

Measuring linguistic complexity independent of plausibility*

JEFF GRUBER

EDWARD GIBSON

Massachusetts Institute of Technology

Massachusetts Institute of Technology

1. INTRODUCTION. Many factors affect the ease or difficulty of understanding a word in a sentence, including the lexical frequency of the word, the syntactic context, the discourse context, the intonation of the sentence, and the plausibility of the situation described by the sentence thus far (see Gibson & Pearlmuter 1998 and Tanenhaus & Trueswell 1995 for summaries of relevant factors and evidence). In order to investigate any one of these factors, it is therefore necessary to control for the others. One factor that has recently begun to be investigated in the psycholinguistics literature is the derivational complexity of different lexical structures. In order to investigate lexical derivational complexity, researchers have compared reading times and sentence-acceptability judgments for different kinds of noun-verb pairs. Unfortunately, these measures are necessarily contaminated by potential differences in plausibility between the real-world situations described by the particular sentences that are used to instantiate the structures. For example, consider McKoon and Macfarland's (2000; henceforth M&M) evidence that externally caused change-of-state verbs are more complex than internally caused change-of-state verbs (see Levin & Rappaport Hovav 1995 for a lexical-semantic distinction between the two types of verbs). M&M demonstrated that participants were slower in deciding that externally caused sentences like 1a were acceptable than they were in deciding that internally caused sentences like 1b were acceptable.¹

- (1) a. External cause: The leaves wilted.
b. Internal cause: The signal faded.

From this evidence, M&M concluded that the externally caused materials were lexically more complex than the internally caused materials. But there is an alternative interpretation of the observed data pattern: that the set of items in the external-cause conditions described less plausible situations in the world on the whole than the set of items in the internal-cause condition. Such a difference is not intuitively obvious in 1, because both events are plausible. But across the items, there may have been a small difference in this dimension.² In general, the two types of explanations—lexical complexity vs. plausibility—will always be possible in comparisons between different sets of items, as in M&M's comparison.

* We would like to thank Melissa Yeh for help in running the experiments that are described here. We would also like to thank the following people for comments on earlier presentations of this work: two anonymous *Language* referees, Dan Grodner, Neal Pearlmuter, and the audience at the 2003 CUNY conference on sentence processing, where this work was presented as a poster. We take full responsibility for all remaining errors.

¹ It is beyond the scope of this paper to discuss the details of the lexical theories that give rise to this prediction. See M&M and Levin & Rappaport Hovav 1995 for such details.

² Other factors that are known to contribute to sentential complexity are less likely confounds. Lexical frequency was controlled in the sets of items that M&M compared, and syntactic context, discourse context, and intonational information were constant across the comparison.

Indeed, a plausibility questionnaire that we conducted with twenty-nine participants on a single survey containing the two sets of items that M&M tested (intransitive and transitive) revealed similar differences between M&M's conditions: the externally caused items were rated as less plausible than the internally caused items for each comparison (intransitive: $F(1,28) = 34.6, p < 0.001$; transitive: $F(1,28) = 7.12, p < 0.02$). Although people were asked to rate the real-world plausibility of the target sentences in this task, participants may have rated only whether the sentences sounded 'good' or not, which probably reflects both plausibility and lexical-syntactic complexity. Thus, like M&M's acceptability-time task, the plausibility-rating task may reflect differences in lexical complexity of the items or real-world plausibility of the events described in the items, or both.

2. A NEW QUESTIONNAIRE METHOD. In order to distinguish the two parameters involved in these judgments—ease/complexity of linguistic derivation of the expression vs. implausibility/plausibility of the situation described by that expression—we designed a new questionnaire format in which participants rated both of these factors independently. In this paradigm, participants were presented with pairs of sentences, many of which were good paraphrases of one another. Participants were asked to rate the degree to which the two sentences described the same situation, rating (i). If they rated them as similar, they were then asked to give three further ratings: (ii) the plausibility of the situation described by both, (iii) the complexity of the manner of expression of this situation for the first paraphrase, and (iv) the complexity of the manner of expression of this situation for the second paraphrase. This method gave us the potential to dissociate plausibility ratings (rating (ii)) from linguistic-complexity ratings (ratings (iii) and (iv)).

Before evaluating M&M's items using this new method, we wanted to see if the method would be sensitive enough to detect complexity differences in two comparisons that are known to be of differing linguistic complexity: (a) subject-extracted vs. object-extracted relative clauses, and (b) active vs. passive sentences.

3. EXPERIMENT 1: SUBJECT- AND OBJECT-EXTRACTED RELATIVE CLAUSES AND ACTIVE/PASSIVE SENTENCES. It is well known that object-extracted relative clauses (RCs) like 2b are more complex than subject-extracted RCs like 2a (see e.g. Gibson 1998 for a recent summary).

- (2) a. Subject-extracted: The boy recited the limerick that appalled the priest.
 b. Object-extracted: The limerick that the boy recited appalled the priest.

Furthermore, passive sentences like 3b are more complex than corresponding active sentences like 3a (Forster & Olbrei 1973, Carrithers 1989, Rohde & Gibson 2003).

- (3) a. The architect built the house.
 b. The house was built by the architect.

In experiment 1, we compared these structures using the new questionnaire method.

3.1. PARTICIPANTS. Twenty-eight native speakers of English from the MIT community were paid \$5 each to complete the survey using pencil and paper. The survey took approximately twenty minutes to complete.

3.2. MATERIALS. The RC sentence pairs were taken from the materials of Stromswold et al. 1996. Sixteen pairs were used, each with an object-extracted and subject-extracted version, as in 2. In the object-extracted structure, the RC modified the subject NP of the main clause, whereas in the subject-extracted structure, the RC modified the object

NP of the main clause. In addition, sixteen active/passive sentence pairs were constructed, having the form of 3. The active sentence was always a simple subject-verb-object sentence, and the passive sentence was the passive form of this sentence. The thirty-two target sentence pairs were combined with forty-eight pairs of sentences from an unrelated experiment and thirty pairs of filler sentences of various types, for a total of 110 sentence pairs. The pairs of sentences in each of the three subexperiments (eighty items in total) were constructed with the intention that the two sentences in a pair would be good paraphrases of one another, as in 2 and 3. The thirty filler sentence pairs were written to be poor paraphrases of one another.

In order to control for ordering effects, we created two sublists of items: one sublist with eight of the subject-extracted items and eight of the active items presented first in each sentence pair, and another sublist with the inverted order. We then created four different random orderings of the list combining these two sublists of items. Four sentence-pair items were presented on each page of a survey. The pages were randomized separately for each participant.

3.3. PROCEDURE. Participants were presented the following instructions in a paper-and-pencil questionnaire:

Instructions:

Read the two sentences **(a)** and **(b)** in each set. Then rate, on a scale of 1 (high) to 7 (low):

- (i) The degree to which the two sentences describe the same situation.

If you judge them to be very closely the same situation, e.g. rating them 1, 2, or 3, then rate, on the same scale,

- (ii) The degree to which the situation described is a plausible or natural situation, i.e. likely to occur.
 (iii) The degree to which sentence **(a)** is an easy, simple or natural way to express that situation.
 (iv) The degree to which sentence **(b)** is an easy, simple or natural way to express that situation.³

Example 1. a. The man gave the dog the bone.

b. The man gave the bone to the dog.

(a) and **(b)** describe the SAME situation. 1 2 3 4 5 6 7 **(a)** and **(b)** describe describe DIFFERENT situations.

If you circled 1, 2, or 3 above (judging **(a)** and **(b)** to be very closely the same situation), then rate also the following:

The situation is PLAUSIBLE. 1 2 3 4 5 6 7 The situation is IMPLAUSIBLE.

In this set you might judge **(a)** and **(b)** as describing very closely the same situation, so you would circle 1 or 2. Because the two describe the same situation, you would also rate the plausibility of the situation described, perhaps circling 2 or 3. Similarly you would also rate whether **(a)** or **(b)** are easy or hard ways to express the situation (not shown).

Participants were provided with three example items like example 1 above, in which they were given explicit advice. In the first two of these items, the two sentences were good paraphrases of one another, whereas this was not the case in the third item.

For consistency, we refer to rating (i) above as the similarity rating between the sentences, rating (ii) as the plausibility rating for the situation, and ratings (iii) and (iv) as simplicity (of expression) ratings.

³ Any one of the terms 'easy', 'simple', and 'natural' alone is not unambiguous. The combination of the three together was intended to pinpoint their intersection, that is, derivational simplicity. Thus, the word 'natural' alone could refer to register, informal, casual language being more 'natural' than formal, specialized language, for example, but this is not likely to be the meaning of 'simple'. In this regard, it is also worth noting that because the verbs in the target sentences were controlled for average frequency and specialized language would almost certainly involve less frequent vocabulary, register is unlikely to be a confound.

- (6) a. The sidewalk crumbled.
 b. The sidewalk got wet.

The list preparation and procedure were the same as in experiment 1.

4.3. RESULTS. Table 1 presents the mean plausibility and simplicity ratings by participants by condition for experiment 2. In the intransitive conditions, participants rated the target item as being similar in meaning to its paraphrase on 583 of the 700 intransitive trials (83%). Broken down by condition, 303 of 350 external-cause trials (87%) were rated as similar to their paraphrases, and 280 of 350 internal-cause trials (80%) were rated as similar to their paraphrases. On these trials, there was no difference in plausibility/naturalness of situation between the internal-cause and external-cause items ($F(1,24) = 0.379, p = 0.54$), but there was a marginal difference in the simplicity-of-expression ratings, such that the internal-cause items were rated as simpler than the external-cause items ($F(1,24) = 3.54, p = 0.07$).

Turning now to the transitive items, on 1,160 of the 1,375 transitive trials (84%), participants rated the target item as being similar in meaning to its paraphrase.⁴ Broken down by condition, 587 of 700 external-cause trials (84%) were rated as similar to their paraphrases, and 573 of 675 internal-cause trials (85%) were rated as similar to their paraphrases. On these trials, there was no difference in plausibility/naturalness of situation between the internal-cause and external-cause items ($F(1,24) = 0.68, p = 0.42$), but there was a significant difference in the simplicity-of-expression ratings, such that the internal-cause items were rated as simpler than the external-cause items ($F(1,24) = 7.57, p < 0.02$).

	INTRANSITIVE, EXTERNAL CAUSE	INTRANSITIVE, INTERNAL CAUSE	TRANSITIVE, EXTERNAL CAUSE	TRANSITIVE, INTERNAL CAUSE
Plausibility	1.60 (0.10)	1.56 (0.12)	1.78 (0.13)	1.72 (0.13)
Simplicity of expression	1.65 (0.10)	1.52 (0.09)	2.10 (0.10)	1.91 (0.12)

TABLE 1. Mean (standard error) plausibility and simplicity-of-expression ratings by condition for experiment 3 (1 = natural/simple, 7 = unnatural/complex).

We performed some additional analyses in order to confirm that the observed differences in simplicity ratings for the target items were not due to potential differences in the paraphrases. One possibility is that participants' simplicity ratings on the target items may indirectly reflect the length of the paraphrase in some way. Such a factor cannot have played a part in the complexity difference observed for the intransitive materials because the length of paraphrases was matched in these items: mean length of external-cause paraphrases: 4.9 words per item; mean length of internal-cause paraphrases: 5.1 words per item. Although the length of paraphrases was not matched in the transitive items (mean length of external-cause paraphrases: 7.6 words per item; mean length of internal-cause paraphrases: 6.7 words per item), analyses on a subset in which paraphrase length was matched (which was obtained by systematically removing long paraphrase items) revealed statistical patterns similar to the original set. In particular, there was no plausibility difference in this subset, but there was still a significant difference in simplicity ratings.

We performed one final analysis in which we examined the simplicity ratings of the paraphrases. We reasoned that it was possible that the ratings on the target items may

⁴ There was a typographical error in a paraphrase of a transitive internal-cause item (*the plans produced yellow blossoms* instead of *the plants produced yellow blossoms*). Our analyses omit this item.

have been affected in some way by the complexity of the paraphrases. In the intransitives, the paraphrases for the external-cause items were significantly simpler than the paraphrases for the internal-cause items (internal-cause rating = 3.38 (standard error = 0.23); external-cause rating = 2.82 (standard error = 0.20); $F(1,24) = 22.7$, $p < 0.001$), but the reverse effect was present in the paraphrases of the transitive items (internal-cause rating = 3.03 (standard error = 0.16); external-cause rating = 3.29 (standard error = 0.21); $F(1,24) = 5.36$, $p < 0.05$). Thus, the complexity of the paraphrases cannot have been driving the effects in both the transitive and intransitive items. We therefore conclude that it is unlikely that the complexity rating for the paraphrase was responsible for the complexity pattern that was observed in the target items in either comparison.

4.4. DISCUSSION. The results of this experiment demonstrated a similar pattern across the intransitive and transitive verbs. In each case, there was no plausibility difference between the external-cause items and the internal-cause items, but there was a difference in their simplicity of expression such that the internal-cause items were rated as simpler than the external-cause items.

5. CONCLUSIONS. The main contribution of this discussion note is to propose an experimental design for measuring lexical derivational complexity independent of the plausibility of the situations described in the sentences. This is important in localizing different sources of complexity of linguistic structures, which helps both in understanding the factors that make language more or less comprehensible and in understanding the content and structure of the linguistic representations themselves. Furthermore, the results of experiment 2, which used the new method, lead us to conclude that M&M's original results are not confounded with plausibility differences. The differences that they observed, that is, in judgment-response time, are likely to be due to lexical-complexity differences, that is, the greater complexity of external-cause, in contrast with internal-cause, change-of-state verbs, as they concluded.

APPENDIX: ITEMS FOR EXPERIMENT 2

GROUP 1: INTRANSITIVE EXTERNAL-CAUSATION VERBS

- | | |
|-----------------------------|------------------------------------------|
| 1a. The wind abated. | b. The wind died down. |
| 2a. His skills atrophied. | b. His skills wasted away. |
| 3a. The residents awoke. | b. The residents got up from sleep. |
| 4a. The concrete crumbled. | b. The concrete broke up in small bits. |
| 5a. The missile exploded. | b. The missile blew up. |
| 6a. The signal faded. | b. The signal became weak. |
| 7a. The crops shriveled. | b. The crops withered and shrank. |
| 8a. The fan vibrated. | b. The fan quivered rapidly. |
| 9a. The smoke dissipated. | b. The smoke spread and disappeared. |
| 10a. The bones fossilized. | b. The bones got preserved in stone. |
| 11a. The string frayed. | b. The string unraveled and broke apart. |
| 12a. His skin reddened. | b. His skin became reddish. |
| 13a. The bullet splintered. | b. The bullet broke in pieces. |
| 14a. The turkey thawed. | b. The turkey became unfrozen. |

GROUP 2: INTRANSITIVE INTERNAL-CAUSATION VERBS

- | | |
|----------------------------|---------------------------------------|
| 1a. The flowers bloomed. | b. The flowers came out. |
| 2a. The roof deteriorated. | b. The roof began to go to pieces. |
| 3a. The seeds germinated. | b. The seeds came alive and grew. |
| 4a. The potatoes rotted. | b. The potatoes got bad. |
| 5a. The beams rusted. | b. The beams broke down by oxidation. |

- | | |
|------------------------------|-----------------------------------------|
| 6a. The economy stagnated. | b. The economy was stalled. |
| 7a. The leaves wilted. | b. The leaves softened and drooped. |
| 8a. The roses withered. | b. The roses dried up and shrank. |
| 9a. Her skin blistered. | b. Her skin produced blisters. |
| 10a. The metal corroded. | b. The metal broke down. |
| 11a. The beaches eroded. | b. The beaches broke down by the water. |
| 12a. The cider fermented. | b. The cider became alcoholic. |
| 13a. The seedlings sprouted. | b. The seedlings began to grow. |
| 14a. His knee swelled. | b. His knee bloated and enlarged. |

GROUP 3: TRANSITIVE EXTERNAL-CAUSATION VERBS

- | | |
|------------------------------------------------------|----------------------------------------------------------------|
| 1a. The police abated violent crime. | b. The police brought down violent crime. |
| 2a. The doctors abated infectious diseases. | b. The doctors brought down infectious diseases. |
| 3a. The illness atrophied the lower leg. | b. The illness wasted the lower leg away. |
| 4a. The stroke atrophied the right brain. | b. The stroke weakened the right brain. |
| 5a. The fire alarm awoke the residents. | b. The fire alarm brought the residents out of sleep. |
| 6a. The telephone call awoke the reporter. | b. The telephone call brought the reporter out of sleep. |
| 7a. The French chef crumbled the cheese. | b. The French chef broke apart the cheese into small pieces. |
| 8a. The nervous teenager crumbled the cigarettes. | b. The nervous teenager broke up the cigarettes. |
| 9a. The scientists exploded the nuclear device. | b. The scientist caused the nuclear device to blow up. |
| 10a. Religious extremists exploded the bomb. | b. Religious extremists blew up the bomb. |
| 11a. The opera singer vibrated the notes. | b. The opera singer caused the notes to waver. |
| 12a. The famous violinist vibrated the strings. | b. The famous violinist made the strings quiver. |
| 13a. The huge radiators dissipated the heat. | b. The huge radiators spread the heat about. |
| 14a. The strong winds dissipated the clouds. | b. The strong winds caused the clouds to spread and disappear. |
| 15a. The scandal frayed party unity. | b. The scandal wore down and unraveled party unity. |
| 16a. The lengthy strike frayed tempers. | b. The lengthy strike caused tempers to wear down. |
| 17a. The wind reddened the baby's face. | b. The wind made the baby's face reddish. |
| 18a. The sunset reddened the evening sky. | b. The sunset colored the evening sky red. |
| 19a. The cook thawed the holiday turkey. | b. The cook unfroze the holiday turkey. |
| 20a. The mild temperatures thawed the soil. | b. The mild temperatures caused the soil to unfreeze. |
| 21a. The senators proliferated government subsidies. | b. The senators increased the number of government subsidies. |
| 22a. The politicians proliferated welfare programs. | b. The politicians increased the number of welfare programs. |
| 23a. Age mellowed the rebel leader. | b. Age made the rebel leader less aggressive. |
| 24a. Marriage mellowed the rock star. | b. Marriage softened the attitude of the rock star. |
| 25a. The solution oxidized the scrap metal. | b. The solution turned the scrap metal into its oxide. |
| 26a. The water oxidized the iron beam. | b. The water turned the iron beam into its oxide. |
| 27a. The weather stiffened his joints. | b. The weather made his joints inflexible. |
| 28a. The starch stiffened his shirts. | b. The starch made his shirts hard. |

GROUP 4: TRANSITIVE INTERNAL-CAUSATION VERBS

- | | |
|-------------------------------------------------|-------------------------------------------------|
| 1a. The plants bloomed yellow blossoms. | b. The plans produced yellow blossoms. |
| 2a. The trees bloomed pink flowers. | b. The trees produced pink flowers. |
| 3a. The heavy traffic deteriorated the bridge. | b. The heavy traffic wore down the bridge. |
| 4a. The harsh climate deteriorated the roads. | b. The harsh climate wore down the roads. |
| 5a. The local florist germinated the seeds. | b. The local florist got the seeds to grow. |
| 6a. The amateur botanist germinated the corn. | b. The amateur botanist got the corn to sprout. |
| 7a. The rare disease rotted the potatoes. | b. The rare disease turned the potatoes bad. |
| 8a. The flood water rotted the carpets. | b. The flood water made the carpets bad. |
| 9a. The constant rain rusted the car. | b. The constant rain made the car rusty. |
| 10a. The salt water rusted the machinery. | b. The salt water made the machinery rusty. |
| 11a. Communism stagnated the country's economy. | b. Communism made the country's economy stall. |

- 12a. The regulations stagnated private investments.
 13a. The intense heat wilted the crowd.
 14a. The bright sun wilted the roses.
 15a. The severe drought withered the crops.
 16a. The late frost withered the tulips.
 17a. The intense sun blistered the paint.
 18a. The hot sauce blistered the tongue.
 19a. Acid rain corroded the building.
 20a. The detergent corroded the fine silver.
 21a. The storm eroded the beach.
 22a. The river eroded the rocks.
 23a. The yeast cultures fermented the beer.
 24a. The apple growers fermented the cider.
 25a. The plants sprouted tender buds.
 26a. The trees sprouted tiny leaves.
 27a. The violent storms swelled the sea.
 28a. The rainy season swelled the rivers.
- b. The regulations kept private investment from growing.
 b. The intense heat caused the crowd to droop.
 b. The bright sun made the roses droop.
 b. The severe drought dried up and shriveled the crops.
 b. The late frost shriveled the tulips.
 b. The intense sun made the paint blistering.
 b. The hot sauce produced blisters on the tongue.
 b. Acid rain ate into the building.
 b. The detergent ate into the fine silver.
 b. The storm broke the beach away.
 b. The river broke down the rocks.
 b. The yeast cultures made the beer alcoholic.
 b. The apple growers turned the cider alcoholic.
 b. The plants produced tender buds.
 b. The trees produced tiny leaves.
 b. The violent storms bloated the sea.
 b. The rainy season bloated the rivers.

REFERENCES

- CARRITHERS, CAROLINE. 1989. Syntactic complexity does not necessarily make sentences harder to understand. *Journal of Psycholinguistic Research* 18.75–88.
- FORSTER, KENNETH I., and ILMAR OLBREI. 1973. Semantic heuristics and syntactic analysis. *Cognition* 23.319–47.
- GIBSON, EDWARD. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68.1–76.
- GIBSON, EDWARD, and NEAL PEARLMUTTER. 1998. Constraints on sentence comprehension. *Trends in Cognitive Science* 2.262–68.
- LEVIN, BETH, and MALKA RAPPAPORT HOVAV. 1995. *Unaccusativity: At the syntax-lexical semantics interface*. Cambridge, MA: MIT Press.
- McKOOON, GAIL, and TALKE MACFARLAND. 2000. Externally and internally caused change of state verbs. *Language* 76.833–58.
- ROHDE, DOUGLAS, and EDWARD GIBSON. 2003. The on-line processing of active and passive structures in English. Poster presented at the 16th Annual CUNY Conference on Human Sentence Processing, MIT, Cambridge, MA.
- STROMSWOLD, KARIN; DAVID CAPLAN; NATHANIEL ALPERT; and SCOTT RAUCH. 1996. Localization of syntactic comprehension by positron emission tomography. *Brain and Language* 52.452–73.
- TANENHAUS, MICHAEL K., and JOHN C. TRUESWELL. 1995. Sentence comprehension. *Speech, Language, and Communication*, ed. by Joanne Miller and Peter Eimas, 217–62. San Diego: Academic Press.

Gruber
 Department of Linguistics and Philosophy
 Massachusetts Institute of Technology
 Cambridge, MA 02139
 [jsgruber@alum.mit.edu]

[Received 8 April 2003;
 revision received 22 December 2003;
 accepted 14 May 2004]

Gibson
 Department of Brain and Cognitive Sciences
 NE20-459
 Massachusetts Institute of Technology
 Cambridge, MA 02139
 [legibson@mit.edu]