All language understanding is a psycholinguistic guessing game – Explaining the still small voice

T.G.Bever
University of Arizona

In Anders, P. (Ed), *Issues in the present and future of reading.* pp. 249-281. Routledge

BACKGROUND

The problem of reading seems straightforward: readers learn to "decode" the visual input into a linguistic representation, and 'then' use that representation as input to normal mechanisms of spoken language understanding. So, to teach effective reading, all we need to do is inculcate an initial linguistic representation of the visual symbols, and language understanding mechanisms will do the rest. Phonics and "whole word" training are examples of this teaching paradigm: teach a kid the sounds of letters or of whole words, and hiser knowledge of auditory sentence processing will do the rest. Simple support for this view is the common idea that as we read, we 'hear' an internal rendering of what we are reading. That is, reading involves first decoding what we see into sounds and then applying our normal processes of speech comprehension to those internally generated sound sequences.

Four decades ago, Ken Goodman published a short article that sparked a revolution in this kind of thinking about reading education and research. The main point was that reading proceeds by the simultaneous integration of all of the reader's linguistic knowledge in ways that affect even the perceived input. That is, there is *no* single level of representation with a necessarily complete mapping from the visual to the imagined acoustic/linguistic world. Letter sequences do not literally force an interpretation; rather, letters and bits of letters are input cues to a reconstructive process, which creates linguistic representations of words, phrases, sentences and their meanings. At first, this view does not seem radical: it is superficially consistent with the view that information flows upward so that information at each level triggers an organization at the next:

/t,h,e, ,d,o,g, ,b,a,r,k,s, ,l,o,u,d,l,y/ ->

the, dog, barks, loudly, ->

(the)det (dog)Noun (barked)verb (loudly)adv ->

((the)d (dog)N)np ((barked)v (loudly)adv)vp ->

(((the)d (dog)N)np ((barked)v (loudly)adv)vp)s

WOOF!

Such models allow for the possibility that information at a given level is incomplete, but just complete enough to trigger the correct interpretation at the next level. For example, with rapid reading the input at the level of letters might be initially incompletely perceived as:

/t,x,e, ,d,x,g, ,x,a,r,k,e,d, ,l,o,x,x,l,y/

This deficient representation would seem still to have enough information to trigger the right words most of the time. Thus, the idea that reading involves accessing multiple levels of linguistic knowledge, was not in itself radical.

Ken's radical idea was that the flow of literary information is not uniformly upward: rather, it is cyclic such that information at each level can inform processes at levels below it. Thus, the following choice of words could have been triggered by the deficient letter sequence representation:

/toe dig marked lovely/

This in turn could trigger a syntactic representation:

(toe dig) (marked lovely)

But, assuming this rather odd syntactic structure, the next level would block the interpretation, because in fact *it makes no sense.* Again, this seems unremarkable on a traditional view – of course, a word sequence that makes no sense is not going to be the one that readers tend to arrive at. But the critical idea is that even the decision about what the reader is 'seeing' at the level of letters is itself modified by higher levels such as the associated meaning that can be built. That is, in the limiting case the reader is using the meaning to modify the perception of the input. The reader is "guessing" what the meaning will be, even when incomplete, and is using the guesses to influence what s/he sees in the letters.

In the same paper, Ken illustrates how powerful the process is by examples of 'miscue analysis' (developed over many years with Yetta Goodman), analysis of errors in young readers that reveal the process as it is building. Examples of such errors show that readers indeed project expectations ahead when they are reading, expectations so strong sometimes that they replace some of what is actually written with different words. But what is striking is that generally the replacement maintains the general meaning of the original: in our example, a young reader might utter:

“the dog was barking loudly”

or

“the dog barked aloud”

or

“the dog barked a lot”

The mere observation of such top-down effects might also seem unremarkable, since they maintain most aspects of the letters, and the meaning of the errors is consistent with the input. But the significant mystery on a bottom-up view of processing, once one notices it, is that the meaning is being projected *ahead* of the reading utterances – how can that be?

Noticing a problem becomes scientifically important when one also notices a possible solution to it. Ken’s solution was to invoke an “analysis by synthesis” model of reading comprehension, outlined below. On this model, readers use the literal input to trigger an initial ‘guess’ about the representation of the sentence at least at an inner level of linguistic representation: that representation then triggers a mapping onto a likely expected surface sequence: the reader matches the expectation to cues from the input. This explains why the mis-readings tend to maintain some of the letter sequences, while also maintaining the meaning. The reader does not hallucinate the entire input based on built-up guesses: rather, s/he uses the guesses to make sure that enough of the letters are accounted for in a coherent meaning.

This interpretation of fluent reading of course has had many implications for reading education and related research. It also explains the phenomenology of silent reading for most of us (St. Augustine’s famous discovery): most people ‘hear’ or imagine a ghostly voice that tracks the input. It has always been a mystery how the talking ghost can speak so fast, (easily 300 wpm for many readers, far faster than any normal speaker) elide over whole phrases, then suddenly alight on the careful internal pronunciation of a single word, and speak with proper phrasing and intonation. Ken was aware even of this conundrum and noted the relevance of the fact that we *can* understand acoustically compressed speech. But the fact remained that the internal speech is not real, yet has the muttered cloak of rapid and linguistically organized reality. How can this be?

ALL LANGUAGE COMPREHENSION IS RECONSTRUCTIVE

One answer is that normal comprehension of spoken language also involves reconstructive formation of mental representations of what we are hearing. That is, even when we are listening to an acoustic speech input, we recreate our own mental representation of it, even at the acoustic level – we don’t notice it as a ghostly echoic

voice (most of the time) in part because the actual acoustics of the input shapes it, and in part, *because our automatic reconstruction of the input directly influences what we think we hear.*

The reconstructive aspect of language comprehension starts at the ground level. Consider the perceptual recognition of simple sounds like /ap/, /at/, /ack/. Say these to yourself or ask a neighbor to do so, without releasing an aspirated consonant at the end, just a sudden stop with the mouth forming the silent /p/ /t/ or /k/. What is remarkable is that each silence is rendered as an internal 'sound' that you believe you (or your neighbor) actually uttered aloud: yet what you 'hear' corresponds to physical *silence.* What differentiates the ghostly internal image of the silence is not the silence itself, but the physically present vowel transition that leads up to it. That transition tells you that the silence is being produced in a particular location of the oral cavity system, which is automatically rendered as a representation of the sound as though it had been actively and independently produced. This example illustrates three related points about the relation of the most basic input level (the acoustic stream) and the perceived output level (the phone sequence).

a) It is non linear: information about what you perceive at the physical point P, can be based on the acoustic shape at some other point.

b) It is reconstructive: e.g., what you perceive to have occurred at point P can actually be missing entirely in the acoustic stream.

c) The reconstruction reshapes the acoustic phenomenology of what you think you "hear".

Such facts as these lead to an early version of reconstructive theory – the *motor theory of speech perception*. On this view, listeners use scattered acoustic cues as an input template, and then reconstruct the vocal motor gestures that would have produced those cues, filling in the missing information, and giving the acoustic sequence an intentional interpretation – *the speaker produced the particular initial acoustic features, by moving hiser vocal system in this particular way, uttering a particular sequence of phones to do so.* That is, the initial stage of decoding input acoustic features into phonetic segments involves a derivation, a model of what phones the speaker was expressing to create those acoustic features. We perceive what the speaker *intended*, reconstructing it by regenerating it from the few cues we initially detect.

The process of the derivational reconstruction of speakers' intentions exists at the interface of every level of linguistic representation. For example, the correct decoding of the phonemes underlying a phonetic sequence itself involves a series of computational operations. In this case the operations do not directly govern the articulatory gestures, but rather govern the organization of phonemic features. Consider the easy and correct recognition of the two phonetic sequences below as the phonemic sequences on the right (in conventional spelling for convenience).

Pa~Dr panter
Paa~Dr pander

(D = tongue flap which neutralizes the t/d distinction)

Several remarkable facts obtain about the phonetic instantiation of the phonemes. First, the /n/ has disappeared from both words, second the t/d distinction has disappeared. Yet, we 'hear' those features as though they were physically present. Somehow the perceptual system reconstitutes the underlying forms, even though the phones do not correspond in direct serial order to the phonemes. The decoding process has to somehow reflect a series of ordered operations that are part of English phonology, which define the *derivation* of the phonetic sequences:

a) **nasalize a vowel before a nasal**
b) **drop a nasal between a nasalized vowel and a following homorganic consonant**
c) **lengthen a vowel before a voiced stop consonant**
d) **change a /t/ or /d/ to /D/ following a stressed vowel and preceding an unstressed vowel**

Again, we see that the input/output relation is nonlinear, reconstructive, and downward flowing: in this case it is based on a set of derivational rules that are language specific. Again, the best solution is a model on which the listener recapitulates the sequence of computational rules to derive the surface form, given some input cues. This is the sort of model proposed by Halle and Stevens (1963) in their groundbreaking formulation of the analysis by synthesis architecture.

The broadest example of such analysis-by-synthesis is at the level of syntax and semantics. In the classic transformational derivational model the formal computation of a sentence 'starts' with a deep structure and semantic representation, which is then transformed by a set of language specific rules into a surface form. Thus similar surface forms can have different intentional deep structures:

They are eager to eat
They are easy to eat

Different surface forms can have the same intentional inner structure:

The dogs were chased by the cats
The cats chased the dogs
The cats happened to chase the dogs
It's the dogs the cats chased
….

And a single surface form can have different inner forms:

The lobsters were ready to eat.

The architecture of grammatical knowledge represents such facts in terms of computational derivations from inner to outer sentence forms, via a series of ordered transformational rules. (Note that the current architectural model of syntax, so called "minimalism" does not change this picture in the relevant respects). The problem for a model of sentence level comprehension is how to map such "vertical" derivations onto the manifest "seriality" of sentences. Recently, Dave Townsend and I corralled the current evidence for an analysis by synthesis answer to this problem (MIT, 2001). The essence of this model is similar to Ken's formulation for reading. The listener grasps cues at various levels of representations, and at each level reconstructs the speaker's intentions by trying out a potential derivation based on the most effective initial cues at the output level in each case – gestural, phonemic, syntactic, semantic… There is considerable evidence for this view and essentially no counter evidence, in the behavioral, acquisition and neurolinguistic literature. Thus, the model that Ken arrived at to explain many puzzling and creative facts about reading actually obtains for the fundamental processes of all language comprehension.

It is useful to emphasize a points about the input stage of this model, since that is where we can expect initial comparison to reading. Ken noted that the background assumptions of Reading presupposed that the input is a clear visual stream, letter by letter, assigned to words as the first stage of reading comprehension. There is a corresponding background discussion today for spoken comprehension: listeners assign syntax *first* and then derive semantics from that. This commonsense view rests on the empiricist logic, that without a syntactic organization of a sentence the meaning cannot be extracted. That is true, but there is no reason that the syntactic organization either has to be complete or even correct. Townsend and I adduced evidence that in fact the completely correct syntax is assigned LAST, not first. On our view, the process goes something like this:

a) **Apply statistically grounded patterns to the input (e.g., in English, the almost universal central pattern is: NpV(Np) = agent predicate patient).**
b) **Use those patterns to assign a likely initial meaning (meaning-1)**
c) **Use the patterns and the likely meaning to trigger a syntactic derivation**
d) **Check the derivation against the input. If it matches, assign the meaning associated with the full syntactic analysis (meaning-2).**

That is, as we put it, *we understand everything twice*. The reason we don't ordinarily notice this is that the processes, drawn out sequentially above, actually can operate in parallel (by projection ahead): in that view, meaning-2 ordinarily wipes out the consciousness of meaning-1.

Many experimental facts are consistent with a model which presupposes two phases of assigning structure and assigning meaning. Here is a simple one with implications for reading. It is well known that very brief interruptions of a sentence (aka "clicks") a#re misperceived as occurring at the boundaries of phrases. Thus a click objectively in the previous sentence at the point marked by a # will tend to be misperceived as occurring before the word /are/. Numerous studies have shown that this phenomenon is truly perceptual, not a response bias, not responsive to serial probabilities, and so on. It shows that an early stage of comprehension involves assigning a surface phrase structure, just as the classic syntax-first model would assume. But in fact, the misperception is limited to those phrases that are frequently ea*sily identified in the surface sequence, and that play a role in the initial pattern identification. Thus, the word /are/ in the above sentence is a characteristic initial word of a predicate, one of the small set of closed class function words, and the perceptual system can respond to that quickly. But in the sentence before the preceding one, the position marked by * is not near an easily identified phrase marker. If the complete phrase structure were available, a click in /ea*sily/ should be misperceived as occurring before the word rather than after it, because of the bracketing:

(are (frequently) ((easily identified)….

But this level of detailed bracketing has no effect on click mislocations. One interpretation is that it is simply too small a phrase to have a detectable effect. But there is a significant fact: if listeners are forced to wait a second after the sentence before indicating where the click occurred, *then* minor phrase boundaries have a significant effect. On our view this is because the initial segmentation is based only on superficial patterns: but the ultimate representation derives from a complete assignment of the syntactic structure, with all of the phrase structure details generated as part of the reconstructive derivational process. This relatively small point has interesting implications for ways to improve reading, outlined in the next section.

IMPROVING READING BY PROMPTING THE INITIAL ORGANIZATION

Ken's idea that reading is reconstructive has had enormous impact on research and educational programs, which everyone at this conference knows far better than I. But our detailed experimental analysis of how the reconstructive process works for spoken language has further specific implications for the improvement of reading by control of formatting. Writing systems in general (but not always historically) have specific ways of indicating segmentation in words, thereby solving a major problem of speech comprehension. Today, we take it as obvious that putting a space between words is a good idea. We also rely on punctuation conventions that can mark major phrases from each other. But what about smaller phrasing such as in the previous sentence, as broken up below:

**We also rely
on punctuation conventions
that can mark major phrases
from each other**

Numerous published studies, starting in the 1960s have shown that indicating phrase boundaries by some marker improves text comprehension. This fact remained a laboratory curiousity for many years without practical value, for three reasons: identifying "phrases" had to be done by actual people; implementing the boundary markers was limited to actual characters or extra whole space, which looked odd if not downright ugly; the notion of what counted as a relevant "phrase" was not well understood or uniform. Modern computer and printing techniques have offered solutions to each of these problems. Printers can be controlled to modify spaces and characters in very small increments that do not result in aesthetic disturbance; "phrases" can be automatically identified by many algorithms; the algorithms themselves provide precise definition of the phrases.

We have been testing the efficacy of a set of automatic programs we have written, called ReadSmart ™ (now patented), which incrementally increase space size between phrases. We have shown that comprehension of ReadSmart te3xts and reading speed improve by roughly 15% each, more for poor readers. We have also found that the texts are enjoyed more by readers and found to be more convincing. In one semester-long classroom study, readers using the phrase-spaced format earned significantly more honor grades, and had significantly fewer failures than readers using the normal format. This entire draft is formatted with such a program.

Why should phrase spacing improve reading? On the traditional view, it is because it reveals to the reader how to segment words together and build the correct surface phrase structure as an initial step in comprehension: this follows from the traditional view that the first step in comprehension is to determine the correct syntactic structure. But our phrase-formatting algorithms in fact do not find the syntactically correct phrase structure – rather, they isolate those kinds of phrases that are easily detected, based on distributional patterns of words and phrases in actual texts. For example, our algorithm phrases the two sentences below differently, as shown by extra spaces in them. Yet, from a linguistic standpoint, they have identical phrase structures as shown by the bracketed examples.

**The large dog was barking at the small cat
The large dog barked loudly at the small cat**

**(the (large dog)) ((was barking) (at (the (small cat))))
(the (large dog)) ((barked loudly) (at (the (small cat))))**

The different analyses assigned by our algorithms follow from the fact that function words such as /was/ and /at/ are easily learned as beginning phrases, while /barked/

is infrequent and will not be recognized by a model that learns phrase boundary cues from texts. This raises a question of theoretical interest: which kind of phrase boundaries are the best to use for implementing segmentation, syntactically correct ones, or those assigned by ReadSmart? With linguistic colleagues to help us assign a correct surface phrase structure to standard font-testing texts, we examined this question carefully. We contrasted the comprehension of phrase-spaced formats based on syntactic vs ReadSmart phrases. The results (published) astounded even us: the ReadSmart-phrased texts were far easier to comprehend, in fact the syntactic-phrased texts were *harder* to understand than normal untreated texts.

This follows from the reconstructive view of reading comprehension, as refined by our consideration of details of the analysis by synthesis model of spoken language. That model involves two phases of structure assignment, an initial one based on readily available cues and patterns, and a later one based on a full syntactic analysis. Our results show that basing visually salient phrase information on readily available cues leads to the best comprehension, thereby giving empirical support to our claims about initial phases of reading comprehension itself. It also gives support to the larger claim that like speech comprehension, reading involves several stages of extracting structure and assigning meaning.

CONCLUSION – THE VOICE WITHIN AND WITHOUT

These considerations offer some perspective on how readers rapidly create a linguistic representation along with the ghostly voice offering an internal rendering of the text. We have undermined the superficial view that the role of the voice is to implement the low level transfer of the print to audition. Rather, it reflects the output of the linguistic system, after reconstruction of the text. It is the same voice that we reconstruct during normal listening – the main difference is that we can hear it because there is no simultaneous acoustic input to overcome our awareness of it.

Ken's was an early voice without in the lingering behaviorist wilderness, proclaiming and demonstrating the computational complexity of even an apparently simple mapping task such as reading. His insightful idea that reading is reconstructive has been a major factor in reading education. This perspective is further supported by 40 years of concentrated research on the comprehension of spoken language.

Reference Note: This preliminary draft is largely free of specific references, which will be added for the final form of the paper. Readers interested in a full discussion might look at Townsend and Bever, 2001, *Sentence Comprehension* (MIT Press). Readers interested in learning more about our patented phrase spacing programs can go to our company website, www.lngtek.com. All readers are invited to contact me directly with comments, complaints, advice, etc. tgb@email.arizona.edu.