

Information Flow in Seneca and English

Author(s): Wallace Chafe

Proceedings of the Eleventh Annual Meeting of the Berkeley Linguistics Society (1985), pp. 14-24

Please see “How to cite” in the online sidebar for full citation information.

Please contact BLS regarding any further use of this work. BLS retains copyright for both print and screen forms of the publication. BLS may be contacted via <http://linguistics.berkeley.edu/bls/>.

The Annual Proceedings of the Berkeley Linguistics Society is published online via [eLanguage](#), the Linguistic Society of America's digital publishing platform.

Information Flow in Seneca and English
Wallace Chafe
University of California, Berkeley

This paper is about some differences between English and Seneca, an Iroquoian language spoken by about 500 people in western New York State. Specifically, it is about differences in how the flow of information through time is handled in the two languages. When I say "flow of information" I have in mind the stream of consciousness of the speaker and hearer, as ideas enter the speaker's mind and are subsequently communicated in some form to the hearer's mind, but I also have in mind the process by which those ideas are converted into language in order to accomplish the communication, the process of on-line verbalization. It is of some interest to compare Seneca and English because they are, in a variety of ways, as unlike each other as two languages can be. Their handling of information flow is one of the most conspicuous ways in which they differ.

In order to understand what I am talking about, you will have to join me in imagining that while a person is talking, first one piece of information and then another is activated in his or her mind. You will have to suppose, furthermore, that the amount of information that can be activated at one time is very small, relative to the total amount that is stored in the speaker's mind. And you will also have to suppose that one piece of active information is replaced by another at fairly regular intervals, it would seem about every two seconds (Chafe 1980:14). Finally, you will have to suppose that each piece of active information is verbalized in what I will call an intonation unit. A typical intonation unit is characterized by an initial pause, a coherent intonation contour ending in a clause-final cadence, and the possession of at least one word that is intonationally prominent. Schematically we can represent an intonation unit as follows, where the three dots indicate the initial pause, the x's indicate words, the accent mark shows that at least one of the words has primary stress, and the comma indicates any one of a set of clause-final intonations:

... x x x \acute{x} x,

The following is a sample of a brief sequence of intonation units in English:

1. ... and then another day,
2. ... it was really $\acute{h}ot$,
3. ... it was in the $\acute{s}ummer$.
4. ... and ... my $\acute{r}oom$ was $\acute{s}mall$.
5. ... it was like ... nine by $\acute{t}welve$ or something.
6. ... it seemed $\acute{s}pacious$ at the $\acute{t}ime$.

One of the things we can notice about these intonation units is that the limited amount of information verbalized in each of them is reflected in the limited number of words each contains. In a sample of about 2,000 English intonation units from casual conversation (see Chafe and Danielewicz, in press), I have found the modal number of words per intonation unit to be five. I interpret this strong preference for very short intonation units as reflecting a strong limitation on the amount of information that can be active in a speaker's mind at one time.

One of the ways the Seneca language differs from English is in being a classic polysynthetic language, which is to say that many of its words have quite a bit more information packed into them than do the words of English. If the just mentioned limitation on the amount of information that can be active in a speaker's mind at one time reflects a universal cognitive constraint, we might then expect that the intonation units of Seneca would be significantly shorter than those of English. That appears to be the case. In a sample of 830 Seneca intonation units in an informal conversation, I found the modal length to be two words rather than five. The following excerpt from this conversation (between two speakers labeled A and B) may provide some idea of what Seneca intonation units are like. Speaker A has been looking at the front page of a newspaper:

7. A. ... tho:há kyʔʔh,
almost they say

8. A. ... wá:tiʔskóʔt nẹ:kẹ:h,
they came to be in water these

9. A. ... honʔwé nẹ:kẹ:h,
it belongs to them this

10. A. ... ka:yátʔshəʔ.
paper

"It says here the owners of this paper,
they almost drowned."

11. B. ... ó:
oh

12. B. ... kweʔtháʔ aotíʔskoʔʔh.
almost they might have come to be in water

"They might have drowned."

13. A. ... Uh huh.
uh huh

14. B. ... ó:.
oh

15. A. ... ka:q̣wó? ne:? o?wátkaha:tho?.
boat it turned over

"Their boat turned over."

It appears, then, that languages may differ in the modal length of their intonation units when measured in terms of words, and that this difference is correlated, as we might expect, with their degree of synthesis, or roughly speaking the amount of information which their words contain (Sapir 1921:127-128).

It is natural to think of degree of synthesis in terms of morphemes per word (Greenberg 1960:185). In this case, the words of Seneca tend to contain relatively large numbers of morphemes, those of English relatively few. One might then suppose that a count of morphemes, rather than words per intonation unit might yield a more stable figure across languages. A preliminary count of small samples of English and Seneca suggests that this may be true, and that both languages tend to show something on the order of eight morphemes per intonation unit. Typical is intonation unit 6 from the English excerpt above:

6. .. it seemed spácious at the time.

Here there are six words but eight morphemes, if seemed and spácious are regarded as containing two morphemes each. Typical also is intonation unit 8 from the Seneca excerpt:

8. A. ... wá:ti?skó?t ne:ké:h,
they came to be in water these

where the first word contains seven morphemes and the second, one. Although further work is called for, it is at the moment a plausible guess that the number of morphemes per intonation unit is a more stable figure across languages than the number of words.

For several reasons, however, it is problematic to assume that morphemes are units of information. For one thing, it is obvious that some morphemes convey quite a bit of information, while others convey little, or even none. To put it in terms of old-fashioned information theory (Shannon and Weaver 1949), the predictability of different morphemes varies greatly. In the Seneca example just quoted, the morpheme represented as -o- near the end of the first word is high in information content, with the meaning "be in water," whereas the sequence -?sk- which immediately precedes it can be considered an "empty morph," present in this word only because the verb root -o- requires an incorporated noun root of some kind. When there is no other noun root, the slot must be filled by this empty one. A little more subtle is the fact that

the "factual" morpheme, represented by the wa- at the beginning of this word, requires the presence of the "punctual" aspect morpheme, represented by the -t at the end. In other words, given that the word began with wa- we can predict that it will end with -t, so that the -t conveys a minimum of new information in the sense that it is maximally predictable.

Another kind of problem is represented by intonation unit 11:

11. A. ... ka:yátqshæ?
paper

This word might be said to contain four morphemes, but it has a unitary meaning for a speaker of Seneca, just as the two-morpheme word "newspaper" has for us. In other words, the division of ka:yátqshæ? into four morphemes may have some etymological validity, but it can hardly be said to reflect any psychological segmentation into smaller meaningful units on the part of a speaker (or hearer).

Despite these problems, we may nevertheless be able to salvage some significance from the provisional finding that both an analytic language like English and a polysynthetic language like Seneca tend to have a similar number of morphemes per intonation unit. For it may well be the case that the relative distribution of the more informative and less informative morphemes in the two languages is the same. Although further research is clearly necessary, it would certainly be of some interest to discover that the number of morphemes per intonation unit is relatively stable across languages.

There is, however, another approach one can take to this question of how much information can be expressed in intonation units. I can sketch it only briefly here; a fuller discussion, limited to English, is available in Chafe (in press). It depends on the notion that a piece of information in a speaker or hearer's mind may be in any one of three different states of activation, which I will call "active," "inactive," and "semi-active." The distinction between active and inactive information may be clear enough already. Active information is that which a person is focusing on, which is in the center of a person's attention at the moment, whereas inactive information is that which is completely out of a person's immediate awareness. For reasons that will appear, however, it is useful to recognize a third, or intermediate, degree of activation, involving information that is in a person's consciousness to a peripheral degree. Information may be in this "semi-active" state for either of two different reasons. It may have been fully active at some earlier point in the discourse, and then have receded for a time into this state. Or it may be semi-active because it is inferrable from some other, related information that has already been activated.

Imagine now, again, a speaker who is activating successive chunks of information. Probably the process of changing a piece of information from the inactive to the active state uses up a

certain amount of mental energy, and the expenditure of that energy takes a certain amount of time. The process itself is what might traditionally be viewed as "recall," or retrieval from long-term memory. Probably at least one of the major reasons for the pause that typically begins an intonation unit is this process of activating previously inactive information. Thus, given our schematic representation of an intonation unit:

$$\begin{array}{c} \dots \text{ x x x } \acute{\text{x}} \text{ x,} \\ / \qquad / \\ \text{X} \qquad \text{Y} \end{array}$$

we can suppose that the activation process begins at point X (the beginning of the initial pause) and is complete at point Y (the end of the initial pause).

It is usually the case that, while some of the information that will be expressed in the upcoming intonation unit will have to be activated in this way, there will also be some information that is already either active or semi-active at point X. In other words there is usually some active, and perhaps some semi-active information that is carried over from one intonation unit to the next. I want now to suggest the hypothesis that during the period between X and Y no more than one concept can be changed from the inactive to the active state. When I say "concept" here, I mean either a referent, such as is conveyed by my room in intonation unit 4, or an event or state in which a referent participates, such as is conveyed by was small. We can call this the "one recall at a time" hypothesis. Its implication is that a particular intonation unit is able to express only one "brand new" concept, or only one concept activated from the inactive state, all others being concepts that were already active or semi-active at point X.

Let us review the English example with this hypothesis in mind. I will indicate in boldface the verbalizations of those concepts which, during the period just prior to the intonation unit in which they appear, were activated from the inactive state:

1. ... and then **another** day,
2. ... it **was** really **hot**,
3. ... it **was** in the **summer**.
4. ... and ... my **room** was **small**.
5. ... it **was** like ... **nine** by **twelve** or something.
6. ... it **seemed** **spacious** at the time.

That is, another day, be hot, be in the summer, be small, be nine by twelve, and seem spacious verbalized concepts that were previously inactive. The stressable syllable of each such item is given intonational prominence. It happens that the one intonationally prominent phrase in this example that does not express previously inactive information is my room. The speaker had been talking about her room shortly before, and it was a concept that

was semi-active at the beginning of the initial pause in 4. This example thus fits the one recall at a time hypothesis. Preliminary investigations of other English material suggest that the hypothesis is a fruitful one.

We can now look at the Seneca example from the same point of view. I will again give in boldface the words that express information recalled from the inactive state:

7. A. ... tho:há kyq?qh,
almost they say

8. A. ... wá:ti?skó?t ne:ké:h,
they came to be in water these

9. A. ... honq:wé ne:ké:h,
it belongs to them this

10. A. ... ka:yátqshæ?
paper

"It says here the owners of this paper,
they almost drowned."

11. B. ... ó:.
oh

12. B. ... kwe?thá? aotí?sko?qh.
almost they might have come to be in water

"They might have drowned."

13. A. ... Uh huh.
uh huh

14. B. ... ó:.
oh

15. A. ... ka:qwq? ne:? o?wátkaha:tho?
boat it turned over

"Their boat turned over."

I assume that the particles and responses that constitute 7, 11, 13, and 14 do not express recalled information, and are thus irrelevant to the present discussion. It is only 15 that might seem to contradict the one recall at a time hypothesis, since it contains both the referent expressed as ka:qwq? "boat" and the event expressed as o?wátkaha:tho? "turned over." Are these in fact two separate concepts that were recalled during the initial pause of 15? I suggest that the speaker was conceptualizing here a unitary "capsizing" event, as in fact is signaled by the stress on

bóat and not on turned over in the English equivalent the bóat turned over (Chafe 1974:115-116). This boat-turning-over concept was in fact a good candidate for noun incorporation, the ultimate way of expressing conceptual unity in Seneca. The speaker might, that is, have said o?wathowókaha:tho?, combining the boat and the turning over within a single word. It is a symptom of the moribundity of Seneca that incorporation is a less and less frequently used device. The inclusion of the noun and verb in a single intonation unit such as 15 might, in fact, be regarded as a kind of poor man's incorporation, easier to accomplish and with similar effect.

Let us return now to intonation units 7-10, and take note of a discrepancy between the original Seneca and the invented, but I think quite reasonable English translation. In the Seneca there are four separate intonation units, whereas the translation (given just below 10) contains only two. It would be awkward in several ways to translate more literally:

"It says here,
they drowned,
it belongs to them,
the paper."

Just now I am concerned with what appears to us to be the excessive fragmentation here. Later we can return to the unusual (for English) order in which the concepts are presented.

There are two reasons why Seneca has more intonation units in 7-10. One is that Seneca is a language in which it is more common than in English to present an evidential or epistemological orientation in a separate intonation unit, as in 7. We actually find a similar phenomenon in written English, where words or phrases like however or in any case are often set off by commas, but spoken English does this less often. Seneca speakers, on the other hand, frequently devote a separate intonation unit to a string of particles. The following example is more extreme than 7:

16. ... ne:? nǎ: shɔ: kho ni:? ne
 assert contrast just also I the

hoqó:kwa: kwe?tha?
over there almost

"And I almost just the one from over there."

This intonation unit fails to contain a single spelled out noun or verb, but from it the hearer understood that the speaker was inclined to subscribe only to the Jamestown newspaper. Perhaps this way of speaking is especially characteristic of small speech communities in which there is a great deal of shared knowledge.

The other device that distinguishes 7-10 from its idiomatic English translation is the expression of "the owners of this paper"

as a single concept in English, but as two in Seneca: essentially hono:wə "they own it" and ka:yátəshə? "paper." It appears that the concept of newspaper owners is a unified concept in English, but one that Seneca speakers must create synthetically. And it must be presented in Seneca in two separate stages: first the idea of ownership, then the idea of the paper. I am suggesting here a kind of Whorfian difference between the two languages, with English expressing the idea of newspaper owners with a minimum of effort, Seneca expressing it with more difficulty. (It is interesting to contrast the effortful expression of this concept with the unified idea of the boat turning over in 15, as discussed above.)

So far I have dealt with various limits on the amount of information that can be included in a single intonation unit in English and Seneca. I have shown that the two languages differ in number of words per intonation unit, but that they appear to be similar in the number of morphemes. I have suggested that both languages are subject to the one recall at a time hypothesis, to the effect that an intonation unit can contain no more than one concept that has been recalled from the inactive state. And I have shown that what is treated or is treatable as a single concept may differ between the two languages. Now I would like to turn from the amount of information expressed in an intonation unit to the order in which the information is presented, both within and across such units. Again we will find interesting differences.

I will begin by assuming that all languages share the strategy of presenting information by bringing up some concept and adding it to some other concept that functions as a starting point. In other words, information is not verbalized in a vacuum, but is attached to other information. The starting point appears as what we generally know as the subject of a clause, the added information as the predicate.

English, as we know, typically expresses the starting point, the subject, at or near the beginning of a clause, and then goes on to add the predicate. Seneca does it differently. The typical Seneca strategy is to begin with an epistemological orientation expressed through one or more particles, sometimes as a separate intonation unit as in 7, sometimes in the same intonation unit as what follows, as in 12. What then follows these particles may be either a subject or a predicate. In a general way the choice seems to depend on which of the two is the most "newsworthy" (Mithun, in press). The following example may give some idea of how this kind of choice operates. The speaker was talking about trains:

17. ... ne:ʔ o:nə o?gáʔse:ʔ,
 assert now it pulled

18. ... okwe:nyó: ne ɔ:kwe ɛyɔteʔnóhtə:h.
 it's possible the people they will get in

"Then they pulled, and people were able to get in."

19. ... ta: ne:? ti ne:kə: ae? niyáwəʔoh,
so assert switch this again what happened

20. ... ne:ne:? twə:níshəte:nyøk,
those in those days

21. ... ne:? ne:kə: ne yatátehso:t,
assert this the he and his grandmother

22. ... wa:yateʔnóhtə:h.
they got in

"So what happened in those days was that a boy and his grandmother got in.

23. ... ne: kaʔsehtowáne:neh.
this in the big car

"In this big car."

24. ... kaʔsehtiyánowéʔkeh.
in the fast car

"In the train."

25. ... wa:yateʔnóhtə: yatátehso:t.
they got in he and his grandmother

"The boy and his grandmother got in."

Intonation units 21 and 22 together express a clause in which the concept of the boy and his grandmother was the subject, that of "getting in" the predicate. The concept of the boy and his grandmother was brand new, previously inactive information, and when such information is presented as a subject it is necessarily allotted its own separate intonation unit. The concept of getting in was at this point semi-active information because it picked up on something already stated in 18. It was less newsworthy than the subject, and thus appeared second.

In 25 the situation was reversed. Here both concepts were semi-active, having both been verbalized before, but having both lapsed into the semi-active state while the speaker focused on the car and the train in 23 and 24. The degree of activation of these two concepts was equal, and was thus irrelevant to determining the order between them.

It is probably the case that when a subject and predicate share the same degree of activation, as in 25, Seneca speakers generally treat the predicate as more newsworthy, expressing it before the subject. This hypothesis also explains the ordering of 8-10, where all the concepts involved were previously inactive, and where there were in fact two cases of predicate preceding subject:

When both the subject and predicate are equivalent in their degree of activation (either both previously inactive as in 8-10 or both previously semi-active as in 25), the issue is decided by a default preference for the predicate as the more newsworthy of the two. However, when a subject and predicate together constitute a single unitary concept, as in 15, there is an arbitrary imposition of a frozen subject-predicate order.

These seem to me to be at least some of the principles that govern the flow of information in this language. The principles cover quite a bit of the data I am familiar with, though doubtless they will need to be modified and extended before they will account for all the relevant facets of Seneca discourse. In the meantime, I hope they may extend somewhat our arsenal of discourse facts and explanations.

References

- Chafe, Wallace. 1974. Language and consciousness. *Language* 50:111-133.
- . 1980. The deployment of consciousness in the production of a narrative. In Wallace Chafe (Ed.), *The pear stories: cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.
- . In press. Cognitive constraints on information flow. In Russell Tomlin (Ed.), *Coherence and grounding in discourse*. Amsterdam: John Benjamins.
- Chafe, Wallace, and Jane Danielewicz. In press. Properties of spoken and written language. In Rosalind Horowitz and S. J. Samuels (Eds.), *Comprehending oral and written language*. New York: Academic Press.
- Greenberg, Joseph. 1960. A quantitative approach to the morphological typology of language. *International Journal of American Linguistics* 26:178-194.
- Mithun, Marianne. In press. Is basic word order a universal? In Russell Tomlin (Ed.), *Coherence and grounding in discourse*. Amsterdam: John Benjamins.
- Sapir, Edward. 1921. *Language: an introduction to the study of speech*. New York: Harcourt, Brace.
- Shannon, Claude E., and Warren Weaver. 1949. *The mathematical theory of communication*. Urbana: The University of Illinois Press.