

Acceptability is assigned independently from grammaticality; the four logical possibilities are illustrated by (4).⁵

(4) It is raining.

#Tom figured that that Susan wanted to take the cat out bothered Betsy out.

*They are running.

*#Tom and slept the dog.

Chomsky formulated this distinction in order to separate irrelevant processing constraints (e.g. limited time and space) from the grammaticality questions which he has been studying. Our hypothesis that a simple device can process language, is then, by definition, a hypothesis about the performance model. Acceptability judgments will bear crucially on the matter.⁶

The problem is to design a parser that *approximates* competence with realistic resources. Unacceptable sentences should be excluded because they require inordinate resources to process; ungrammatical sentences should be rejected because they violate competence idealizations (or approximations thereof). The design criteria are summarized below:

- (5) What are some reasonable performance approximations?
- (6) How can they be implemented without sacrificing linguistic generalizations?

1.2 The FS Hypothesis

We will assume a severe processing limitation on available short term memory (STM), as commonly suggested in the psycholinguistic literature ([Frazier79], [Frazier and Fodor78], [Cowper76], [Bresnan78], [Kimball73, 75], [Chomsky61]). Technically a machine with limited memory is a finite state machine (FSM) which has very nice computational properties when compared to an arbitrary PM. Most importantly, a FSM requires less time and space in the worst case. There are some other advantages which we have not explored

5. These examples are taken from [Kimball73]. A hash mark (#) is used to indicate unacceptability; an asterisk (*) is used in the traditional fashion to denote ungrammaticality.

6. Just as Chomsky idealized grammaticality from other unexplained irrelevant factors, it will be useful to idealize acceptability. In this work, we are most interested in time and space behavior *in the limit* as sentences grow; we will not address borderline cases where judgments tend to be extremely variable. This move is often taken in complexity arguments which study limiting growth, but ignore constants (borderline cases).

in detail. For example, it is easier to run a FSM in reverse. This may have some important implications if one were attempting to build a single model for both production and generation as suggested in [Kay75].⁷

When discussing certain performance issues (e.g. center-embedding),⁸ it will be most useful to view the processor as a FSM; on the other hand, competence phenomena (e.g. subjacency)⁹ suggest a more abstract point of view. Because of a lack of TM resources, the processor cannot literally apply rules of competence; rather, it resorts to more computationally realistic approximations. Whenever a competence idealization calls for inordinate resources, there will be a discrepancy between the competence idealization and its performance realization.

1.2.1 Center-embedding

Chomsky and Bar-Hillel independently showed that (arbitrarily deep) center-embedded structures require unbounded memory [Chomsky59a,b] [Bar-Hillel61] [Langendoen75]. As predicted, center-embedding is severely compromised in performance; it quickly becomes unacceptable, even at relatively shallow depths.

(7) #[The man [who the boy [who the students recognized] pointed out] is a friend of mine.]

(8) #[The rat [the cat [the dog chased] bit] ate the cheese.]

7. Trivially all physical machines are FSMs. The FS hypothesis is interesting, though, because the memory limitation is so severe (i.e. two or three clauses) that it is a crucial issue in many practical situations. Similar comments can be made about modern computers. Most engineers would model a typical large computer system as a TM. However, it would be hard to think of a computer as a TM if it had only 1 bit of memory. How much memory does it take before a FSM is best modeled as a TM? The answer may depend on the current price of memory. What once seemed unreasonable, may not be so unrealistic today.

8. A center-embedded sentence contains an embedded clause surrounded by lexical material from the higher clause: [_S x [_S ...] y], where both x and y contain lexical material.

9. *Subjacency* is a formal linguistic notion which constrains the applicability of a transformation. (Informally, subjacency is a locality principle: all transformations must be local to a single cyclic node (e.g. clause) or to two adjacent cyclic nodes.) We offer subjacency as an example of a competence idealization. In general, though, it is extremely difficult to prove that a particular phenomenon is necessarily a matter of competence. We have no proof that subjacency is a competence universal, and similarly, we have no proof that center-embedding is a processing universal. Our assessments are most plausible, though conceivably, they might be incorrect.