

Sentence processing: General background

Shravan Vasishth

April 16, 2015

This is a short (and incomplete) survey of very basic issues that come up in sentence comprehension research. The goal is to simply familiarize you with the general background for this course.

1 Introductory remarks

Online sentence comprehension clearly involves rapid and complex computational processes that go all the way from lexical access to syntactic/semantic processing, to discourse level processing. Sentence comprehension research attempts to

- identify the constraints on sentence comprehension
- define process models of how exactly sentence comprehension unfolds in the mind/brain

Historically (and simplifying only a little bit), sentence processing research has generally attacked this question by focusing on two major issues: complexity, and ambiguity. Both these issues are interwoven with research on **working memory** in **cognitive psychology**: in particular, constraints on forgetting, and individual differences in memory capacity. But this connection has not been explored as closely as it could have been.

Background reading: (Frazier, 1987).

2 Complexity

2.1 Grammar-based explanations

One of the first attempts at addressing the central puzzle of sentence comprehension was carried out by Miller and Chomsky (Chomsky & Miller, 1963). Their **derivational theory of complexity** asserts that the operations of grammar directly affect online processing difficulty. This idea quickly ran into trouble (Slobin, 1966), but has experienced a revival in recent years. A classic example is the passive versus active construction:

- (1) John bought a book.
- (2) A book was bought by John.

The passive, one could argue, is harder to process because of more complex syntactic operations involved in building a passive.

The idea that grammatical constraints might be responsible for processing difficulty has been pursued in detail by Colin Phillips' group in Maryland (see his home page for papers on this issue). In particular, the processing of island structures and syntactic constraints like c-command have been argued to be used by the parsing in online parsing.

Example: Island constraints.

Syntactic islands are structures that disallow extraction of wh-phrases. These include relative clauses, wh-clauses, factive clauses, subjects, adjuncts, and coordinate structures.

- (3) a. *What did the agency fire the official [that recommended —?
b. *Who do you wonder [whether the press secretary spoke with —?
c. *Why did they remember [that the corrupt CEO had been acquitted —?
d. *What did the fact [that Joan remembered — surprise her grandchildren?
e. *Who did Susan watch TV [while talking to — on the phone?
f. *What did [[the Senate approve —] and the House reject the bill?

It seems that the parser keeps track of whether an island configuration is being processed, which implies that the parser has access to this grammatical

constraint. Stowe's filled-gap effect and a follow up experiment illustrate this.

The filled-gap effect (Stowe, 1986): slower reading time at *us* in a vs b.

- (4) a. My brother wanted to know *who* Ruth will bring us home to at Christmas
- b. My brother wanted to know *if* Ruth will bring us home to Mom at Christmas

2.2 Working-memory based explanations

2.2.1 Similarity-based Interference

Miller and Chomsky were probably the first to raise the possibility that **structural similarity** (a working memory problem) could be a source of processing difficulty.

2.2.2 Decay

It has been widely believed (Just & Carpenter, 1980) that increasing distance between a dependent and a head makes processing more difficult (at the point where dependency is resolved), presumably because representations decay in memory. This idea, referred to in shorthand as locality, has received a lot of attention in the literature, see e.g., (Gibson, 2000), (Lewis & Vasishth, 2005).

It has been an open problem that, rather than decay over time, forgetting could be induced due to interference. Recently, psychologists (e.g., (Lewandowsky, Oberauer, & Brown, 2009), (Berman, Jonides, & Lewis, 2009)) have argued that decay in working memory could just be interference. Showing that locality effects are due to interference and not decay has been a big open problem in sentence comprehension research.

3 Ambiguity

Funny headlines like these often appear in newspapers:

- (5) Two sisters reunited after 18 years in checkout counter

A major reason for the funny “reading” of this sentence is a tendency on our part to incorrectly attach the prepositional phrase to mean *after spending 18 years in a checkout counter*.

But it’s not just newspaper headlines, such ambiguities occur in all kinds of day-to-day conversation; often we don’t even notice anything wrong. It’s as if our parsing system does not consciously consider the silly interpretations (but see (Tabor, Galantucci, & Richardson, 2004)). But the ambiguity is there, the human parsing mechanism is just really, really good at “pruning out” or ignoring inappropriate alternatives. Interestingly, sometimes the parsing mechanism chooses the wrong alternative.

- (6) a. I’m eating a pizza with a friend
- b. I’m eating a pizza with a fork and knife

Several factors can guide the parser’s decisions. One of them is prosody and punctuation (even in silent reading (Fodor, 2002), (Frazier, Carlson, & Clifton Jr., 2006)). In the example below, we tend to treat *the chair* the the direct object of *moved*, until we reach the word *broke*. At this point, a REANALYSIS is required. An intonational phrase boundary after *moved* (or a comma) would have saved us from making that initial mistake.

- (7) a. After the student moved the chair broke.
- b. After the student moved, the chair broke.

Prosodic breaks (or commas) can also determine whether relative-clause attachment occurs high or low:

- (8) a. The brother of the student who went to Japan on a holiday was left-handed.
- b. The brother of the student, who went to Japan on a holiday, was left-handed.

Here’s another example of grammatical role assignment going temporarily awry. Since *the lawyer* is more likely to be the one doing the explaining, people tend to get temporarily fooled (at *explained*) into treating *the lawyer* as the agent. Adding an explicit syntactic cue prevents this temporary mis-parse from occurring. For an interesting debate on this, see (Trueswell, Tanenhaus, & Garnsey, 1994), (Clifton et al., 2003).

- (9) a. The lawyer explained the problem got angry.

- b. The lawyer who was explained the problem got angry.

Pronoun disambiguation can depend on prosody (contrastive focus), among many other things:

- (10) a. John saw Bill and he ...
b. John saw Bill and *he* ...

4 Early attempts at explaining sources of ambiguity and complexity

4.1 Frazier's heuristics: Minimal attachment and late closure

Frazier proposed two **heuristic principles** that guide parsing decisions. These were mainly directed at explaining garden-path sentences.

Minimal attachment: Choose the structurally simplest analysis (the one with the fewest additional nodes)

- (11) The lawyer knew the answer was wrong.

Late closure: Integrate current input into current constituent (when possible).

- (12) After the student moved the chair broke.

The **limited capacity of working memory** has always been a central motivating factor in the development of theories of ambiguity and complexity in sentence processing. E.g., Frazier suggested that Minimal Attachment and Late Closure are reflexes of a **constrained capacity working memory** system.

Regarding Late Closure, she says (Frazier 1979, 39):

“It is a well-attested fact about human memory that the more structured the material to be remembered, the less burden the material will place on immediate memory. Hence, by allowing incoming material to be structured immediately, Late Closure has the effect of reducing the parser’s memory load.”

Similarly, regarding Minimal Attachment (Frazier 1979, 40):

“[T]he Minimal Attachment strategy not only guarantees minimal structure to be held in memory, but also minimizes rule accessing. Hence, [Minimal Attachment is also an ‘economical’ strategy] in the sense that [it reduces] the computation and memory load of the parser.”

Just as in the complexity debate, there has been an attempt to characterize **computation and memory load** in sentence processing using key findings from cognitive psychology. **Decay** and **interference** (cf. (Brown, 1958), (Peterson & Peterson, 1959), (Keppel & Underwood, 1962), (Waugh & Norman, 1965)) have played an important role in the models developed by Just and colleagues, and Lewis and colleagues.

4.2 Sausage machine

Frazier and Fodor’s “sausage machine” (Frazier & Fodor, 1978) was partly inspired by Miller’s famous paper (Miller, 1956).

Key idea: during real-time processing a phrase packaging device would package incoming items (each item could be a word or a syllable or morpheme) into units of 5 or 6 items; many consider this to be reminiscent of Miller’s (1956) seven plus or minus two items (cf. (Cowan, 2001)).

Example: Consider the two center-embedding structures shown below.

- (13) a. The rat the cat the dog bit chased ran away.
 b. The beautiful young woman | the man the girl loved | met on a
 cruise ship in Maine | died of cholera in 1962.

In (13a), the three NPs, *the rat*, *the cat*, and *the dog* are combined into a single phrase package, resulting in a conjoined reading that is incompatible with the following verb(s). By contrast, (13b) can be processed more easily due to ‘better’ packaging (indicated by vertical lines), which allows easier structure-building.

4.3 The garden-path model (Frazier et al 1982)

- This was based on the seven principles of parsing proposed by Kimball (1973) and the Sausage Machine model (Frazier and Fodor 1978).

- It uses MA and LC as heuristic principles to build a single initial parse (**strictly serial processing**).
- If the parse is incorrect, backtrack and re-do (this takes time).

(14) The horse raced past the barn fell

4.4 The serial-parallel parsing debate

Note that in principle there is no reason why parsing should be strictly serial. Consider the options:

- Serial: compute a single analysis, and if that fails, backtrack and compute new analysis.
- Parallel:
 - Ranked: Compute (=generate) all analyses in parallel, but rank them (e.g. by likelihood).
 - Prune: compute (=generate) parallel parses, but choose n-best using, e.g., beam search.
 - Don't prune at all—generate all possible structures and then compute some complexity function to quantify the uncertainty about the next word (Hale, 2001) or the rest of the sentence (Hale, 2006).

It is an open question whether humans use serial versus (some form of) parallel parsing (Gibson & Pearlmutter, 2000), (Lewis, 2000), (Hopf, Bader, Meng, & Bayer, 2003), although the evidence seems to be consistent with some limited serialism (although parallel structure building is difficult to rule out). The whole debate about serial vs parallel processing may be a non-problem; the two positions are hard to disentangle (Smolensky & Legendre, 2006a, 2006b).

5 Computational (implemented) models of sentence processing

In sentence processing there is currently nothing comparable to eye-movement control models like E-Z Reader (Reichle, Rayner, & Pollatsek, 2004) and

SWIFT (Engbert, Nuthmann, Richter, & Kliegl, 2005). Most of the work has been done in small-scale simulations, but this is now changing.

Some examples (not an exhaustive list!):

- Just and colleagues: READER/CAPS (Thibadeau, Just, & Carpenter, 1982), (Just & Varma, 2002), and 4CAPS (paper not yet released)
- Connectionist and constraint-based (MacDonald & Christiansen, 2002)
- Cognitive Architecture-based approaches: SOAR (Lewis, 1993), ACT-R (Lewis & Vasishth, 2005)
- Probabilistic/information-theory based parsing models (Jurafsky, 2002), (Hale, 2001), (Boston, Hale, Patil, Kliegl, & Vasishth, 2008).
- Models combining working memory and probabilistic/information-theoretic parsing constraints (Boston, Hale, Vasishth, & Kliegl, 2011).

Several (but not all) of these modeling approaches have attacked the problem of ambiguity resolution and/or complexity from somewhat different perspectives.

References

- Berman, M., Jonides, J., & Lewis, R. L. (2009). In search of decay in verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(2), 317.
- Boston, M. F., Hale, J. T., Patil, U., Kliegl, R., & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research*, 2(1), 1–12.
- Boston, M. F., Hale, J. T., Vasishth, S., & Kliegl, R. (2011). Parallel processing and sentence comprehension difficulty. *Language and Cognitive Processes*, 26(3), 301–349.
- Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, 10, 173–189.
- Chomsky, N., & Miller, G. (1963). Introduction to the formal analysis of natural languages. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology, volume ii* (pp. 269–321). John Wiley.

- Clifton, C., Juhasz, B., Ashby, J., Traxler, M. J., Mohamed, M. T., Williams, R. S., et al. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language*, 49, 317-334.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1), 87-114.
- Engbert, R., Nuthmann, A., Richter, E., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, 112, 777-813.
- Fodor, J. D. (2002). Psycholinguistics cannot escape prosody. In *Proceedings of the 1st International Conference on Speech Prosody* (pp. 83-88). Aix-en-Provence.
- Frazier, L. (1987). Sentence processing: a tutorial review. In M. Coltheart (Ed.), *Attention and performance xii* (pp. 559-586). Lawrence Erlbaum Associates.
- Frazier, L., Carlson, K., & Clifton Jr., C. (2006). Prosodic phrasing is central to language comprehension. *Trends Cogn Sci*.
- Frazier, L., & Fodor, J. D. (1978). The sausage machine: A two-stage parsing model. *Cognition*, 6, 291-325.
- Gibson, E. (2000). Dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain: Papers from the first mind articulation project symposium*. Cambridge, MA: MIT Press.
- Gibson, E., & Pearlmutter, N. J. (2000, March). Distinguishing serial and parallel parsing. *Journal of Psycholinguistic Research*, 29(2), 231-240.
- Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics*. Pittsburgh, PA.
- Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science*, 30(4).
- Hopf, J., Bader, M., Meng, M., & Bayer, J. (2003). Is human sentence parsing serial or parallel? Evidence from event-related brain potentials. *Cognitive Brain Research*, 15(2), 165-177.
- Jurafsky, D. (2002). Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic linguistics* (pp. 39-96). MIT Press.
- Just, M., & Carpenter, P. (1980). A theory of reading: From eye fixations

- to comprehension. *Psychological Review*, 87(4), 329–354.
- Just, M., & Varma, S. (2002). A hybrid architecture for working memory: Reply to MacDonald and Christiansen (2002). *Psychological Review*, 109(1), 55–65.
- Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior*, 1, 153–161.
- Lewandowsky, S., Oberauer, K., & Brown, G. (2009). No temporal decay in verbal short-term memory. *Trends in cognitive sciences*, 13(3), 120–126.
- Lewis, R. L. (1993). *An architecturally-based theory of human sentence comprehension*. Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA.
- Lewis, R. L. (2000). Falsifying serial and parallel parsing models: empirical conundrums and an overlooked paradigm. *Journal of Psycholinguistic Research*, 29(2), 241–248.
- Lewis, R. L., & Vasishth, S. (2005, May). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29, 1–45.
- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: A reply to Just and Carpenter and Waters and Caplan. *Psychological Review*, 109(1), 35–54. Retrieved from <http://psych.wisc.edu/ugstudies/Psych733/MacDonaldChristiansen.pdf>
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81–97.
- Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual items. *Journal of Experimental Psychology*, 61, 12–21.
- Reichle, E., Rayner, K., & Pollatsek, A. (2004). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, 26(04), 445–476.
- Slobin, D. I. (1966). Grammatical transformations and sentence comprehension in childhood and adulthood. *Journal of Verbal Learning and Verbal Behavior*, 5.
- Smolensky, P., & Legendre, G. (2006a). *The harmonic mind: From neural computation to optimality-theoretic grammar, volume 1*. MIT Press.
- Smolensky, P., & Legendre, G. (2006b). *The harmonic mind: Linguistic and*

- philosophical implications, volume 2*. MIT Press.
- Stowe, L. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language and Cognitive Processes*, 1(3), 227–245.
- Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, 50, 355–370.
- Thibadeau, R., Just, M., & Carpenter, P. (1982). A model of the time course and content of reading. *Cognitive Science*, 6(2), 157–203.
- Trueswell, J. C., Tanenhaus, M., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, 33, 285–318.
- Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review*, 72, 89–104.