

B. T. S. Atkins, Oxford University Press
Beth Levin, Northwestern University
Grace Song, Northwestern University

Making Sense of Corpus Data: a Case Study¹

Abstract

This paper considers how corpus data is currently used in lexicography, and how it might be further exploited. We compare corpus evidence with entries in corpus-based dictionaries. The corpus evidence is difficult to understand and systematize. The dictionary entries are rich, but their coverage of the patterns is patchy. How can the facts in the corpus be made more amenable to systematic lexicographical treatment? We offer a theoretical analysis of the relevant aspect of verb behaviour and suggest that such analyses could facilitate the lexicographer's task and systematize lexicographical description.

In the past decade, English lexicography has benefited from the availability of large-scale text corpora and good corpus query software, and lexicographers in all large UK reference publishers now use this type of resource. English dictionaries have benefited from this wealth of evidence, but the potential of text corpora has not been exploited to the full, because:

- corpus-based dictionaries have to sell against their corpus-free competitors, and therefore editorial costs cannot be budgeted at a realistic rate to allow for in-depth corpus exploration;
- even if time were not constrained, lexicographical skills do not necessarily include the grasp of metalexicography and theoretical linguistics needed to tease out the corpus evidence.

Good lexicographers are good at writing dictionaries. Good theoretical linguists are good at linguistic analysis. Both types of skill are needed, if dictionaries are to benefit from the rich lexicographical evidence available in electronic text corpora.

We focus on one tiny lexico-semantic area which highlights the lexicographical problem of describing semantically related words in such a way as to make clear the similarities and differences among these words. We take as examples two verbs of sound emission, *rattle* and *rumble*, looking at the descriptions of the relevant senses of these verbs in three

of the leading English corpus-based dictionaries for learners: CIDE, COBUILD and LDOCE².

1. Problems presented by the corpus evidence

These semantically-related verbs should ideally receive coordinated lexicographical treatment, despite the fact that their behaviour is not identical: the corpus reveals that they differ with respect to transitivity. Specifically, we find one transitivity pattern common to both, but three patterns unique to *rattle*. The shared pattern expresses their primary sense: the emission of a specific sound, denoted by the verb itself. The sound event is expressed by an intransitive verb with the emitter – or source – of the sound as subject.

- [1] a. The windows rattled.
b. The trucks rumbled.

The three patterns unique to *rattle* found in the corpus are:

- [2] a. She heard ... the plastic rattling in the wind...
b. A gust of wind rattled the sash window.
c. The woman... rattled the doorknob in vain.

The composed sentences in [3] illustrate the patterns found in the corpus:

- [3] a. The doorknob rattled in the wind.
b. The wind rattled the doorknob.
c. The man rattled the doorknob.
d. The wind made the doorknob rattle.
e. The man made the doorknob rattle.

What sets these uses apart from those in [1] is the presence of the cause of the sound emission, expressed in a prepositional phrase ([3a]), or as the subject of *rattle* ([3b], [3c]) or the subject of *make* ([3d], [3e]). In [3a], the verb is still intransitive, with the source of the sound (the “sound emitter”) as subject; in [3b] and [3c], the verb is transitive, with sound emitter as the direct object, and the cause of the sound event as the subject. However, two transitive senses of *rattle* must be distinguished – one [3b] with an inanimate subject, and the other [3c] with an animate subject. These two types do not behave in the same way: those like [3b] with an “inanimate cause” subject have a causative paraphrase with *make*

plus intransitive *rattle*, as in [3d]. Those like [3c], with an “animate cause” subject, have no comparable causative expression. Though [3e] is parallel in structure to [3d], it is not a paraphrase of [3c].

To summarize, the patterning of verbs like *rattle* may be schematized as follows:

- SUBJ/SOURCE V as in [1a]
- SUBJ/SOURCE V PP/INAN-CAUSE as in [3a]
- SUBJ/INAN-CAUSE V DOBJ/SOURCE as in [3b]
- SUBJ/ANIM-CAUSE V DOBJ/SOURCE as in [3c]

The distribution of the transitivity-related patterns of the two verbs is shown in Table 1, which includes references to the appropriate illustrative sentences.

	rattle	rumble
SUBJ/SOURCE V	yes [1a]	yes [1b]
SUBJ/SOURCE V PP/INAN-CAUSE	yes [3a]	no
SUBJ/INAN-CAUSE V DOBJ/SOURCE	yes [3b]	no
SUBJ/ANIM-CAUSE V DOBJ/SOURCE	yes [3c]	no

Table 1. Transitivity-related patterns found in the corpus

2. Why *rattle* differs from *rumble*: the theory

A linguistic analysis of verbs of sound (Song in press) suggests that two factors determine the behaviour of the verbs: (i) whether the sound is emitted externally to or internally by the source, and (ii) whether the cause of the sound event is animate or inanimate.

2.1 External versus internal sound source

Most verbs denoting an action which originates within an entity (e.g., *laugh*, *speak*, *run*) are intransitive [4a], and have no transitive use [4b] with a causative interpretation to parallel the normal overt causative construction [4c].

- [4] a. Jenny laughed.
b. *We/The joke laughed Jenny.
c. We/The joke made Jenny laugh.

In contrast, verbs of change of state, including *break*, *melt*, and *open*, require an external cause to bring about the changes they denote. They have both a transitive use [5a] expressing a causative sense and an intransitive use [5b] expressing the caused event only.

- [5] a. The window broke.
b. Jane/The earthquake broke the window.

Sound emission can also be viewed from this perspective; a sound can be produced either apparently autonomously within the sound emitter (cf. *rumble*, *whir*), or by some contact with its external surface (cf. *jingle*, *clatter*). Some sounds – such as a rattle – can be produced either way: internally, as by machines with moving parts (*Slowly the boiler began to rattle*), or externally, through contact with or between surface parts (*She was ... expecting a key to rattle in the lock*).

Verbs like *rumble*, describing sounds with an internal source of production, occur only in the intransitive [6a]; they have no transitive, causative counterparts [6b] and the cause of the sound event is commonly expressed as the subject of *make* [6c]:

- [6] a. My stomach rumbled.
b. *Hunger rumbled my stomach.
c. Hunger made my stomach rumble.

However, verbs like *rattle*, describing sounds produced externally by means of surface contact, may be used transitively with a causative interpretation to describe the production of the sound, as shown in [2b].

2.2 Animate versus inanimate cause subjects

The distinction between these types of cause is introduced in Section 1, where it is illustrated by the two transitive uses of *rattle*. When the cause

of the sound event is inanimate (for instance, a force such as wind, an explosion, or an earthquake), the verb may be used either transitively with the inanimate cause as subject of the sound emission verb ([2b]), or intransitively, with the emitter as subject ([2a]). However, when the cause is animate, as in [2c], there is no intransitive counterpart.

To clarify this difficult point: when an inanimate cause brings about an externally produced sound such as a rattle, we can construe the sound emission as either internally or externally produced. In the presence of natural forces, physical objects can appear to rattle “on their own”, thus allowing the situation described by a transitive sentence ([3b]) also to be expressed by an intransitive sentence ([3a]). However, when the sound is caused by an animate being acting upon the sound emitter, the emitter is never understood as producing the sound independently. Thus transitives with animate cause subjects lack intransitive counterparts. For these reasons the intransitive *The window rattled* is more acceptable as a paraphrase of “Some force caused the window to rattle” than of “Some person caused the window to rattle”.

The licensing of intransitive paraphrases therefore depends on the type of cause. In the presence of inanimate forces, physical objects can appear to rattle “on their own”, thus allowing the situations described by these transitive sentences to be expressed also by intransitive sentences. This is not so when the sound is caused by an animate being acting upon the sound emitter: no intransitive counterpart is possible since the emitter is not understood to be producing the sound independently.

3. Dictionary treatment

How do the three dictionaries treat the two verbs with respect to the transitivity patterns? As expected, their entries for *rumble* indicate that this verb is always intransitive, and appropriate example sentences are given. In the case of the more complex *rattle*, the transitivity-related information in the entries is summarized in the Table 2³, using the schemas established in Section 1.

As Table 2 shows, all three dictionaries note transitive uses of *rattle*, but they imply, by their use of codes and examples, that every transitive use has a corresponding intransitive paraphrase; COBUILD even states this explicitly with the code “V-ERG.” However, the transitive and

intransitive pair of examples in CIDE are appropriately chosen: *The explosion rattled the cups* and *The cups rattled*.

	COBUILD	CIDE	LDOCE
A. SUBJ / SOURCE V			
indicated by codes:	V	I	I
indicated by definitions:	yes	yes	yes
illustrated in examples:	yes	yes	yes
– subj / source	dishes	cups,voice	windows
B. SUBJ / INAN-CAUSE V DOBJ / SOURCE			
indicated by codes:	Vn	T	T
indicated by definitions:	no	yes	yes*
illustrated in examples:	no	yes	no
– subj / cause		explosion	
– dobj / source		cups	
C. SUBJ / ANIM-CAUSE V DOBJ / SOURCE			
indicated by codes:	Vn	T	T
indicated by definitions:	yes	no	yes*
illustrated in examples:	yes	no	yes
– subj / cause	he		the beggar
– dobj / source	the cage		coins

Table 2⁴: Dictionary treatment of sound senses of *rattle*

4. Possible applications for practical lexicography

The dictionary entries discussed here are not inadequate entries, taken individually. They are carefully constructed and appropriately exemplified. But they could do more to tell the advanced learner of English about how the language works, and give a truer picture of the similarities and differences of semantic neighbours like *rattle* and *rumble*, if they were based on a systematic analysis of the verbs of sound emission, as summarized in Figure 1, where typical sentences for the “output” boxes are:

- Box [1] *The shutters rattled.*
 Thunder rumbled in the distance.
- Box [2] *The wind rattled the windows.*
 The windows rattled in the wind.
- Box [3] *She rattled the doorknob*

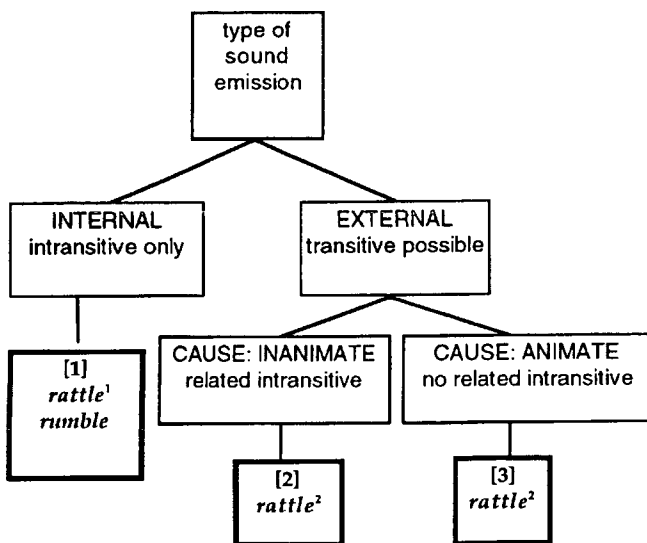


Figure 1. Sound emission verb schema

This approach would result in a set of related entries which are clearer and more informative than any currently available.

5. Conclusion

We have tried to give a systematic explanation of the transitivity alternations of one small semantic class of verbs whose entries in three corpus-based dictionaries do not explain clearly the facts found in the corpus. We believe that the systematically contrastive approach advocated here would form a sound basis for a fine-grained description of semantic groups in, for instance, a dictionary of synonyms or an electronic thesaurus. We know that in the commercial environment of dictionary-making, lexicographers have not enough time to do lexicography like this. One reason why publishers are wary of corpora is that lexicographers are reluctant to skim the surface of the citations, and find it frustrating to have to abandon the corpus evidence before the last little lexicographical tangle is unravelled. Theoretical linguists have a role to play in building a new-age dictionary, by setting out facts like these briefly and clearly, so that dictionary makers may apply them in their lexical descriptions.

Notes

- 1 Levin and Song's research was supported by US National Science Foundation Grant DBS-9221993.
- 2 See References for the full titles of these dictionaries and the Appendix for relevant extracts.
- 3 In the two cases marked with an asterisk, the animacy of the subject is left unspecified.
- 4 See Levin (1993) for more detail on such alternations.

References

- CIDE: *Cambridge International Dictionary of English*, Cambridge: Cambridge University Press, 1995.
- COBUILD: *Collins Cobuild English Dictionary*, London: HarperCollins Publishers, 1995.
- LDOCE: *Longman Dictionary of Contemporary English*, London: Longman Group UK Ltd., 1995.
- Levin, B. (1993) *English Verb Classes and Alternations*, University of Chicago Press, Chicago.
- Song, G. (In press) "Adicity, Causation, and Lexical Aspect", *Proceedings of ESCOL 1995*.

Appendix 1 Extracts from dictionary entries

rattle (*obj*) *v* • *The cups rattled* (=made a noise like a series of knocks) *as the waitress laid the table.* [I] • *The dying man's voice rattled in his throat.* [I] *The explosion rattled the cups* (=caused them to make a noise like a series of knocks) *on the shelf.* • *The car rattled* (=made a noise as it travelled) **along** *the road / down the street / over the cobblestones.* [I] • *My car engine is making a strange rattling noise.*

From: CIDE 1995

rattle 1	When something rattles or when you rattle it, it makes short sharp knocking sounds because it is being shaken or it keeps hitting against something hard. <i>She slams the kitchen door so hard I hear the dishes rattle ... He gently rattled the cage and whispered to the canary ... Somewhere close at hand a train rattled by ... The truck pulled away, and she listened to the rattling noises fade down the lane. ...</i>	v-erg v vn vadv v-ing
-----------------	---	---

From: COBUILD 1995

rattle v 1 [I,T] to shake, or make something shake, with quick repeated knocking noises: *The windows rattled in the wind. The beggar was rattling coins in an old mug.* **2** [I] to move quickly, making a rattling noise: [+**along / past / over etc**] *The cart rattled along the stony road. ...*

From: LDOCE 1995

rumble v [I] to make a continuous low sound • *Please excuse my stomach rumbling – I haven't eaten all day.* • *The tanks rumbled* (=moved slowly and noisily) *across the battlefield. ...*

From: CIDE 1995

rumble	A rumble is ...	
	2 If a vehicle rumbles somewhere, it moves slowly forward while making a loud, continuous, throbbing noise. <i>A bus rumbled along the road at the top of the path... A line of tractors rumbled onto the motorway through a cordon of police... The air reeked of kerosene and huge aircraft rumbled overhead. ...</i>	verb vadv/prep
	4 If something rumbles , it makes a low, throbbing noise. <i>The sky, swollen like a black bladder, rumbled and crackled... Speeches rumbled within the walls of the churches.</i>	verb v
	5 If your stomach rumbles , it makes a vibrating noise, usually because you are hungry. <i>Her stomach rumbled. She hadn't eaten any breakfast</i>	verb v

From: COBUILD 1995

rumble *v* **1** [I] to make a series of long low sounds, especially a long distance away from you: *We could hear thunder rumbling in the distance.* **2** [I always + adv/ prep] to move slowly while making this sound: [+ **along** / **past etc**] *A tank rumbled past.* **3** if your stomach rumbles, it makes a noise, especially because you are hungry. ...

From: LDOCE 1995