
LCTG 12—Convergence of Computation and Linguistic Theo

1

Implications for Linguistic Theory

**A methodological moral
or: Moral methodology**

Formalisation forces explicitness.

3

Chomsky and Shannon 50 years on

- The course and the subject started from Chomsky 1957 (Syntax as Kernel Sentences, Singulary and Double-based Transformations) and Shannon 1948 (Probabilistic Information Theoretic models of redundancy).
- 50 years on, most theories of grammar look pretty much like Syntactic Structures, apart from having gotten rid of Singulary Transformations.
- In computational linguistics, virtually all approaches now involve a probabilistic component.
- The most successful methods *integrate* these two elements, as in Collins' work and the Pereira reading.

2

The Tension between Descriptive Coverage and Explanation

Descriptive coverage

- Linguists want to be able to characterise the linguistic phenomena of any language . . .
 - What are these phenomena?
 - How do we know when we have the tools to model them?

4

Perspicuity

- Phenomena should be modeled in a “perspicuous” way

The reader can easily determine that to duplicate the effect of [the phrase-structure and transformational rules for generating the legal verbal sequences of English] without going beyond the bounds of a [Context Free Grammar] of phrase structure, it would be necessary to give a fairly complex statement. Once again, [...] we see that significant simplification of the grammar is possible if we are permitted to formulate rules of a more complex type than those that correspond to a [Context Free system]. *Chomsky: Syntactic Structures, p. 41*
- Warning: perspicuity is a subjective measure.

5

What is the object we are trying to model?

- What is syntax, and what phonology/morphology/semantics?
- What is competence, and what performance?

7

The Tension between Coverage and Explanation (Contd.)

Explanation

- We want to find a way of characterising the grammars of human languages that distinguishes them from other systems. Then the formalism itself would constitute an explanation for some features of human language.
 - *What* features of human language, exactly?
 - Most famous example: constraints on multiple dependencies.

6

Some Convergences in the Theory of Competence

- Lexicalisation of grammars
- The abandonment in Minimalism of any syntactic level of description
- The (re)recognition of the importance of characterising relevant domains of locality for particular types of relation

8

Some Convergences with the Theory of Performance

- Stochastic methods give a principled treatment of problems associated with parsing real-world language, especially when combined with linguistically adequate grammar formalisms, as Chomsky anticipated (1957, p.17).
- One of the most promising lines of investigation involves the use of grammar formalisms which are *lexicalized* and which capture *long-range dependencies*, like those discussed in this course.
- The problem can be stated as modelling, estimation and approximation.
- Performance is still poor (around 90% accuracy on most measures).
- The main problem is **sparsity of data**: 1 million words of human annotated data is not nearly enough to see all words with all categories.
- Exciting possibility: If the lexicon is the only language specific element of the grammar, is it possible to induce a lexicon automatically by looking at very large amounts of text such as the web?

9

Is This The Way People Parse Language?

- The existence of garden path sentences might suggest that it is:
 - (1) a. The horse raced past the barn fell.
b. The flowers sent for the patient died.
- However, such garden paths can be overridden or induced by *context*:
 - (2) a. Harry and his brothers were racing their horses all over the farm to see which was fastest.
b. The horse raced past the barn fell.
- The short term effect of context cannot be captured in global higher-order statistics over corpora. It requires the involvement of knowledge representations and inference.
- It remains possible, even likely, that such knowledge processes are probabilistic in nature.

10

Competence and Performance Revisited

- Theories of this kind are entirely modular, and the grammar can be identified independently of the statistical model that assigns probabilities to strings and the algorithm that searches for parses.
- Their principles generalize quite readily to more powerful grammars like CCG, TAG, HPSG, LFG, with one important qualification concerning the multiple dependencies captured in these grammars, which are not easy to deal with in generative models.
- However, in the end Competence and Performance are a package deal: Grammars must not only be linguistically adequate, but must also support parsing mechanisms that capture human performance.