COMPLEXITY METRICS AND PRODUCTION OF COMPLEX SENTENCES[1]

Hintat Cheung and Susan Kemper

University of Kansas

Psycholinguists have frequently attempted to formulate ways of measuring the complexity of different sentences (Crain & Schankweiler, 1988; Ford, 1983; Frazier, 1988; Smith, 1988; Watt, 1970). Complexity metrics have been theoretically derived from specific linguistic theories, experimentally devised form models of syntactic processing, and empirically developed from research on sentence processing. Complexity metrics are important research tools since they enable researchers to, for example, order experimental stimuli from least to most complex, examine developmental trends in children's mastery of complex construction, or make cross-linguistic comparison as to the relative complexity of grammatical constructions in different languages.

The most widely-known attempt to develop a complexity metrics was labeled by Fodor, Bever and Garrett (1974) as the Derivational Theory of Complexity (the DTC). The DTC tried to equate the complexity of sentences with the number of transformations, required in the then-current model of generative transformational grammar, intervening between the sentence's deep structure and its surface structure. Experimental findings which indicated that not all transformations increase processing complexity (see Fodor, Bever & Garrett for a review) and theoretical changes in syntactic theory (Bresnan, 1982) lead to the abandonment of the DTC.

Nonetheless, a variety of approaches to measuring syntactic complexity have been taken since the abandonment of the DTC. The most common approach does not postulate a general complexity metric but contrast children's or adult's processing of alternative grammatical construction (see for example, Clancy, Lee, & Zoh; Frazier and Fodor, 1978; Shapiro, Zurif, & Grimshaw, 1987; Smith & van Kleeck, 1986). Other researchers have attempted to develop metrics which will generally apply to sentences in order to scale the sentences as to their relative difficulty for production or comprehension. Some, like mean length of utterance (MLU) (Brown, 1973) and mean clauses per utterance (Kemper, Kynette, Rash, Sprott, & O'Brien 1989) measure sentence length. Others, like the formula developed by Botel and Granowsky (1972), assume a priori that grammatical construction differ in there relative difficulty. A final class of metrics examine structural aspects of sentences and attempt to quantify the processing demands of various sentence structures (Yngve, 1960; Frazier 1985).

The present study was to investigate the utility of different complexity metrics and it was undertaken as part of a study of age-group differences in adults' language. Kemper (1988) has suggested that there is an age-related decline in the complexity of adult's language. Previous researches by Susan Kemper and her co-workers (Kynette & Kemper, 1986) have established that elderly adults in their 70s and 80s are less likely than young adults to produce sentences with multiple embedded clauses, especially left branching clauses.

This asymmetry in the production of left- and right-branching sentences is linked to a similar asymmetry in comprehension; elderly adults have more difficulty recalling (Kemper, 1987b) and imitating (Kemper, 1986) left-branching sentences. Left branching sentences are presumed to be more difficult to process (Fodor, Bever, and Garrett, 1974) because they impose more demand on working memory to retain and manipulate grammatical constituents than do right-branching sentences.

As evidence for this linkage between working memory and the production of embedded sentences, Kemper and Rash (1988) computed the Yngve depth (Yngve, 1960) of a sample of adults' sentences. Yngve (1960) assumed that the production of a sentence imposed demands on a limited capacity working memory in order to retain planned but not yet articulated grammatical constituents. The depth of any word in a sentence represents how many planned grammatical constituents have not yet been realized during the left-to-right production of the sentence. In general, sentence embedding, particular left-branching embeddings, increase the Yngve depth of sentences since words within the embedded sentence are at greater depth than words in the main clause.

Kemper and Rash (1988) showed that Yngve depth declines with the age of the speaker. Kemper and Rash (1988) also found that Yngve depth is correlated with adults' backward digit span (Wechsler, 1958) while Kemper et al (1989) found that adult's backward digit is correlated with MCU and the production of left-branching clauses; adult's with larger backward digit spans produce sentences with more embedded clauses, particularly left-branching clauses, and greater Yngve depth. This finding implies that the age-related decline in adults' production of complex sentences, particularly left-branching sentences, is due to age-related declines in the capacity of working memory, as measured by backward digit span.

Frazier (1985) has challenged Yngve as a valid measure of syntactic complexity and suggested that an alternate metric which was explicit motivated by consideration of the complexity of sentence processing operation. Frazier's count differs from Yngve depth in two ways: first, sentence embeddings are explicit acknowledge as sources of complexity and, hence, increase the complexity of a particular sentence; second, the complexity is computed over three-word sequences such that a cluster of many processing decisions contributes more to the complexity of a sentence than a distributed sequence of processing decisions.

The following experiment was undertaken in order to empirically compare Yngve depth to other complexity metrics. The measures included: MLU, traditionally used in the child language literature to measure linguistic development (Miller and Chapman, 1981), MCU, developed by Kemper et al (1989) to measure adults' linguistic development, the Botel & Granowsky (1972) formula developed as an alternative to readability formulas to measure the difficult of texts, two alternative ways of measuring Yngve depth, and two variants of Frazier's count.

Language samples. The language sample were taken from the oral narratives analyzed by Kemper et al (1989). Five narratives were randomly selected from each group so that there were five narratives from college students ages 17-24 years of age, five from adults aged 60-69 years, five from adults aged 70-79 years, and five from adults aged 80-90 years. Each narrative was told by a native speaker of English. The age of each speaker and each speaker's forward and backward digit span from the Wechsler Adult Intelligence Scales (Wechsler, 1958) were available. For this small sample, age correlated $r(20) = -.61$ with backward digit span and $r(20) = -.39$ with forward digit span. The two digit span measures were correlated $r(20) = +.91$.

Analyzes. A total of 100 sentences was analyzed. Five consecutive sentences were selected from the middle portion of each narrative to yield a sample of continuous speech. Only complete sentences were selected for the analysis. Seven different complexity measures were then obtained for each sentence:

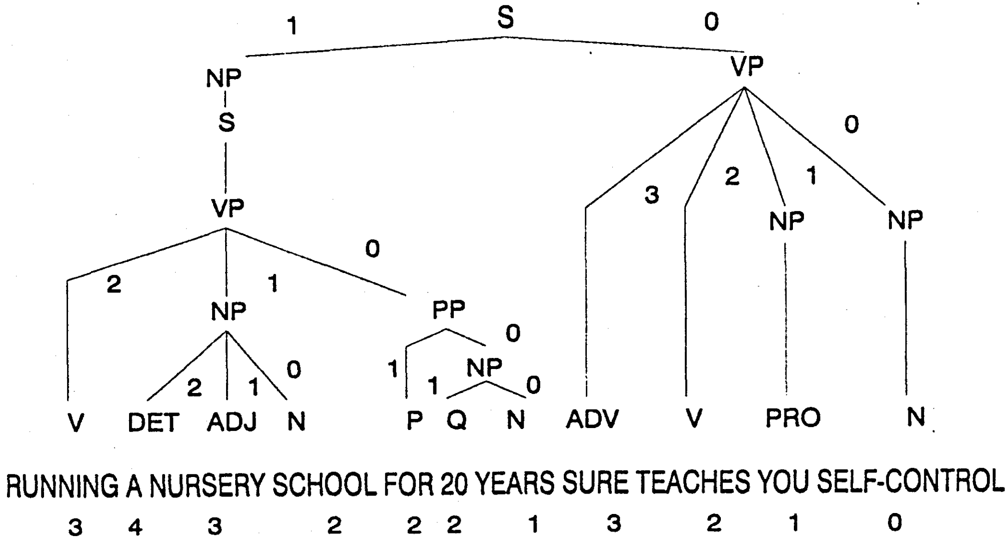
MLU: The number of words per sentence was determined and the mean length of each speaker's sentence was calculated.

MCU: The number of syntactic clauses per sentence was determined by counting each main clause and each embedded or subordinate clause. The mean number of clauses per utterance was calculated for each speaker.

Directional Complexity: The rules given by Botel and Granowsky (1972) were applied to each sentence to determine directional complexity. These rules assigned 0, 1, 2, 3 points to various sentence patterns and structures: 0-point structures include subject-verb, subject-verb-object, and subject-verb-infinitive constructions, interrogative sentences, and coordinate clauses joined by *and;* 1-point structure include sentence with both direct and indirect objects, noun modifiers such as adjectives and possessives, adverbials, coordinate clauses joined by *but*, *or* etc., gerunds used as subjects, and infinitive complements to subject-verb-object clauses; 2-point structures include comparatives, subordinate clauses, infinitives used as subjects, and passives; 3-point structures include wh- and that- clauses used as subjects. The average directional complexity of each speaker's utterances was then calculated.

Yngve depth: Both the total Yngve depth and the maximum Yngve depth of each sentence was determined according to the procedures given by Yngve (1960). Figure 1 illustrates the calculation of the Yngve depth. Yngve depth was determined by first performing a surface phrase structure analysis of the sentence to construct a syntactic tree with nodes and branches and then numbering the branches below each node from right to left starting with zero. The depth of each word was the sum of all the branches connecting the word to the root or top-most node of the sentence.

Figure 1: Calculation of Maximal and Total Yngve Depth.



Maximal depth = 4

total depth = 23

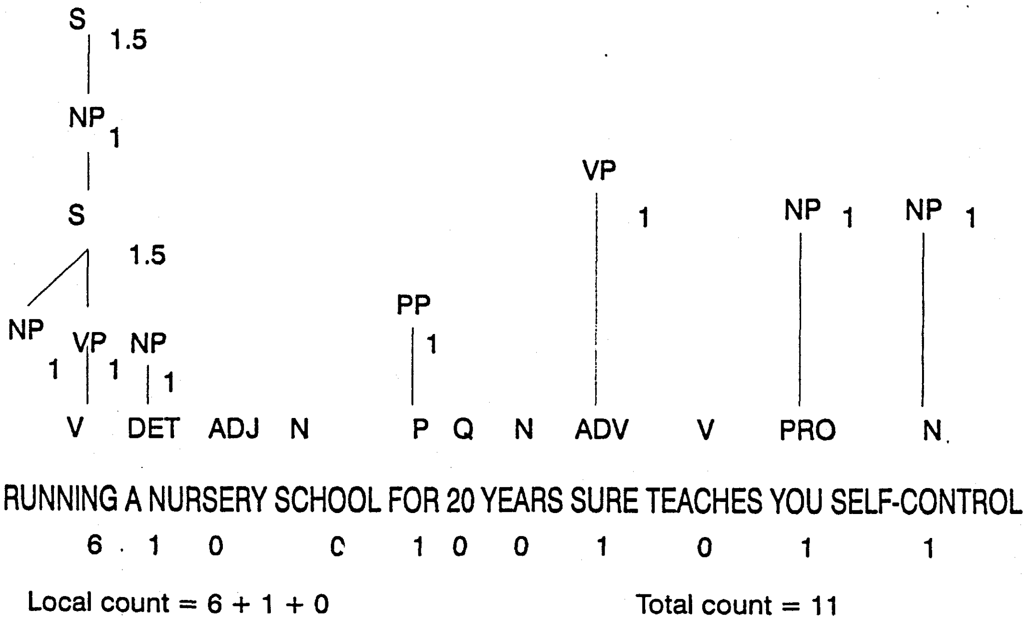
Two Yngve depth measures were determined for each sentence: (a) maximal Yngve depth is the largest number associated with any word in the sentence and (b) total Yngve depth is the sum of all depth counts for each word in the sentence. Maximal Yngve depth was, therefore, a "local " measure which was independent of sentence length; total Yngve depth was confounded with the number of words in the sentence. The average maximum Yngve depth and the average total Yngve depth were computed for each speaker.

Frazier count: Two measures, local node count and total count, were derived from the rules given by Frazier (1985). Figure 2 illustrates the calculation of the local Frazier and total Frazier counts. The Frazier counts were based on a surface phrase structure analysis in which all (nonterminal) nodes in the phrase structure of the sentence were assigned a point value of 1 except for sentence nodes and sentence-complement nodes which were assigned a point value of 1.5. Counts for each word were then determined by summing up the points assigned to all of the nodes dominating each word in the sentence.

As implied by the analyzes given in Frazier (1985), nodes in the phrase structure of a sentence were counted if the sentence was being parsed from left to right in a deterministic manner, as in the parser developed by Marcus (1980) and discussed by Berwick and Weinberg (1984). Consequently, nodes were assigned to, e.g., possessive markers and deleted noun phrases which introduce new syntactic constituents or which are required in order to connect each new word to the preceding structure. For example, in Figure 2, a gerund is used as the subject of the main clause and the entire gerund is treated as a noun phrase. The gerund is given 1 point as a noun phrase, 1.5 point for the embedded S node since the gerund is the subject of the main sentence.

Two variants of the Frazier counts were computed. The local Frazier count was determined by summing the node points for each sequence of three adjacent words and identifying the largest such sum in the sentence. This three-word window is assumed to reflect the capacity of the parser to hold partially analyzed constituents (Marcus, 1980). The total Frazier count was determined by summing all nodes points for all of the words in each sentence. The local Frazier count, therefore, reflects the concentration of grammatical constituents; the total Frazier count was confounded

Figure 2: Calculation of Frazier Local and Total Counts.



with the length of the sentence. Average local Frazier and total Frazier counts were obtained for each speaker.

Results. The seven complexity measures were compared by performing a series of one-way ANOVA with age group of the speaker as the between-subjects factor. Each measure produced a significant age effect, as listed in Table 1 and the linear component of the age effect was significant in each case.

Table 1. Mean complexity and ANOVA results for the language samples from college students and adults in their 60s, 70s and 80s.

	College	60s	70s	80s	Age *F*(3,16)	Linear *F*(1,16)
MLU	16.1	12.0	11.3	8.0	5.83**	16.41**
MCU	1.9	1.6	1.5	1.2	4.92*	5.59*
Directional	2.6	1.6	1.3	1.2	2.89*	7.13*
Total Yngve	28.0	18.4	15.5	8.8	7.20**	20.87**
Maximal Yngve	3.4	2.6	2.8	2.2	4.98*	11.61**
Local Frazier	14.9	12.2	10.8	7.3	4.04*	11.83**
Total Frazier	5.4	5.1	4.8	4.6	3.69*	10.79**

* $p < .05$, ** $p < .01$

Table 2 presents the matrix of correlations produced by the seven complexity measures as well as the speakers' age and digit spans. All the complexity measures were significantly correlated; the speakers' age was negatively correlated with MCU. Older speakers produced sentences with fewer embedded clauses, thus lowering MCU. Backward digit span was positively correlated with the maximal Yngve depth, maximal Frazier count, and MCU and forward digit span was correlated with maximal Yngve depth and MCU; speakers with greater digit spans produced sentences which contained more clauses which were more complex as measured by maximal Yngve depth and local Frazier count.

Table 2. Matrix of correlations among the complexity measures and the speakers' age and digit spans.

	MLU	MCU	Direction	Max Yngve	Total Yngve	Local Fraz.	Total Fraz.
MLU	-						
MCU	.82	-					
Directional	.73	.67	-				
Maximal Yngve	.87	.75	.64	-			
Total Yngve	.96	.83	.75	.87	-		
Local Frazier	.74	.78	.63	.64	.70	-	
Total Frazier	.95	.89	.68	.84	.95	.76	-
Age	-.19	-.53	-.07	-.34	-.27	-.29	-.27
Digits Backwards	.42	.55	.50	.63	.49	.53	.51
Digits Forwards	.40	.52	.48	.56	.44	.50	.51

Stepwise regression was then used to determine whether the speakers' age and forward and backward digit span accurately predicted the complexity of their sentences. In each case but one, a significant linear regression equation was obtained with backward digit span as the sole predictor of complexity. Adding the speaker's age and forward digit span did not improve the fits of these regression equations. MLU was the exception; neither age, backward digit span, forward digit span nor any combination of these three predictors produced a significant regression equation for MLU. Table 3 summarizes these results. Backward digit span accounted for between 27% and 39% of the variance in these complexity measures with the exception of MLU.

Table 3. Results of regressing the complexity measures on the speakers' backward digit span.

	R	R^2	$F(1,18)$
MCU	.55	.31	7.97
Directional	.50	.25	6.08
Maximal Yngve	.63	.39	11.57
Total Yngve	.49	.24	5.79
Local Frazier	.52	.27	6.89
Total Frazier	.52	.27	6.5

Discussion

Despite computational differences, all seven complexity measures yielded the same relative ordering of the twenty speech samples. Age-related decline in complexity emerged regardless of which complexity metric was used to scale the adults' language samples. The decline is, therefore, due, in part, to a reduction in the length of the adults' utterances as measured by MLU and MCU and, in part, due to a loss of grammatical structures which contribute to the Botel and Granowsky (1972) Directional Complexity formula, Yngve depth, and the Frazier count. The correlational analyses suggest that the age-related decline occurs because of a loss of sentence embeddings, particular left-branching embeddings, due to limitations of working memory. Sentence embeddings impose demands on working memory for the simultaneous construction and manipulation of multiple syntactic constituents; hence, backward digit span, a measure of the capacity of working memory, is significantly correlated with MCU, maximal Yngve depth, and the local Frazier count which are sensitive to the production of multi-clause sentences. The capacity of working memory appears to decline with advancing age (Baddeley, 1986); consequently, adults become less able to construct complex syntactic structures with embedded gerunds, that-clause, wh-clause and infinitives. This will lower MCU, maximal Yngve depth, and the local Frazier count.

In looking for the determinants of sentence processing difficulties, psycholinguists have identified many contributing syntactic factors either by systematically contrasting sentences with different syntactic properties or by developing formulae for ordering sentence as to their overall complexity. The choice of a complexity metric for research purpose will depend upon practical considerations. For most language samples, MLU and MCU can be easily computed; however, MLU, while widely used to scale children's language acquisition, shows little variation over the adult years and may not be sensitive to developmental difference once the basics of morphology and syntax have been mastered (Kemper et al, 1989; Klee & Fitzgerald, 1985). MCU has limited utility for the study of the early stages of language acquisition since young children do not begin to master the syntax of embedding until rather late in the acquisition period (Limber, 1973).

While MLU and MCU can be computed with some ease, the other complexity metrics require skilled analysis for their application. The Directional Complexity formula of Botel & Granowsky (1972) requires that the researcher carefully examine each sentence for a wide range of different syntactic constructions and assign appropriate points values to these constructions. The Yngve and Frazier analyses require that the research perform a surface phrase structure analysis of the sentence. For language samples of e.g. 50 utterances, these analyses can be time consuming and their accuracy must be verified by another skilled syntactican. The Frazier analysis is more difficult to execute than the Yngve analyses since it attempts to emulate a deterministic, left-to-right parser. The analysis must detect and fill in gaps in the structure of the sentence whenever noun phrases have been deleted, or fronted. Therefore, MLU and MCU seem to be the more ready to-use measurement for studies involve a large amount of speech samples; Yngve depth, Frazier count, Directional Complexity can provide a better index in controlling levels of structural complexity as they are sensitive to smaller structures. In sum, the five different syntactic complexity measures we have examined are sensitive to age-group difference and the decision of using which of those should be made according to practical factors.

Note

[1]This research was supported by grant R01AG06319 from the National Institute of Aging to S. Kemper

References

Baddeley, A. 1986. Working memory. Oxford: Oxford University Press.

Berwick, R. C. & A. C. Weinberg. 1984. The grammatical basis of linguistic performance: Language use and acquisition. Cambridge, MA: MIT Press.

Botel, M. & A. Granowsky. 1972. A formula for measuring syntactic complexity: a directional effect. Ele. Ed. 49.513-516.

Bresnan, J. (Ed.), 1982. The mental representation of grammatical relations. Cambridge, MA: MIT Press.

Brown, R. 1973. A first language. Cambridge, MA: Harvard University Press.

Clancy, P.M., H. Lee & M. H. Zoh. 1986. Processing strategies in the acquisition of relative clauses: Universal principles and language specific realizations. Cognition. 24.225-262.

Crain, S. & Schankweiler, D. 1988. Syntactic complexity and reading acquisition. Critical approaches to readability (pp. 167-192) ed. by A. Davison & G. Green.. Hillsdale, NJ: Erlbaum Associates.

Fay, D. 1980. Performing transformations. Perception and production of fluent speech (pp.441-469) ed. by R. L. Cole. Hillsdale, NJ: Erlbaum Associates.

Fodor, J. A., T. G. Bever, & M. Garrett. 1974. The psychology of language. New York: McGraw-Hill.

Ford, M. 1983. A method for obtaining measures of local parsing complexity throughout sentences. J.Verb.Learn.Verb.Beh. 22.203-218.

Frazier, L. 1985. Syntactic complexity. Natural language parsing: Psychological, computation, and theoretical perspectives (pp. 129-189) ed. by D. R. Dowty, L. Karttunen, & A. M. Zwicky. Cambridge: Cambridge University Press.

Frazier, L. 1988. The study of linguistic complexity. Critical approaches to readability (pp. 167-192) ed. by A. Davison & G. Green. Hillsdale, NJ: Erlbaum Associates.

Frazier, L. & J. D. Fodor. 1978. The sausage machine: a new two-stage parsing model. Cognition. 6.291-325.

Jackenoff, R. 1977. X-Bar syntax: a study of phrase structure. Cambridge, MA: MIT Press.

Limber, J. (1973). The genesis of complex ideas. Cognitive psychology and the acquisition of language ed. by T. Moore. New York: Academic Press.

Kemper, S. 1986. Imitation of complex syntactic constructions by elderly adults. App.Psychol. 7.277-287.

Kemper, S. 1987a. Life span changes in syntactic complexity. J.Ger. 42.323-328.

Kemper, S. 1987b. Syntactic complexity and the recall of prose by middle-aged and elderly adults. Exp.Ag.Resc. 13. 47-52.

Kemper, S. 1988. Geriatric psycholinguistics: Syntactic limitations of oral and written language. Language and memory in old age (pp.58-76) ed. by L. Light & D. Burke. Cambridge, MA: Cambridge University Press.

Kemper, S., D. Kynette, S. Rash, R. Sprott, & K. O'Brien. 1989. Life-span change to adults' language: Effect of memory and genre. App.Psychol. 10.49-66.

Kemper, S. & S. Rash. 1988. Speech and writing across the life-span. Practical aspects of memory: Current research and issues (pp.107-112) ed. by M. Gruneberg. P. E. Morris, & R. N. Sykes. Chichester: Wiley.
Klee, T. & M. D. Fitzgerald. 1985. The relation between grammatical development and mean length of utterance in morphemes. J.Chld.Lng. 12.251-269.
Kynette, D. & S. Kemper. 1986. Aging and the loss of grammatical forms: a cross-sectional study of language performance. Lng.Comm. 6.43
Marcus, M. 1980. A theory of syntactic phenomena of English, Vol.1 and 2. Chicago: University of Chicago Press.
Miller, J. F., & R. S. Chapman, R. S. 1981. The relation between age and mean length of utterance in morphemes. J.Sp.Hear.Resc. 24.154-161.
Shapiro, L. P., E. Zurif, & J. Grimshaw. 1987. Sentence processing and the mental representation of verbs. Cognition. 27.219-246.
Smith, C. S. 1988. Factors of linguistic complexity and performance. Critical approaches to readability (pp. 167-192) ed. by A. Davison & G. Green. Hillsdale, NJ: Erlbaum Associates.
Smith, C. S. & A. Van Kleeck. 1986. Linguistic complexity and performance. J.Chld.Lng. 13.389-408.
Yngve, V. 1960. A model of a hypothesis for language structure. Proc.Am.Phil.Soc. 104. 444-466.
Watt, W.C. 1970. On two hypotheses concerning psycholinguistics. Cognition and the development of language ed. by J. R. Hayes. New York: Academic Press.
Wechsler, D. 1958. The Measurement and Appraisal of Adult Intelligence. Baltimore: Williams & Wilkins.