The University of Georgia

ARTIFICIAL INTELLIGENCE CENTER

CASPR Research Report 2006-01

HOW COMPLEX IS THAT SENTENCE?
A PROPOSED REVISION OF THE
ROSENBERG AND ABBEDUTO D-LEVEL SCALE

Michael A. Covington, Congzhou He,
Cati Brown, Lorina Naçi, and John Brown

caspr

COMPUTER ANALYSIS OF SPEECH
FOR PSYCHOLOGICAL RESEARCH

How Complex Is That Sentence?

A Proposed Revision of the

Rosenberg and Abbeduto D-Level Scale

MICHAEL A. COVINGTON, CONGZHOU HE, and CATI BROWN

University of Georgia

LORINA NAÇI

Cambridge University

JOHN BROWN

GlaxoSmithKline Plc

ABSTRACT

We present a revision of the D-Level sentence complexity scale of Rosenberg and Abbeduto (1987). The original scale left some sentences unrated. However, some types of analysis require finding the average D-Level of the sentences in a text. Accordingly, we have generalized the scale so that every sentence receives a score.

Further, on the basis of recent psycholinguistic research, we have re-categorized sentences with semi-auxiliaries (e.g., *be gonna*) and raising verbs (*seem*). We have also swapped Rosenberg and Abbeduto's levels 5 and 6, since experimental results show that these levels were originally in the wrong order.

Our proposals are motivated by linguistic structure and psycholinguistic data. A brief experiment suggests that the revised scale has better reliability than the original scale.

INTRODUCTION

This paper presents a revision of the Rosenberg and Abbeduto D-Level scale for rating the complexity of sentences.

Perhaps the surest way to judge syntactic complexity is to observe the sequence in which children acquire the ability to use various types of sentences; the most complex sentence types, by definition, are those that children acquire last. Rosenberg and Abbeduto (1987) surveyed research on language acquisition and constructed a seven-step developmental level (D-Level) scale to rate sentence complexity. As far as we know, it is the only acquisition-based sentence complexity scale in current use.

Recent research has shown a strong correlation between low sentence complexity in early life and symptoms of Alzheimer's disease in old age (Snowdon et al., 1996; Kemper et al., 2001). Because of the importance of this result, a critical review of the complexity scales that were used to obtain it is in order. D-Level is one such scale.

The original 7-point D-Level pertained only to various kinds of complex sentences. Cheung and Kemper (1992) added a zeroth level for simple sentences. We have reviewed Rosenberg and Abbeduto's original work as well as more recent research in order to develop a revised D-Level scale that is theoretically sound and practically useful.

The original D-Level scale left some sentences unrated. That was acceptable because Rosenberg and Abbeduto measured a child's developmental level by noting the maximum complexity that the child reached, not the average complexity of a set of sentences. However, work such as Snowdon's requires averaging the complexity of all the sentences in a text. Accordingly, we have generalized the scale so that every sentence receives a score.

Based on recent psycholinguistic research, we have re-categorized sentences with semi-auxiliaries (e.g., *be gonna*) and raising verbs (*seem*). We have also swapped Rosenberg and Abbeduto's levels 5 and 6, since many experimental results show that these levels were originally in the wrong order.

A brief experimental test indicates that the new scale has greater reliability than the original scale.

With the revised D-Level scale, we have shored up scoring uniformity, brought the original scale up to date based on current research, and made sure that all types of sentences are covered while maintaining the ease of use that was a major advantage of the original scale.

THE ORIGINAL D-LEVEL SCALE

Rosenberg and Abbeduto's original D-Level scale is shown in Table 1. It does not rate every sentence; rather, it identifies certain landmarks, sentence-types that indicate a particular developmental level. A person's speech is rated by looking for the highest-numbered sentence type that regularly occurs, not by trying to rate all the sentences. Although the scale is based on studies of child language, Rosenberg and Abbeduto actually used it to rate impaired adults. Since then, it has been used for a number of other purposes. Neither the reliability nor the validity of the original scale has been reported.

REVISING THE D-LEVEL SCALE

Our revision of the D-Level scale addresses two issues: the ability to rate every sentence and the appropriate arrangement of levels.

Extending the scale to rate every sentence

Many psycholinguistic experiments rate the complexity of a particular text or speech sample rather than the overall competence of an individual. For this to be practical, it is necessary to assign a rating to every sentence. To make this possible, we have filled in the scale as described in the following paragraphs.

The original D-Level scale was based partly on actual observations of language acquisition, and partly on grammatical generalizations. The revised scale continues to have this mixed heritage. As far as we can determine, the acquisition of most parts of the grammar of English has not been mapped in detail. Accordingly, we follow Rosenberg and Abbeduto's precedent of filling in the scale through generalizations about grammar.

Simple sentences are rated at level 0. We follow Cheung and Kemper (1992) in adding a Level 0 for simple sentences that do not contain any of the higher-numbered indications of complexity. As we explain later, these include not only simple NP+VP (subject-predicate) structures, but also the corresponding simple *wh*-questions.

Elliptical sentences are rated only by what is actually uttered. When a sentence is incomplete, we rate the complexity of what is actually uttered, not the complete sentence that could be considered its target. Thus, elliptical (incomplete) sentences with only an NP, or only the parts of a simple sentence, will be rated at Level 0, and more complex elliptical sentences will be rated at face value (without reference to items recoverable from the context). For example, in the dialogue:

A: *Who said that John came to the party?*

B: *Mary.*   or   *Mary did.*

the second speaker gets credit only for Level 0 (a simple sentence), not the Level 3 complex sentence that could be claimed to be its target (*Mary said that John came to the party*). The rationale for this decision is that children commonly understand sentences more complex than they can produce.

Questions are rated the same as the corresponding declarative sentences.   There are two kinds of questions in English, simple yes/no questions (*Is he here?*) and "*wh*-questions" formed with the "*wh*-words" (*who, what, which, whose, where, when, why, how*). Both kinds are very common in children's speech but were not mentioned in the original D-Level scale.

Although questions normally emerge slightly later than simple statements, we assign them to the same level as their correspondingly structured declarative counterparts.

The alternative would have been to place them one level higher, which is undesirable for two reasons. First, it would require a Level 8 for the questions corresponding to Level 7 sentences. Second, the formation of a question adds, at most, only a small amount of complexity to the sentence, and in some cases none at all (e.g., *John sang* vs. *Who sang?*).

Thus a simple question such as *Where are you going?* is rated the same as its Level 0 declarative counterpart (*You are going somewhere*), and the complex question *Who did he say that John met at the party?* is Level 3 together with its corresponding statement *He said that John met Mary at the party.*

Coordinate structures are assigned Level 2.   Noun phrases and simple sentences joined by *and* were listed in Level 2 of the original scale, but other coordinate structures were not included. We supplement Level 2, accordingly, with coordinate verbal, adjectival and adverbial structures. Sentences of this type include: *He sang and jumped on the way home; She is the prettiest and cleverest little girl; Where and when did he see me?* Besides *and*, there are six coordinating conjunctions: *or, but, yet, so, for, nor.* We also include in this level structures conjoined by the four correlative conjunctions: *both...and, neither...nor, either...or, not (only)...but.*

Wh-complements are moved to Level 3.   A *wh*-complement is a subordinate clause introduced by a *wh*-word, such as *where it is* in a sentence such as *Remember where it is?*

The original scale placed *wh*-complements in Level 2, along with coordinate structures; all other subordinate structures were assigned higher levels.

The original scale therefore claims that *I remember where it is* (original Level 2) is simpler than *I remember it was here* (original Level 3). This does not seem justified. The two sentences are very much alike; each includes a finite subordinate clause; and, if anything, the sentence without the *wh*-word is semantically simpler.

Accordingly, we place finite *wh*-complements in Level 3 and nonfinite *wh*-complements in Level 4, along with similar complements not marked with *wh*-words.

Nominalizations in object position are assigned Level 3.   Nominalizations are sentences converted into abstract noun phrases, such as *the enemy's destruction of the city* (from *the enemy destroyed the city*). In the original scale,

nominalization was mentioned only in Level 5, for nominalizations occurring as subjects. Nominalizations in object position were not mentioned.

An obvious solution is to classify nominalizations in object position in Level 3, like clauses in object position. This preserves the generalization, reflected throughout the scale, that complex objects are simpler to process than complex subjects (due to right-branching rather than left-branching).

However, we must take care with the definition of "nominalization." Not all abstract nouns derived from verbs actually reflect structures comparable to an embedded clause. It would be inappropriate to rate *Construction is in progress* as Level 5 simply because its subject is a derived noun. We therefore count a noun phrase as a "nominalization" only when it includes explicitly the subject and/or object of the root verb (thus *the city's construction of the roads,* but not *the construction*).

Subject extraposition is included in Level 3. Extraposition is the use of *it* in subject position, with the understood subject following at the end of the sentence: *It puzzled me that he left* (vs. *That he left puzzled me,* without extraposition).

The original D-level scale leaves extraposition unclassified. However, there is a widespread consensus that the purpose of extraposition is to make sentences easier to process by moving a complex structure to the right, thus changing left-branching into right-branching (Yngve, 1960; Frazier & Fodor, 1978; Levin & Garrett, 1990; Anagnopoulos et al., 1990; Mazuka & Lust, 1990; see however Frazier & Rayner, 1988). Therefore, it is more reasonable to group such extraposed sentences with the right-branching structures at Level 3 than with the left-branching structures at Level 5.

Non-finite and verbless complements are in Level 1 if they implicitly share the main clause subject, and Level 4 otherwise. Several kinds of non-finite and verbless embedded clauses have been grouped together by theoreticians as "small clauses." Their hallmark is that they do not contain a verb with the normal inflections marking number and tense. They may contain a verb marked with –*ing* or –*ed*, an infinitive (*to* + V), or a verbless structure in which "to be" is understood. These structures share many similarities cognitively. Some examples:

I expect <u>you to pass the exam</u>.   (infinitive)

I saw <u>him walking the dog</u>.   (–*ing* clause)

I want <u>these things painted</u>.   (–*ed* clause)

I want <u>them out of the house</u>.   (verbless small clause)

Originally, Level 4 comprised all –*ing* complements. We propose to add to Level 4 all the other types of small clause complements with explicit subjects of their own. However, when an –*ing* clause has the same understood subject as the main clause, we place it in Level 1 along with the corresponding infinitive complements. Careful observation of one two-year-old subject has convinced us that *Try brushing her hair* is acquired at the same stage as *Try to brush her hair,* and likewise *I like dancing* is comparable to *I like to dance,* certainly not three levels later.

Non-finite adjuncts are in Level 6. Nonfinite adjuncts such as *when done, before doing, having done,* and *in hopes of … -ing,* were not rated in the original scale. Level 4 included nonfinite –*ing* complements, and we have just added other nonfinite complements to it. But adjuncts are not the same. The difference

is that instead of supplying an element required by the main verb, an adjunct expresses an entirely optional, additional element of meaning.

The original Level 6 included adjuncts which are finite subordinate clauses introduced by *when, before,* and the like. It is reasonable to place nonfinite adjuncts into the same category, since they are syntactically and cognitively similar.

Level 7 requires two different levels of embedding. The original Level 7 sentences are defined by "having more than one kind of embedding in a single sentence." This definition is ambiguous. The sentence *I was sad, and I didn't go, but he came to me.* has two coordinated structures with *and* and *but* respectively. Are there two kinds of embedding or one in this sentence? We clarify the definition by requiring two different levels of embedding in Level 7 sentences. According the new definition, the above example is still level 2. However, the sentence *I was sad, but I didn't want to tell him.* is Level 7, since it includes a Level 1 infinitive object and a Level 2 coordination.

Rearranging levels on the basis of psycholinguistic evidence

The original D-Level scale does not entirely fit the psycholinguistic data on which it was based, nor the results of later experiments by others. Accordingly, we propose some rearrangements.

Semi-auxiliaries are moved to Level 0. Alongside auxiliaries (*has, does, may, can, will, could,* etc.), English has a widely recognized class of "semi-auxiliaries" or "auxiliary-like verbs" whose syntax and meaning are often very similar to that of true auxiliaries. They "express modal or aspectual meaning" and "are introduced" by either *have* or *be* (Quirk et al, 1985, p. 143). The most commonly

used semi-auxiliaries include *be going to* (which doesn't mean "go" at all), *be able to, be about to, be likely to,* and *have to* (which is the only semi-auxiliary beginning with *have*) (Quirk et al., 1985, pp.143-145; Greenbaum, 1996, section 5.33).

In the original D-Level scale, Rosenberg and Abbeduto (1987, p. 26) classified *I am going to meet John* as a Level 1 infinitive complement just like *Try to brush her hair.* This is correct if the former sentence means, "I am on a journey to meet John," but more commonly, *am going to* is simply a future tense marker and should be included in Level 0, just like *will* and *can.*

Raising structures are placed in Level 3. A raising structure is one in which the subject of the embedded clause has been "raised" to become subject of the main clause, such as *He seems* [*to like it*] (from *He likes it*).

The original D-Level scale would lead us to classify raising structures such as *Mary seems to John* [*to be sad*] as Level 1 infinitive complements, since the embedded verb is an infinitive. This, however, does not conform to children's linguistic development. Acquisition studies show a noticeable increase in sentence complexity from non-raising structures like *It seems to John* [*that Mary is sad*] to their corresponding raising structures like *Mary seems to John* [*to be sad*], and young children's language features a lack of raising structures. In fact, "the CHILDES database (MacWhinney & Snow, 1985) failed to turn up an indisputable example of raising out of an infinitival clause in any child under age of 5" (Frank, 1998, p. 263). If the non-raising sentence *It seems to John* [*that Mary is sad*] is ranked Level 3, then the corresponding raising structure *Mary seems to John* [*to be sad*] should be ranked at least as high.

Levels 5 and 6 are swapped.   The sources that Rosenberg & Abbeduto used to develop the D-Level scale indicate that Level 5 may well be more advanced than Level 6. According to Miller (1981), subordinate clauses marked with *if* (Level 6 in the original scale) are observable in children chronologically prior to gerundive structures (Level 4). Bowerman (1979) specifically mentions Level 6 items such as *because, so, when, if, or, but* and *while,* but notes that extended subject structures as in Level 5 are missing from child-corpora in general.

Rosenberg and Abbeduto's own study also indicates that Level 5 and Level 6 may be improperly ranked. Most glaringly, there are no instances of Level 5 sentences in any of their seven subjects. However, there are Level 6 and Level 7 sentences within their corpus. Also, a breakdown of levels of the constituent phrases making up Level 7 sentences supports a possible switch. Relative frequency of these constituents shows "a systematic relationship between development level and [constituent] use," with one exception (Rosenberg & Abbeduto, 1987, p. 27). The percent of Level 6 constituent phrases actually falls in between the percentages for Level 3 and Level 4.

The stark contrast between the extreme rarity of Level 5 structures and the frequent and early occurrence of Level 6 structures strongly suggests that Levels 5 and 6 have been misplaced. The simplest remedy is to switch Level 5 and Level 6. This switch not only approximates child-acquisition findings, but also provides an adequate explanation of why actual mildly retarded adult individuals' usages of constituent sentences within Level 7 had the above-mentioned irregularities.

THE REVISED D-LEVEL SCALE AND ITS RELIABILITY

To summarize, our revised D-Level scale can be understood as follows:

Level 0: simple sentence

Level 1: non-finite clause as object without overt subject

Level 2: coordinated structure

Level 3: finite clause as object (and equivalents)

Level 4: non-finite clause as object with overt subject (and
equivalents)

Level 5: finite or non-finite adjunct clause

Level 6: complex subject

Level 7: more than one structure of Level 1-6

Table 2 shows the new scale, with examples.

After the revised scale was finished, a small-scale experiment was conducted to test the reliability of the revised scale and to compare it with the reliability of the original scale. Four raters participated in the experiment independently. The raters include both linguists and non-linguists, both native speakers of English and non-native speakers. Five different paragraphs were chosen for the experiment, which consisted of 37 sentences from different styles of literature ranging from newspapers to children's books. The raters were asked to rate the sentences using both the original scale and the new scale. There was no restriction on the time used for rating the sentences. Based on the ratings from the four raters, the inter-rater reliability of the revised scale was 0.93, whereas that of the original scale was 0.86. The result is within our expectation, since unlike the original scale, our revised scale assigns proper levels to all sentences, thus reducing subjectivity.

With the revised D-Level scale, we have reinforced scoring uniformity, brought the original scale up to speed with current work and made sure that all types of sentences are covered without creating a scale that is too difficult or time-consuming to be cost-effective. We welcome the application of the revised D-Level scale in all kinds of psycholinguistic study to verify both its validity and technical details.

ACKNOWLEDGMENTS

REFERENCES

Anagnopoulos, C., Norman, S., Cheung, H., & Kemper, S. (1991). Syntactic Complexity and Adults' Listening Comprehension: A Task Comparison. In F. Ingemann (Ed.), *Mid America Linguistics Papers* (pp. 43-57). Lawrence: Department of Linguistics, University of Kansas.

Bowerman, M. (1979). The acquisition of complex sentences. In P. Fletcher & M. Garman (Eds.), *Language Acquisition* (pp. 285-306). Cambridge: Cambridge University Press.

Cheung, H., & Kemper, S. (1992). Competing complexity metrics and adults' production of complex sentences. *Applied Psycholinguistics, 13,* 53-76.

Frank, R. (1998). Structural complexity and the time course of grammatical development. *Cognition: International Journal of Cognitive Science, 66 (3),* 249-301.

Frazier, L., & Keith, R. (1988). Parameterizing the language processing system: Left- vs. right-branching within and across languages. In J. Hawkins (Ed.), *Explaining LanguageUniversals* (pp. 247-279). Oxford: Basil Blackwell.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition, 6,* 291-325.

Greenbaum, S. (1996). *The Oxford English Grammar*. Oxford: Oxford University Press.

Kemper, S., Rice, K., & Chen, Y. J. (1995). Complexity metrics and growth curves for measuring grammatical development from five to ten. *First Language*, *15,* 151-166.

Kemper, S., Greiner, L. H., Marquis, J. G., Prenovost, K., & Mitzner, T. L. (2001). Language decline across the life span: Findings from the Nun Study. *Psychology and Aging, 16(2),* 227-239.

Lee, L. L. (1974). *Developmental Sentence Analysis*. Evanston: Northwestern University Press.

Levin, H., & Garrett, P. (1990). Sentence structure and formality. *Language in Society, 19,* 511-520.

MacWhinney, B., & Snow, C. (1985). The child language data exchange system. *Journal of Child Language, 12,* 271-296.

Mazuka, R. & Lust, B. (1990). On parameter setting and parsing: predictions for cross-linguistic difference in adult and child processing. In L. Frazier & J. De Villiers (Eds.), *Language Processing and Language Acquisition* (pp. 163-205). Dordrecht: Kluwer Academic Publishers.

Miller, J. F. (1981). *Assessing Language Production in Children*. Baltimore: University Park Press.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.

Rosenberg, S., & Abbeduto, L. (1987). Indicators of linguistic competence in the peer group conversational behavior of mildly retarded adults. *Applied Psycholinguistics, 8,* 19-32.

Scarborough, H. S. (1990). Index of Productive Syntax. *Applied Psycholinguistics*, *11,* 1-22.

Snowdon D. A., Kemper, S. J., Mortimer, J. A., Greiner, L. H., Wekstein, D. R., & Markesbery, W. R. (1996). Linguistic ability in early life and cognitive function and Alzheimer's disease in late life: Findings from the Nun Study. *Journal of the American Medical Association, 275(7),* 528-532.

Yngve, V. H. A. (1960). A model and an hypothesis for language structure. *Proceedings of the American Philosophical Society, 104,* 444-466.

TABLES

Table 1: Original D-Level scale (Rosenberg & Abbeduto 1987).

Level 1   Infinitival complement with same subject as main clause
Try <u>to brush her hair</u>. I am going <u>to meet John</u>.

Level 2   Sentence with *wh*-complement
Remember <u>where it is</u>.   I will tell you <u>what's happening</u>.

Sentences joined by coordinating conjunctions
<u>I brought candy</u> and <u>Peter cleaned up</u>.

Subjects consisting of 2 NPs joined by a conjunction
<u>John and Mary</u> left early.

Level 3   Relative clause modifying object of main verb
The man scolded the boy <u>who stole the bicycle</u>.

Complement clause serving as object of main verb
John knew <u>that Mary was angry</u>.

Level 4   *-ing* form as complement
I felt like <u>turning it</u>.

Comparative with object of comparison
John is <u>older than Mary</u>.

Level 5   Relative clause modifying subject of main verb
The man <u>who cleans the rooms</u> left early today.

Embedded clause serving as subject of main verb
<u>For John to have left Mary</u> was surprising.

Nominalization serving as subject of main verb
<u>John's refusal of the drink</u> angered Mary.

Level 6   Sentences joined by a subordinating conjunction
They will play today <u>if it doesn't rain</u>.

Level 7   More than one kind of embedding in a single sentence
John decided <u>to leave Mary</u> <u>when he heard</u> <u>that she was seeing Mark</u>.

Table 2: Our revised D-Level scale.

Level 0	Simple sentences, including questions	*The dog barked.* *Did the dog bark?* *Where are you going?*
	Sentences with auxiliaries and semi-auxiliaries	*This <u>may have</u> solved it.* *He is <u>going to</u> take the bus.*
	Simple elliptical (incomplete) sentences	*The dog over there.* *He did.*
Level 1	Infinitive or *-ing* complement with same subject as main clause	*Try <u>to brush her hair.</u>* *Try <u>brushing her hair.</u>* *I felt like <u>turning it.</u>*
Level 2	Conjoined noun phrases in subject position	*<u>John and Mary</u> left early.*
	Sentences conjoined with a coordinating conjunction	*<u>I came early</u> but <u>Peter arrived late</u>.*
	Conjoined verbal, adjectival, or adverbial constructions	*He <u>sang and jumped</u> on the way home.*
Level 3	Relative (or appositional) clause modifying object of main verb	*The man scolded the boy <u>who stole the bicycle</u>.*
	Nominalization in object position	*Why can't you understand <u>his rejection of the offer</u>?*
	Finite clause as object of main verb	*John knew <u>that Mary was angry</u>.* *Remember <u>where it is</u>?*
	Subject extraposition	*<u>It</u> was surprising <u>for John to have left Mary</u>.*
	Raising	*<u>John</u> seems to Mary <u>to be happy</u>.*
Level 4	Non-finite Complement with its own understood subject	*I expect <u>him</u> <u>to go.</u>* *I want <u>it</u> <u>done today</u>.* *I saw <u>him</u> <u>walking the dog</u>.* *I consider <u>John</u> <u>a friend</u>.* *I want <u>these animals</u> <u>out of my house</u>.*
	Comparative with object of comparison	*John is <u>older than Mary</u>.*
Level 5	Sentences joined by a subordinating conjunction	*They will play today <u>if it does not rain</u>.*
	Nonfinite clauses in adjunct (not complement) positions	*Cookie Monster touches Grover <u>after jumping over the fence</u>.* *<u>Having tried both</u>, I prefer the second one.*

Level 6	Relative (or appositional) clause modifying subject of main verb	*The man <u>who cleans the rooms</u> left early.*
	Embedded clause serving as subject of main verb	*<u>For John to have left Mary</u> was surprising.*
	Nominalization serving as subject of main verb	*<u>John's refusal of the drink</u> angered Mary.*
Level 7	More than one level of embedding in a single sentence	*John decided <u>to leave Mary</u> <u>when he heard</u> <u>that she was seeing Mark</u>.*