8

Some Sentences on Our Consciousness of Sentences

Thomas G. Bever and David J. Townsend

"My end is my beginning."
—Luciano Berio, 1975

"All the . . . furniture of earth . . . have not any subsistence without a mind."
—George Berkeley, 1710

"Consciousness . . . [is] . . . a river. . . . Like a bird's flight it seems to be made of an alternation of flights and perchings. . . . Consciousnesses melt into each other like dissolving views."
—William James, 1890

"Instead of . . . a single stream [of consciousness], multiple channels . . . do their various things, creating Multiple Drafts as they go . . . some [drafts] get promoted to further functional roles . . . by the activity of a virtual machine in the brain."
—Daniel Dennett, 1991

"I see everything twice! I see everything twice."
—Joseph Heller, 1961

Overture: How Many Times Do We Perceive a Thing?

We have a problem common to those who depend on eyeglasses: each morning, there is a ritual hunt to find where we left them the night before. The Catch-22 in this, of course, is that without our glasses on, it is hard to see them, since they are a composite of spidery metal and maximally invisible lenses. We ritually wander around the house checking my habitual domestic locations, like a dog on optimistic visits to its most likely backyard bone mounds. Inevitably, we find the glasses (the first day that we do not do so may be our last).

The fact that this is an everyday perceptual event has allowed us to analyze it into those parts of which we can be conscious. We are sure that when we finally find them, we "see" the glasses twice. The first time is fleeting, but with a clear memory of an event that appears "preperceptual." We believe we saw them the first time only after we see them for the second time, a split second later. In ordinary life, the second percept is what we think of as the real percept. The first percept is almost immediately suffused by the second, and is primarily conscious as a memory.

The phenomenological difference between the two representations is striking. The first is a foggy mixture of a skeletal configuration and one or two clear parts—an earpiece, or nosepiece connected to a chimerical outline of the rest—the combination of global and particular is "the glasses" at a conceptual level rather than in visual detail. The second representation is explosive, sharp and complete overall, no part stands out more than the others, it has suddenly become the coherent visual entity, "glasses." The formation of this representation is the perceptual event that we normally would describe as, "Then, we found our glasses."

Fortunately, we are not aware of seeing everything two times. The daily eclectic search for my glasses is ingrained, and perhaps has allowed us to peer into its stages more thoroughly than into the random events of quotidien life. The cautious reader (or a scholar of the downfall of Titchnerian introspectionism) may object that our memory is tricking us, that the first "prepercept" is merely a shadow of the in a retrospective backwash. Perhaps so, but the fundamental point would remain potentially true of all perceptual acts: we perceive everything twice, once as a metonymously schematic object, once as clear image of the object.

Let's turn to sentence comprehension and see what it might tell us about the possible reason for multiple representations in perception. What follows is a brief précis of our just-published book (Townsend and Bever, 2001). We show that the long-neglected analysis-by-synthesis model of comprehension can offer an explanation for the role of successive representations in sentence comprehension. At the end of this summary, we suggest some hypothesis about the implications of this for consciousness of sentences and other objects.

Theme 1: Sentences Sound Clearer

Cl#ss#c#l s#m# th#n n r#nd#m t th# s#nt#nc#s a th#t s#mpl#rr #s wh#n w#rds f#nd#ng #n w#rds ndrst#nd #r# t# #rd#r. That is, #t #s a cl#ss#c#l f#nd#ng th#t w#rds #n s#nt#nc#s #r# s#mpl#r t# r#c#gn#z# th#n th# s#m# w#rds #n r#nd#m #rd#r. When the words are presented in a noisy acoustic background to make the perception difficult (or with visual noise added as in the preceding sentences), people get more words right when they are in sentence order. But most important for this discussion, the words also sound (or look) physically clearer as one takes them in and recognizes them.

The apparent sensory punch of words-in-sentences may not be surprising, for a number of reasons. First, sentences carry meaning, which we can relate to separately existing knowledge. It is intuitively reasonable that accessing sentence meaning provides a redundant source of information about the words. An empirical problem with this explanation is that it is also relatively easy to recognize words in sentences with little meaning ("The slithey toves did gyre and gymbal") or even with counterintuitive meaning ("The electrolytic mailman bit the religious dog"). Furthermore, even if it is intuitive, it is a scientific mystery how accessing the meaning at the sentence level could enhance the physical clarity of the individual words.

It is also a fact that sentences have formalizable syntactic structure that might provide a different kind of redundant information enhancing the acoustic clarity of the words. In "the big city-i was rich", there is a surface hierarchical segmentation (indicated by the relative size of spaces between words). In many theories, there is also an abstract level of representation that links distant phrases and canonical locations together. This can underlie important differences in sentences that appear the same superficially: in "the big city-i was attacked [i]", "city-i" is linked to the usual patient position immediately after the verb by the abstract coindexed trace element, "[i]." This dual representation for "city" captures the intuitive fact that it is both the topic of the sentence, and the patient of the verb. There are numerous formal models of how this linking is computed. Many models presuppose a form of "derivation"—in the most transparent case, the sentence formation involves a stage in which the words are initially placed in the marked location and subsequently moved to their apparent location (e.g., "city" would be originally in postverbal patient position of "attack" at a point when "attack" has no subject, and is then moved to the beginning of the sentence). This movement creates the notion of "derivation," in which the structure is built and deformed in a formal sequence of operations. Some models overlay the distant relations between phrases with a different kind of structure, but they are arguably the equivalent of the more transparent kind of derivational sequence of operations.

In any case, all sentence structure models involve considerable abstract symbolic manipulation. This complexity is a mixed theoretical blessing if our goal is to use syntactic structure as the explanation of why sentences sound clear. On the one hand, syntactic structure might afford another kind of representation that relates the words to each other. But it also involves a lot of computational work, which one might expect would actually take mental resources *away* from recognizing the words. And, as in the case of meaning, the syntactic model itself does not explain how syntactic relations between adjacent words and distant phrases enhance the clarity of the words themselves.

Theme 2: Syntax Is Real

Syntax accounts for an essential mental fact: with little or no training, people can distinguish grammatical from ungrammatical sentences in their language, often regardless of meaning. English speakers know that there is something wrong with "The electrolytic mailman bite the religious dog" which has nothing to do with its unusual meaning. On occasion, syntacticians rely on extremely subtle or unconvincing grammaticality discriminations; this often tempts psychologists to dismiss the mental relevance of the entire linguistic enterprise. But the clear cases of grammaticality distinctions remain, and support the relevance of syntactic structures and at least the broad outline of how they are computed.

Psychology aspires to be an experimental science, so despite the many mental facts syntax accounts for, psychologists have doggedly attempted to prove that syntax is "really" real. The goal is to show that syntax plays a role in all language behaviors, not just rendering grammaticality intuitions. Jacques Mehler's dissertation (Mehler, 1963) is a classic example, demonstrating that the abstract level of syntactic structure plays an important role in how sentences are remembered. Other studies (many by Mehler as well) revealed that both surface and abstract structures play a role in ongoing sentence comprehension.

For this discussion, the most important aspect of linguistic structure is the postulation of an abstract level at which structurally related phrases are adjacent, before they are moved to distant surface positions. This movement leaves behind a coindexed trace as a relic of the original location and relation (or a silent copy of it in some syntactic theories), and thus the complete syntactic structure provides two representations of such an element. An empirical prediction of this is that elements that have an extra trace should be relatively salient during comprehension: For example, "big" in "the big city-i was attacked [i]" should be more salient than in "the big city-i was rich." This is the case: a variety of experiments have shown it takes less time to recognize that "big" was in trace-sentences like the first than in trace-free sentences like the second. Most important for this discussion is the fact that the relative saliency of elements with trace becomes clear only after the proposition is complete: probes presented right at the trace position itself generally do not show the relative saliency immediately.

The struggle to prove the "psychological reality" of syntax is not over. Grammatical formalisms change and compete like weeds. The hapless experimental psychologist is often left behind in the formal dust. We have chosen the experiments on "trace" because that way of representing phrase movement from one level to another has corresponding formalisms in all versions of transformational grammar. Indeed, while the comprehension trace experiments have appeared only in the last decade, the corresponding predictions could have been equally well motivated by the 1957 model of syntax in *Syntactic Structures* (Chomsky, 1957). (E.g., the critical word would be represented both in the "kernel" and passive "surface" sentence structure, and hence more salient psychologically. Such studies were not done then, perhaps because of the fact that "cognition" in the sixties was preoccupied with memory and not perception.)

The upshot of much of the last forty years of psycholinguistics since Mehler's dissertation is that comprehension involves assignment of syntactic representations. Any model of language understanding must take this into account.

Theme 2, Variation: Associations Are Real

The cognitive revolution of the 1960s was a triumph of rationalist structuralism over behaviorist associationism. Abstract structures, such as the initial level of representation of sentences, were recognized by all as a dangerous challenge to the "terminal metapostulate" of behaviorism—the claim that all theoretical terms and all corresponding inner psychological entities are "grounded" in more superficial entities, ultimately in explicit behaviors and stimuli. There is no obvious way in which inner syntactic forms are extracted from outer ones. Indeed, syntactic theory suggests that the relation is in the opposite direction. This sharpened the focus of research on the behavioral role of such inner forms, and as the evidence mounted, behaviorist structures lapsed into the obscurity of irrelevance.

But arguments against behaviorism are not direct arguments against associationism. Associationism is merely the doctrine that all mental relations are associative rather than categorical, usually built up out of repeated experiences or co-occurrences. And the apparent arguments for associationism are compelling. No one can deny that a vast proportion of everyday life is based on habits. We rely on them all the time, we see how we can train them in animals, and we use the notion as a powerful explanation of many forms of natural behavior.

One of the arguments against the idea that absolutely everything is based on associative habits is the complexity of normal behaviors, such as the comprehension of language. Furthermore, the existence of inner levels of representation would seem to defy associationistic explanation. The traditional answer by associationists has been that we underestimate the subtlety of behavior that a large complex associative network can compute. Even inner representations might be connected by myriads of associations so complex as to mimic the appearance of hierarchies, part-whole relations, relations at a distance, and other kinds of categorical facts. This is the thrust of a major school of modern connectionism—to apply relatively complex networks of associatively connected units, and thereby explain away apparent categorical and rulelike structures as the result of "pattern completion" and the exigencies of associative information compression. The idea is that with enough experience, associative networks produce behavior that looks structural in nature, but we allegedly "know" it really is not because we manufactured the machine ourselves.

That, of course, is just one of many human conceits. We manufacture electric generators, but we do not make the laws or the electrons that the laws govern—rather, we construct machines that manipulate those phenomena to our own ends. Similarly, we might someday wire up an associative network as massively interconnected and complex as a human brain, and watch it learn language. The language-learning process in the machine would be as much a mystery as in a human child. We would be back to the 1960s' cognitive psychology joke: "someday the artificial intelligencers will create a plausibly human robot, and then we will have to practice psychophysics, theoretical psychology, and experimental "neuro"science on it to figure out how it works." We will have a new division of international psychological associations: "robopsychology."

Similar modesty is indicated about the mimicking achievements of today's modest connectionist models. The ability of a model to converge on 90 percent accuracy in computing a specific syntactic structure after thousands of training trials, or to differentiate lexical classes based on repeated exposure to local contexts, is an important achievement. It stands as an existence proof, that statistically reliable information is available in the language stimulus world that can support the induction of recognition patterns. But to claim that such achievements show we can account for actual categorical syntactic structure is not warranted. It would be like claiming that the amazing height of a medieval church spire shows that humanity can reach the heavens.

Yet, whatever the limits of associationism, habits exist and dominate most of life. Any adequate model of comprehension must find the appropriate computational locus for their operation and influence.

Development: Analysis by Synthesis

We are left with two truths. most of the time we do what we usually do, using experientially based superficial habits. But sometimes we create novelty, using categorical computations that reflect sequential symbolic operations. How can we reconcile these politically opposed, but valid approaches to language behavior in a model of perception?

It is fashionable today to assert that the best models of behavior are "hybrids" between connectionist and symbolic systems. Such models combine the two kinds of approaches by definition, but the real test is the particular architecture that does this. The architecture is constrained by several features:

1. Associative information operates on relatively superficial representations and is immediately available.
2. Readily available surface information includes the lexical items in sequence, a rough phrasing structure, and a likely meaning.
3. Syntax is derivational—it involves a series of steps in building up sentence structure, which can obscure the initial computational stages in the surface form.

The first two points suggest that comprehension can proceed directly from surface cues to a primitive syntax, and associative relations between words and phrases that converge on particular meanings. This is exactly what many connectionist modelers assume, as they build their toy models and extrapolate them to the entire language. For example, the sentence "the city was rich" might be interpreted directly, based on separate analyses of the words, and a rough phrase structure, as "NP be PREDicate." "Rich" is a lexical item which carries the semantic information that its syntactic subject is semantically a stative experiencer. Similarly, "the city was ruined" can be analyzed in the same way. Though it looks like a passive form, "ruined" is actually a lexical item, which acts like a normal adjective; for example, it can modify nouns, as in "the ruined/rich city lay before us." Accordingly, the word itself carries the semantic information that the syntactic subject is a stative patient.

"The city was attacked" can be initially understood on the same superficial template, "NP be PREDicate." But the syntactic analysis is actually wrong, even if the associated meaning is roughly correct. "Attacked" is not a lexical item, as reflected in the oddity of "*the attacked city lay before us" (contrast that with the acceptable, "the ruined/rich city lay before us"). Thus, the correct syntactic analysis should reflect the movement of "city" from patient to subject position, as represented in the surface form, "the city-i was attacked [i]." What is a possible model for assigning this derivationally based syntactic structure?

The derivational nature of syntax makes it difficult to go back from the surface form to the original compositional stages. Like an airplane, the grammar runs in "forward" only. And, as with an airplane, the only way to retrace where it started and how it arrived at a current position is to go back to a likely starting point and recapitulate the potential journey, marking the route and checking to make sure it arrives at the correct destination. In the case of syntax, this means re-creating possible derivations from likely starting points, and checking to be sure they arrive at the right surface structure. The initial analysis provides a number of pointers as to the likely starting configuration. Individual lexical items are recognized, as well as major phrases, along with a potential meaning that relates the phrases in functional categories. This sets rich constraints on the possible starting configuration and derivation to a surface form. In fact, it corresponds well to the initial derivational stage in current "minimalist" theories of syntax, which start with a "numeration" of lexical items and functional categories, and build up the surface form from there (see Chomsky, 1995).

So, we have two phases in comprehension: an initial analysis of a likely meaning, and a recapitulation of the complete derivational syntactic structure consistent with the form and meaning. This invites a comprehension architecture which proceeds in several major stages:

a. Analyze the string into lexical sequences broken into major phrases.

b. Analyze the structure in (a) for a likely meaning using canonical syntactic patterns and associative semantic information.

c. Take the output of (a) and (b) as input to a syntactic derivation.

d. Compare the output of (c) and (a): if they are identical, then the meaning in (b) and the structure in (c) are correct. If they are not identical, reanalyze (a) and reinitiate the process.

In words, this is equivalent to a traditional "analysis-by-synthesis" model, which enjoyed much currency for a while. On this model, candidate derivations are computed and matched to a temporarily stored representation of the surface input. When there is a match, the derivation is assigned as part of the overall representation. Such models account for the active nature of perception, but most have had a major flaw that contributed to the loss of interest in them. Since the starting form of a derivation is "abstract," it is unclear how a surface analysis can constrain potential derivations to be anywhere near correct. Our current version lacks this weakness: the analysis into lexical items and functional categories gives exactly what is needed to start a derivation in minimalist syntax that is likely to converge on the correct surface analysis.

In principle, the model works. The question is, does it correspond to real facts about comprehension? Our book presents the model and an eclectic compilation of existing and new data in its favor. Some salient features of the model and supporting arguments and predictions that we note are the following:

1. A complete syntactic structure is assigned fairly late in the process. Past studies of the formation of a complete phrase structure have suggested that it is assigned only after a fraction of a second. For example, mislocation of interrupting "clicks" is controlled by fine details of phrase structure only after a brief interval. Similarly, as mentioned above, evidence for the saliency of movement traces occurs only a few hundred milliseconds after their actual position. Finally, ungrammaticality based on movement constraints is detected relatively slowly (both in explicit grammaticality judgment tasks, and reflected in evoked brain response patterns). All three facts follow from the view that the complete syntactic details, including derivationally based traces, are assigned as part of the syntactic recapitulation of structure.

2. Initially, the meaning is computed from an incomplete, and sometimes incorrect syntax. A number of researchers have suggested that meaning can be derived from syntactically deficient analysis. Not only is this an assumption of much of connectionist modeling and also an assumption of a number of current symbolic comprehension models, it also appears to be characteristic of much comprehension by children. A typical and frequently cited example is the idea that passives are initially comprehended as "NP BE Predicate," in which the PREDicate is itself a complex adjective. This coordinates exactly with the proposal we have made. In general, it is striking to note that constructions with Nounphrase trace are just those that also have parallel nontrace constructions, as below:

The city-i was attacked [i]: the city was ruined
The city-i was likely [i] to attack: the city was eager to attack
The city-i appeared [i] to attack; the city rose to attack

This may be a chance matter, but it is also predicted by our model: Nptrace can occur only in constructions which can be initially understood by the (mis)application of an independently supported perceptual template.

3. We "understand" everything twice: once on the initial pass, and once when assigning a correct syntactic derivation. First, one might ask, If the initial meaning is successfully derived from surface cues, why does the system recapitulate a correct syntax at all? If associative and preliminary structural information is sufficient for discovery of a meaning, why gild the lily? The answer is reflected in the fact that Dehli copra pricing techniques are quite ununderstandable . . . well, no, it is actually reflected in the fact that you were able to understand that proposition about Delhi and copra prices, despite the fact that the syntax is odd and the meaning is even odder and totally out of context. The functional role of recapitulating the syntax and checking the meaning is to make sure that the initial comprehension was correct. Otherwise, we would be consigned only to understand frequent constructions that convey common and predictable meanings that make sense in the local context.

The formation of dual meaning representations is the most theoretical claim at the moment, and is largely untested. It certainly is intuitive that we understand sentences immediately as they are apprehended. Thus, the claim that we initially understand sentences online substantiates the first pass in the model. The second phase of forming a complete syntax and meaning is more controversial, but subject to empirical test. A simple demonstration is the fact that it can take some time to realize that a superficially well-formed sentence is defective. Consider your intuitions as you peruse a run-on sentence like "More people have gone to Russia than I have." At first, it seems plausible and you think you understand it. This follows from the fact that it meets a superficial template and appears to have a meaning. But then, as you reconstitute it, it does not compute, and you realize that it is not actually a sentence, and you are actually quite confused about what it really means.

There are also predictions of the two-phase model that we can test experimentally. For example, it predicts that we go through two phases of understanding syntactic passive sentences like "the city-i was attacked." In the first phase, we actually misconstrue the city as experiencer of a stative predicate; in the second phase, we understand "city" correctly as the patient of an action. We are currently testing this prediction in a variety of ways.

Recapitulation: Sentences Sound Clear because Comprehension Involves the Formation of Two Surface Representations

We now return to the classic and enduring fact about sentences mentioned in the first section: their constituents actually sound extraclear. We note that the analysis-by-synthesis model affords a direct explanation of this. In comprehension we form and compare two surface representations of the sentence: one as part of the initial analysis, one as a result of the syntactic reconstitution. Thus, comprehension contributes an extra surface representation to the perception of the word sequence, as well as adding a check for the identity of the two representations. While the details await more specification, the model involves multiple manipulations of surface representations, which affords an explanation of why the sentences seem so clear physically.

Coda: Is Recapitulation of Object Representation an Important Component of "Consciousness?"

Our finale goes back to our overture. We have sketched a comprehension model which chases its tail, at least once—it starts with a meaning, and then uses that to begin a syntactic derivation which rearrives at the same surface form and enriched meaning. This presents the comprehension process as a set of emerging representations which converge. We have sketched the representations as occurring in series, but, of course, they could be partially computed in parallel in many actual cases. In this respect, we can infer that our model of comprehension is an instance of Dennett's "multiple stage" model of consciousness (Dennett, 1991). We think we understand a sentence and are conscious of it as a unified experience. Yet, analysis and experiment suggest that this apparently unified process is actually composed of converging operations of quite different kinds.

We are left with the puzzle of why we phenomenologically give priority to the final stage of processing. Why does the "real" perception of our eyeglasses of which we are conscious seem to be the final one? Why does the "real" comprehension of the sentence seem to be the one that includes a syntactically complete representation? There are several possible answers.

- The "latest" draft seems to be the most real by co-opting the earlier ones.
- The most complete draft seems to be the most real.
- The draft that depends most on internal computations seems to be the most real.

In a real machine, these explanations tend to co-occur. But I maintain the top-down computational bias Mehler and I acquired in the heady days of the sixties "cognitive revolution." So we favor the last alternative. The appearance of reality depends most strongly on internally generated representations of it.

This is not mere rhetoric. It can be tested. And probably will be. As in other cases, the precision of linguistic theory allows us to quantify the computational complexity of sentences. In this way, we can test the relative effect of complexity on conscious clarity of the words in sentences. The prediction is bold: the more complex a sentence is, the more acoustically clear the words in it (so long as it can be understood). This is somewhat counterintuitive, which is all to the better if it turns out to be true.

Thus, we end at our beginning. Jacques, salut, and many more!

References

Chomsky, N. (1957). *Syntactic structures.* The Hague: Mouton.

Chomsky, N. (1995). *The minimalist program.* Cambridge, MA: MIT Press.

Dennett, D. (1991). *Consciousness explained.* Boston: Little, Brown.

Mehler, J. R. (1963). Some effects of grammatical transformations on the recall of English sentences. *Journal of Verbal Learning and Verbal Behavior, 2,* 346–351.

Miller, G. A., Heise, G. A., and Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology, 41,* 329–335.

Miller, G. A., and Isard, S. (1963). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior, 2,* 217–228.

Townsend, D. J. and Bever, T. G. (2001). *Sentence comprehension.* Cambridge, MA: MIT Press.