed in the false affirmative condition. The mismatch on the first "find and compare" operation between the inner propositions causes the reinitiation of the comparison process, resulting in one extra comparison above the base number. The mismatching constituents are tagged, and the index is set to false. After the reinitiation, the tagged inner constituents are compared and they match. The next comparison. between polarity markers, also results in a match. So,

the response *false* is executed after a total of k + 1 "find and compare" operations.

The exact predictions for the six conditions are worked out in Table 1, where a plus (+) reflects a match and a minus (-), a mismatch. The latency predictions are proportional to the proposed number of "find and compare" operations. The number of comparisons, and hence the latency, should increase linearly from true affirmatives (k), to false affirmatives (k + 1), to false predicate negatives (k + 2), to true predicate negatives (k + 3), to false denials (k + 4), to true denials (k + 5).

The experiment in English (Carpenter & Just, 1975) obtained a linear increase in latencies among the six conditions, constituting strong support for the constituent comparison model and the notion of a single underlying iterative operation. Figure 2 shows this result along with the best fitting straight line. The latencies increased an average of 200 msec for each additional comparison. The latencies in this experiment include the time for reading and representing sentences as well as comparison processes. However, a second experiment in English showed that latencies still increase linearly when reading and representation time is eliminated (Carpenter & Just, 1975). Thus, this supports the assumption that the underlying comparison process is a primary determinant of latencies in this task.

EXPERIMENT I

The purpose of Experiment 1 was very simply to determine whether the model accounts for verification latencies in Chinese. The experiment was identical to the one in English, except that the sentences were in Chinese.

Method

The experiment was a verification task in which the subject was timed while he read a sentence, looked at a picture, and then decided whether the sentence was true or false with respect to the picture. The stimulus sentences were translations of the three sentences in Table 1, as well as six more where the adjectives *black* or *green* replaced the adjective *red*. The picture was an

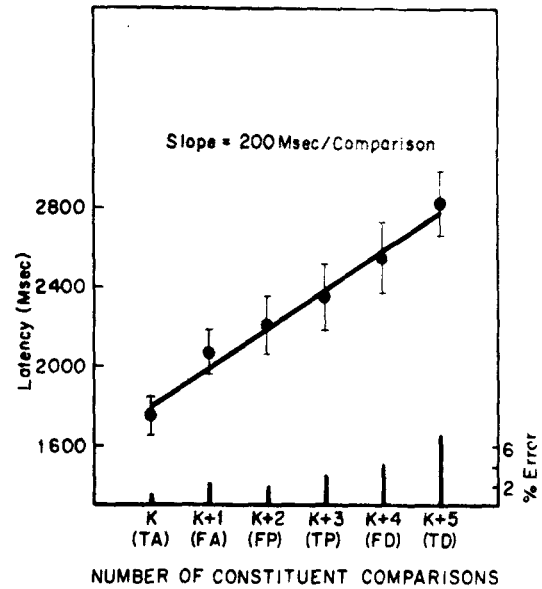Figure 1. The constituent comparison model.

Figure 2. The fit of the constituent comparison model to the data for English sentences (from Carpenter & Just, 1975).

array of 16 dots of one color, either red, black, or green. Each affirmative sentence could be false with respect to two pictures and true with respect to one picture; consequently, that one true sentence-picture combination was presented twice. The sentence-picture combinations for the false negative conditions were presented twice for the same reason. Thus, there was a total of 36 sentence-picture combinations. Each stimulus sentence consisted of Xeroxed copies of printed Chinese characters. Each sentence subtended about 10 deg of visual angle. The 4 by 4 array of dots, which was drawn below the sentence with a felt tipped pen, subtended 2.5 x 2.5 deg of visual angle. The stimulus card was viewed in a tachistoscope at a distance of 58 cm.

Examples of the Chinese sentences used in the experiment, as well as the English equivalents of the characters, are shown in Figure 3. The last character in each clause does not have an English equivalent. This character is an expletive that is often used as an ending for declarative statements. As the figure indicates, the Chinese equivalent of the word *these* consists of two characters. The Chinese sentences were written in a left-to-right format in the experiment, since Chinese students at American universities seem to be as adept at reading the left-to-right format as the top-to-bottom format. The English equivalents presented in Figure 3, of course, were not shown to the subjects.

The position of the negative character, relative to the copula verb, is different in the two types of negative sentences. In a denial sentence, the negative follows the copula verb and immediately precedes the character for *correct*. This expression was used because the Chinese characters for *not correct* constitute an entity, somewhat like the English *incorrect*. The alternative expression, translated "This is not correct," would be grammatically acceptable, but a more stilted way of speaking. The negative character in the predicate negative precedes the copula verb. This is the usual way of expressing negation and also reflects the fact that the negation of the adjective, *red*, does not constitute an entity. Thus, negation in these Chinese sentences is expressed slightly differently than in their English equivalents. However, the constituent comparison model specifies a matching process between semantic representations that is not directly dependent on surface structure. So this difference should not lead to a difference in the comparison process.
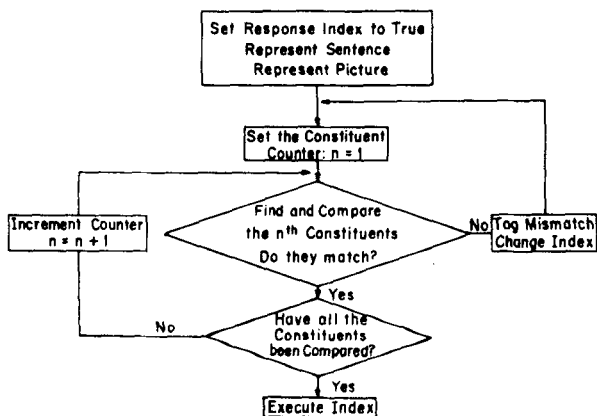
Affirmative

這　是　对　的，　這　些　点　是　红　的
↓　　↓　　↓　　　　↓　　↓　↓　　↓　　↓

THIS IS CORRECT , THESE DOTS ARE RED

Predicate Negative

這　是　对　的，　這　些　点　不　是　红　的
↓　　↓　　↓　　↓　　↓　　↓　↓　↓　↓　↓

THIS IS CORRECT , THESE DOTS NOT ARE RED

Denial

這　是　不　对　的，　這　些　点　是　红　的
↓　　↓　　↓　　↓　　↓　　↓　↓　↓　↓　↓

THIS IS NOT CORRECT, THESE DOTS ARE RED

Figure 3. Examples of the stimulus sentences used in Experiment I and their literal English equivalents.

The subject initiated a trial by pushing a switch, and 500 msec later the stimulus card was presented and remained in view until he responded. The subject's decision was made using a two-button decision apparatus. The assignment of dominant hand to "true" button was balanced across subjects. The subject was given feedback about the correctness of his answer only during the practice session, which consisted of 10 trials selected at random from the ensemble of 36. After the practice, all 36 cards were shuffled and the subject completed 5 blocks of 36 trials. The testing session lasted about 40 min. The subjects were 12 paid students and staff of Carnegie-Mellon University who were native Chinese speakers.

Results

The latencies for error trials (4.7%) were discarded, and the subject's score for each condition was the mean of his correct latencies for that condition. As predicted, the mean latencies increased linearly with the number of hypothesized constituent comparisons. More precisely, latencies increased an average of 210 msec for each additional constituent comparison (standard error = 34 msec). Figure 4 shows this result, along with the best fitting straight line. The model accounts for 91.8% of the variance among the six means $[F(1/55) = 115.80, p < .01]$. The residual 8.2% is just barely significant $[F(4/55) = 2.59, p < .05]$. The root mean squared deviation (RMSD) of 107 msec is small relative to the 210-msec parameter. The standard errors for the six means ranged from 156 to 272 msec, generally increasing with the means. This analysis confirms the major hypothesis that verification time increases linearly with the number of constituent comparisons.

The error rates for the six conditions were correlated with the latencies (r = .80), as shown in Figure 4. This

correlation indicates that the probability of error increases with the number of hypothesized operations.

Discussion

The results show a remarkable similarity between the processing in Chinese and English. The time per constituent comparison, 210 msec, is very close to the 200 msec found for English. Thus, processing rates and modes of processing are similar, even though the two languages come from very different language families.
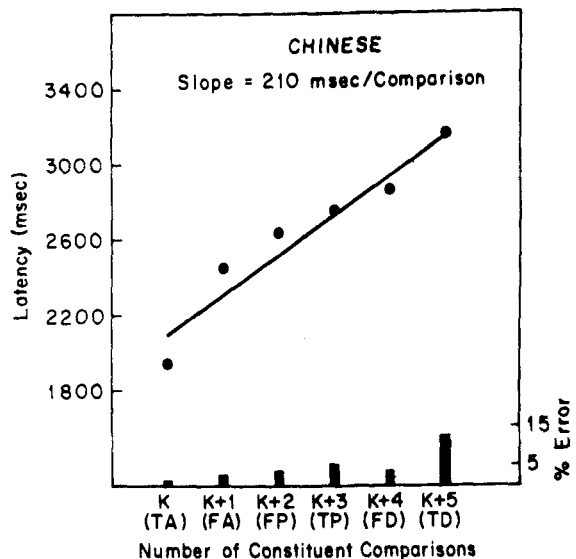

Figure 4. The fit of the model to the data for Chinese sentences.

Since the surface structure of the Chinese sentences was similar to that of the English sentences, it might be argued that the processing is determined by the surface structure order of the constituents. However, our past research has shown that varying the surface structure of English sentences (for example, by using sentences like *That the dots aren't red is true* and *That the dots are red isn't true*) does not alter the linearity of the latencies (Carpenter & Just, 1975). This suggests that the surface structure similarities are not responsible for the basic results. Rather, the two kinds of negative sentences have different internal representations that are somewhat independent of the surface structure. Thus, even if one could locate a language in which the two types of negatives were expressed very differently from English, the processing should remain the same. The next experiment investigated Norwegian which, although an Indo-European language like English, also differs slightly in how negation is expressed. The effect of both this surface structure variation and the effect of quantification was examined in Experiment II.

EXPERIMENT II

The purpose of Experiment II was to examine the verification of sentences in Norwegian, and additionally, to examine how quantifiers are processed in the context of complex sentences. This latter issue has not been examined in English or in any other language. The quantifiers referred to are either a large subset, i.e., *mange* (many), *14 av de 16* (14 of the 16), or to a small subset, i.e., *noen få* (a few, literally, "some few") and *2 av de 16* (2 of the 16). The quantifiers modified the subject of the subordinate clause, for example, *It's true that many of the dots are red.* The experiment examined how the processing differed for the two types of quantifiers.

As mentioned previously, there is a difference in the surface structure of negation in Norwegian and English. In English, the negative morpheme generally occurs after a copula verb, in both subordinate and superordinate clauses (e.g., *It is not true that the dots are red* and *It is true that the dots are not red*). In Norwegian, the negative morpheme, *ikke* occurs after the verb (*stemmer*) in a superordinate clause (*Det stemmer ikke at prikkene er røde*), but before the verb (*er*) in a subordinate clause (*Det stemmer at prikkene ikke er røde*). The position of the negative morpheme relative to the copula verb resembles the Chinese sentences rather than the English ones. However, the model deals with a matching process between semantic representations that do not reflect these surface structure differences. Thus, these differences should not affect the basic comparison processes in the verification of the sentences.

The representation of the affirmative sentence, *It is true that many of the dots are red*, may consist of a proposition where *(larger subset of dots)* is represented as the subject and *(red)* is the predicate, so the sentence would be represented *(AFF, (RED, LARGE SUBSET))*. For a predicate negative like *It's true that many of the dots aren't red* the postulated representation is *(NEG, (RED, LARGE SUBSET))*. For a denial like *It isn't true that many of the dots are red* the proposed representation is *(NEG, (AFF, (RED, LARGE SUBSET)))*.

The picture encoding process is postulated to be somewhat different now. In Experiment II, the picture contained two subsets of dots, a large subset of red dots and a small subset of black dots, or vice versa. It is presumed that the picture is coded so that the subject of the picture representation, *(LARGE SUBSET)* or *SMALL SUBSET)*, is the same as the subject (argument) of the sentence representation. For example, if the sentence contains the quantifier *many*, the large subset and its color would be encoded from the picture, and serve as the argument and predicate of the picture representation. If the picture contained a large subset of red dots, it would be represented *(RED, LARGE SUBSET)*.

The assumption that picture coding is contingent on the sentence has been independently confirmed in several ways. First of all, if the picture is presented before the sentence, then the verification latencies for quantified sentences change appropriately (Just & Carpenter, 1971) Secondly, if people are told how to code the picture, the verification latencies are consistent with this picture-coding hypothesis (Just & Carpenter, 1971). Thirdly, and perhaps most importantly, people scan the picture differently depending on whether the sentence contains *many* or *a few*. This was confirmed by monitoring eye fixations during the verification task. Carpenter and Just (1972) found that, following a quantifier like *many*, people fixated the larger subset, but after a quantifier like *a few* or *a minority* they fixated primarily on the smaller subset. Thus it is reasonable to assume that the picture is coded in terms of the quantifier from the sentence representation.

This contingent picture coding hypothesis leads to the prediction that the verification latencies should increase linearly from true affirmatives (k), to false affirmatives (k + 1), to false predicate negatives (k + 2), to true predicate negatives (k+ 3), to false denials (k + 4), to true denials (k + 5).

A second prediction is that the verification latencies for quantifiers referring to a small subset should be longer (by a constant amount) than those referring to a large subset. This prediction is an empirical one, based on previous studies (Just & Carpenter, 1971) which found longer verification latencies for simple sentences that had small subset quantifiers. The proposed
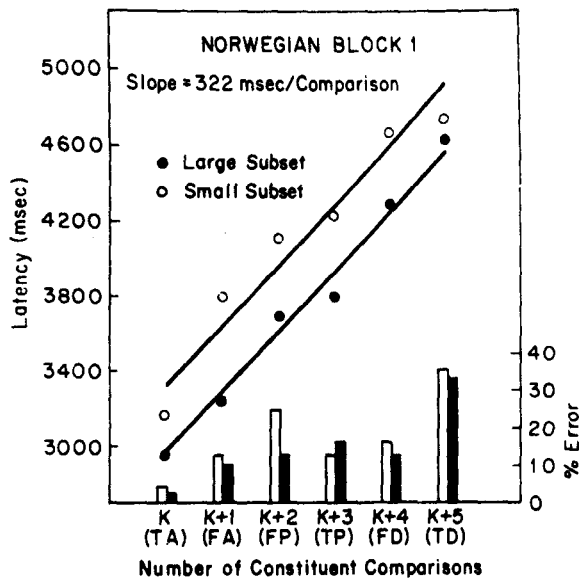
Figure 5. The fit of the model to the data for quantified Norwegian sentences (Block 1).

additivity of the two effects assumes that the advantage of large subset quantifiers occurs at some stage other than the comparison stage. Thus, the results for the two types of quantifiers in this experiment should show a difference in intercept, but not in slope, the time per comparison. Specifically, affirmative quantifiers referring to the small subset (e.g., *a few, 2 of the 16*) should show a greater intercept than sentences with affirmative quantifiers referring to the larger subset (e.g., *many, 14 of the 16*). This prediction concerns only affirmative quantifiers, which must be distinguished from other quantifiers like *few, hardly any*, and *scarcely any*. These latter quantifiers have different linguistic properties (cf. Klima, 1964) and are processed differently from the affirmative quantifiers studied in the current experiment (Carpenter & Just, 1972; Just & Carpenter, 1971).

Method

The method was basically similar to that in Experiment I except that the stimulus materials and instructions were in Norwegian. The Norwegian sentences have the superordinate *Det stemmer*, which translates as *It is so*, and quantifiers that denoted either a large subset (*many, 14 of the 16*) or a small subset (*a few, 2 of the 16*). Only the color terms *red* and *black* were used, and they interchanged roles in the sentences and pictures. Each sentence was paired with a picture of 14 red and 2 black dots, as well as with a picture of 14 black and 2 red dots. Each sentence-picture pair was presented the same number of times. In summary, the design included the six information conditions, quantifiers referring to two set sizes, two exemplars of each type of quantifier, and two color predicates, resulting in 48 different stimuli.

Each sentence subtended about 24 deg of visual angle, and was typed on an index card. The 4 by 4 array of dots, which was drawn with a felt pen to the right of the sentence, subtended 4 x 4 deg. The two different-colored dots always occupied the same

position ((3,3) and (4,4)) in the array. The stimulus card was viewed in a tachistoscope at a distance of 30 cm.

The subject pushed a switch to initiate a trial, and 1 sec later the stimulus card was presented for 3 sec. The subject was timed (in msec) from stimulus onset while he read the sentence, examined the dot array, and decided whether the sentence was true or false with respect to the picture. After the practice, all 48 cards were shuffled and the subject did 3 blocks of 48 stimuli. The whole session lasted about 35 min. The subjects were 12 Norwegian students at the University of Oslo.

Results (Block 1)

Previous research with English sentences (Carpenter & Just, 1975) indicated that the processing strategy for verifying predicate negative sentences sometimes spontaneously changes with practice. Inspection of the data indicated that a change occurred after the first block of trials, and so we will initially discuss only the results of the first block.

The latencies for erroneous responses were discarded, and the person's score for each condition was the mean of his correct latencies for that condition. The data were collapsed over sentences that differed only in their color predicates *red vs. black*. In addition, there were no systematic differences between the results for *many* and *14 of the 16* or between *a few* and *2 of the 16*. Therefore, the data for the two exemplars were combined.

As predicted, the mean latencies increased with the hypothesized number of constituent comparisons for both kinds of quantifiers. More precisely, latencies increased an average of 322 msec for each additional retrieval and comparison operation (standard error = 39 msec). Figure 5 shows that result, along with the best fitting straight lines for quantifiers referring to the small subset and those referring to the large subset. The model accounts for 96.2% of the variance among the 12 means $[F(2/121) = 55.11, p < .001]$. The residual 3.8% is not significant $[F(9/121) < 1]$. The root mean squared deviation (RMSD) is 114 msec, small relative to the 322-msec parameter. The standard errors for the 12 means ranged from 162 to 473 msec, generally increasing with the means. This analysis confirms the major hypothesis that verification time increases linearly with the number of constituent comparisons.

The 332-msec parameter is larger than the 200-msec parameter for English and 210-msec parameter for Chinese. However, the experiments are not entirely equivalent because the sentences for English and Chinese speakers did not contain quantifiers. Moreover, subjects in the English and Chinese experiments had seven or five times as much practice, and practice does decrease the slope (Carpenter & Just, 1975).

Quantifiers that referred to a large subset were verified an average of 354 msec faster than those that referred to a small subset. However, this difference just failed to reach significance at the .05 level $[F(1/11) = 4.65]$. As predicted, there was no difference

between the slopes for the two types of quantifiers [t(11) < 1].

The overall error rate in the first block of trials was 16.1%, and the errors were distributed across the 12 conditions as shown in Figure 5. There is a positive correlation between the mean latencies and probabilities of error in the 12 conditions (r = .80), indicating that a speed-accuracy tradeoff is unlikely.

RECODING STRATEGIES

With increased practice in the task, true predicate negatives are sometimes verified faster than false predicate negatives (Carpenter & Just, 1975). This result suggests that a change in the processing strategy occurs with practice. In particular, it is possible that people spontaneously start to recode predicate negative sentences into functionally equivalent affirmative forms before comparing them to the picture representation. For example, *It's true that a few of the dots aren't red* might be recoded as *It's true that a few of the dots are black.* For unrecoded sentences, the negative marker will mismatch the corresponding affirmative marker in the picture representation. By contrast, a recoded negative sentence will match. Thus, recoding removes a mismatch between polarity markers of the sentence and picture representation. The data for Blocks 2 and 3 showed evidence of such recoding, and so they were analyzed separately.

The occurrence of recoding can be inferred when true negatives are verified faster than false negatives. The evidence that supports this inference is twofold. First, subjects sometimes spontaneously report using a recoding process, and in these cases their latencies usually show the characteristic recoding latency pattern (Wason, 1961). Secondly, subjects can be explicitly instructed when and how to recode negative sentences, and in these circumstances their latencies exhibit the characteristic recoding latency pattern (Young & Chase, Reference Note 1). Certain experimental conditions, such as extended practice, binary alternatives (e.g., open-closed), a delay between the sentence and the picture (cf. Trabasso, 1972; Carpenter, 1973; Carpenter & Just, 1975), increase the probability that subjects will spontaneously recode negatives. However, none of these conditions alone seems to be necessary or sufficient. Moreover, it is not clear why predicate negatives are recoded while denials are not. However, given that recoding has occurred, the original model can account for how long recoding will take and how long it will take to compare the recoded representation.

The data for English-speaking subjects showed that recoding time systematically adds to comparison time. The results suggested that operations involved in recoding a negative sentence resemble the operations involved in retrieving and comparing constituents during

the comparison process. According to such a recoding model, the constituents are serially retrieved, and replaced by other symbols, e.g., *(NEG, (RED, LARGE SUBSET))* by *(AFF,(BLACK,LARGE SUBSET))*. The model is based on the same "find and compare" operation found in the primary verification model.

The assumptions underlying the recoding model are presented elsewhere (Carpenter & Just, 1975), so we will briefly work through the recoding process and summarize the predictions. Suppose a person has read a true predicate negative sentence like *It's true that many of the dots aren't red*, represented *(NEG, (RED, LARGE SUBSET))*, and examined a confirming picture of mostly black dots. During the recoding process, the inner proposition of the sentence representation, *(RED, LARGE SUBSET)* would be replaced with *(BLACK, LARGE SUBSET)*. Then the embedding negation marker would be replaced with an affirmation marker, for a total of two "find and compare" operations during the recoding. After recoding, the sentence representation is identical to the picture representation, and so only requires k constituent comparisons. Thus, the total verification time for a true predicate negative should involve k + 2 operations. Recoding a false predicate negative should be very similar, involving 2 "find and compare" operations. Then during the comparison process, there would be a mismatch on the color predicate, resulting in k + 1 comparisons, and a grand total of k + 3 operations. Thus, if a recoding strategy is used, true predicate negatives should be verified faster than false predicate negatives.

The pattern of verification latencies for affirmatives and denials does not change with practice, so the predictions for affirmatives and denials remain as before. In summary, the number of operations necessary to verify a true affirmative is k; for a false affirmative, k + 1; for a true predicate negative, k + 2; for a false predicate negative, k + 3; for a false denial, k + 4; and for a true denial, k + 5. So when recoding takes place, latencies should still increase with the number of operations.

Results (Blocks 2 and 3)

The data for blocks 2 and 3 were collapsed and treated similarly to the data in Block 1. The latencies for erroneous responses were discarded and the score for each condition was the mean of the correct latencies for that condition. The mean latencies increased with the number of constituent operations for both kinds of quantifiers. More precisely, latencies increased an average of 278 msec for each additional constituent operation (standard error = 51 msec). Figure 6 shows this result, along with the best fitting pair of straight lines. Notice that the abscissa of the graph has been changed. This axis still represents an increasing number of operations. However, true predicate negatives now are
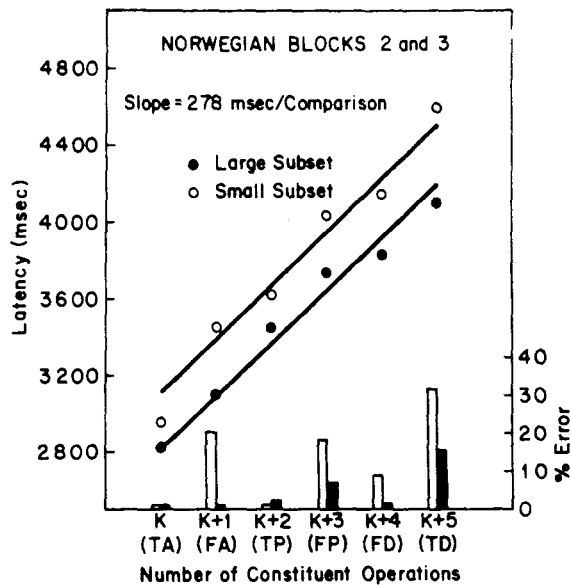
Figure 6. The fit of the model to the data for quantified Norwegian sentences (Blocks 2 and 3).

hypothesized to involve fewer constituent comparisons than false predicate negatives. The model accounts for 96.5% of the variance among the 12 means [F(2/121) = 51.26,p < .001]. The residual 3.5% is not significant [F(9/121) < 1]. The RMSD of 95 msec is small relative to the 278-msec parameter. The standard errors for the 12 means ranged from 207 to 406 msec, generally increasing with the means. The analysis indicates that the data support the model.

In Blocks 2 and 3, true predicate negatives were verified faster than false predicate negatives. This is consistent with the hypothesis that most people had spontaneously adopted a recoding strategy in Blocks 2 and 3. A methodological difference between Experiments I and II may explain why the Norwegian predicate negative sentences were recoded after some practice but the English and Chinese sentences were not. In the Norwegian experiment, only two colors, red and black, were used. Thus, *not red* was functionally equivalent to *black.* By contrast, in the English and Chinese studies there were three colors, red, black, and green, so that *not red,* for example, was not functionally equivalent to *black.* Recoding is more likely to occur if the alternatives are binary (cf. Trabasso, 1972), and the occurrence of recoding only in the Norwegian experiment is consistent with Trabasso's observation.

The slope diminishes somewhat with practice from 322 msec in Block 1 to 278 msec in Blocks 2 and 3 [t(11) = 1.03, n.s.] but the linearity is preserved. This result was also true of a number of other studies reviewed and reported elsewhere (Carpenter & Just, 1975).

As predicted, the quantifiers that referred to a large

subset were verified an average of 300 msec faster than those that referred to a small subset [F(1/11) = 7.67, p < .05]. One possible explanation for this effect is that quantifiers that refer to a small subset are linguistically more complex (marked) and so they take longer to read and represent. Such a linguistic complexity argument might be applied to the *many-a few* pair, but it is difficult to see how *2 of the 16* is linguistically more complex than *14 of the 16.* However, it is possible that the internal representation of small subset quantifiers is in some way more complex than the representation of large subset quantifiers. Thus, the process of deriving the sentence representation may account for the verification time difference. Alternatively, the difference may arise in the construction of the picture representation (cf. Just & Carpenter, 1975). Since the sentence and picture are hypothesized to have the same quantifier represented, the extra time for a small subset quantifier, if there is an extra time, could arise in the construction of either representation. Finally, the extra time could also arise in picture scanning and encoding. It may take longer to visually locate the smaller subset, by virtue of its lesser perceptual prominence. Whatever the locus of this quantifier effect, it is clear that it does not interact with the main comparison process. The two effects are statistically independent: as predicted, there was no difference between the slopes for the two types of quantifiers [t(11) = 1.16, n.s.]. Inspection of Figures 5 and 6 confirms this statistical conclusion. The two effects seem to occur at different stages in the processing.

The overall error rate in Blocks 2 and 3 was 9%. As the distribution of errors shown in Figure 6 indicates, there were more errors for the quantifiers that referred to a small subset. Some of these errors may have been due to an ambiguity in the sentence. For example, if 14 of the 16 dots in the picture are red, and the sentence says *It's true that 2 of the 16 dots are red,* then the sentence can be judged to be true or false, depending on whether the quantifier is interpreted as *at least 2 of the 16* or *only 2 of the 16.* Two subjects seem to have interpreted the quantifiers referring to the smaller subset to mean *at least 2 of the 16...* and *at least a few...* because they accounted for two thirds of the "erroneous" responses in the false affirmative condition for quantifiers referring to the small subset. Just to be sure that these two subjects did not distort the latency results, the latencies were analyzed for the small subset quantifiers for the 10 remaining subjects, and the results and conclusions remained intact.

The pattern of "errors" for these two subjects lends independent support to the recoding model. The "erroneous" interpretation of the small subset quantifiers should lead to errors in false affirmative cases (*At least 2 of the 16 dots are red* paired with 14 red and 2 black dots). If predicate negatives are recoded into an affirmative form then they should exhibit the same error pattern as affirmatives, while denials may exhibit a

different error pattern. The data confirmed these predictions. These two subjects have a 70% error rate for false affirmatives, and a 63% error rate for false predicate negatives. The true conditions show no errors. By contrast, they have only a 4% error rate for false denials, but 63% for true denials. Thus the error data for these subjects are consistent with the hypothesis that predicate negatives were recoded but denials were not.

GENERAL DISCUSSION

Sentence verification latencies in English, Chinese, and Norwegian all conformed to a single processing model, suggesting at least a limited universality of the underlying mental operations. The results indicate that the control structure governing sentence verification is the same in the three languages examined. Moreover, the processing rates were within the same range in all three studies. The time per operation was very similar in English and Chinese (200 and 210 msec, respectively), and slightly longer for Norwegian (322 msec in Block 1, 278 msec in Blocks 2 and 3) where there was less practice and the sentences and pictures were more complex. So both the control structure and the speed of the underlying mental operations are similar across the three languages.

The ubiquity of the find and compare operation in sentence verification is not surprising, given that very similar operations occur in other cognitive tasks. For example, Sternberg's (1969) model of the context recall task is based on a scan and compare operation that resembles the find and compare operation of the constituent comparison model in several respects. In the context recall task, a subject memorizes a list of digits, then is given a probe digit from the list, and his task is to report the digit that followed the probe. The response latencies suggested there was a serial, iterative scanning-to-locate process, which consisted of comparing the probe digit to each digit in the memorized list until a match was found. The duration of the scan and compare operation was estimated at 240 msec, comfortably close to the estimates in the present study. Thus, these fundamental operations appear in different tasks and in different languages. Such commonalities, limited as they currently are, point the way toward an explanation of language universals through the discovery of processing universals.

Considerable research has examined certain aspects of human cognition that are "culture free". For example, psychological studies have documented the existence of cross-language similarities in affective meaning (Osgood, 1964), metaphor (Asch, 1961), phonetic symbolism (cf. Brown, Black, & Horowitz, 1955), color terms (Berlin & Kay, 1969), and the acquisition of semantic structures (Slobin, 1970). More recently, there have been investigations of similarities in the way different languages are processed, such as the current study and

Slobin's (1971) investigation of the strategies employed by children during acquisition. A separate tradition of research has examined linguistic properties common to all languages (cf. Greenberg, 1966). Additionally, there have been some examinations of how these observed linguistic properties relate to universals of the physical environment (Bierwisch, 1967; Clark, Carpenter, & Just, 1973). Ultimately, cognitive psychology must look beyond linguistic universals, and search for processing universals. To this end, cross-language studies seem to be both a necessary and fruitful approach.

REFERENCES

Asch, S. E. The metaphor: A psychological inquiry. In M. Henle (Ed.), Documents of gestalt psychology. Berkeley: University of California Press, 1961.

Berlin, B., & Kay, P. Basic color terms. Berkeley: University of California Press, 1969.

Bierwisch, M. Some semantic universals of German adjectivals. Foundations of Language, 1967, 3, 1-36.

Brown, R., Black, A. H., & Horowitz, A. E. Phonetic symbolism in natural language. Journal of Abnormal and Social Psychology, 1955, 50, 388-393.

Carpenter, P. A. Extracting information from counterfactual clauses. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 512-521.

Carpenter, P. A., & Just, M. A. Semantic control of eye movements during picture scanning in a sentence-picture verification task. Perception & Psychophysics, 1972, 12, 61-64.

Carpenter, P. A., & Just, M. A. Sentence comprehension: A psycholinguistic processing model of verification. Psychological Review, 1975, 82, 45-73.

Clark, H. H., Carpenter, P. A., & Just, M. A. On the meeting of semantics and perception. In W. G. Chase (Ed.), Visual information processing. New York: Academic Press, 1973.

Greenberg, J. H. Language universals. The Hague: Mouton, 1966.

Jackendoff, R. S. An interpretive theory of negation. Foundations of Language, 1969, 5, 218-241.

Just, M. A., & Carpenter, P. A. Comprehension of negation with quantification. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 244-253.

Just, M. A., & Carpenter, P. A. The semantics of locative information in pictures and mental images. British Journal of Psychology, 1975, in press.

Klima, E. S. Negation in English. In J. A. Fodor & J. J. Katz (Eds.), The structure of language. Englewood Cliffs, N. J.: Prentice-Hall, 1964.

Lenneberg, E. H. Biological foundations of language. New York: Wiley, 1967.

Osgood, C. E. Semantic differential technique in the comparative study of cultures. American Anthropologist, 1964, 66, 171-200.

Slobin, D. I. Universals of grammatical development in children. In G. B. Flores d'Arcais & W. J. M. Levelt (Eds.), Advances in psycholinguistics. London: North-Holland Publishing Co., 1970.

Slobin, D. I. Developmental psycholinguistics. In W. O. Dingwall (Ed.), A survey of linguistic science. Maryland: University of Maryland Press, 1971.

Sternberg, S. Memory scanning: Mental processes revealed by reaction-time experiments. American Scientist, 1969, 57, 421-457.

Trabasso, T. Mental operations in language comprehension. In J. B. Carroll & R. O. Freedle (Eds.), Language comprehension and the acquisition of knowledge. Washington, D. C.: V. H. Winston & Sons, 1972.

Wason, P. C. Response to affirmative and negative binary statements. British Journal of Psychology, 1961, 52, 133-142.