Mathematising Linguistics
–
The Case of Transformational Theory

Thomas Meier

Munich Center for Mathematical Philosophy
LMU Munich

21 de Octubre, 2013

Aim of this talk:

- Outline the phase of theory change from Bloomfield to Harris
- Discuss Harris' and Chomsky's formalization of syntax and their introduction of transformational theory as a case of mathematization.

Why discuss mathematization?

→ Central part of the debate about maturity of scientific disciplines. Implies many connections to discussions on theory change and hence to the debate on scientific realism.

Content

Mathematization: Mathematical Explanation in Science

Recent publications on the topic of mathematical explanation, driven by the intuition that a mathematical explanation of an empirical phenomenon proceeds by identifying *structural* facts and *in virtue of* ignoring the (causal) details:

- Batterman, R. (2001). *Devil in the Details*, OUP, chap 4. "Asymptotics and the Role of Minimal Models", *BJPS*, 61:1-25
- Pincock, C. (2007). "A Role for Mathematics in the Physical Sciences", *Nous* 41(2):253-275.
- Bueno, O. and Colyvan, M. (2011). "An Inferentialist Conception of the Application of Mathematics". *Nous*, 45(2):345-374.

→ My work is part of the working group at the MCMP Munich: *Mathematical Explanation in Science*.

Mathematization II

- As such, the mathematical account of scientific explanation has the potential to undermine the pretensions of causal accounts, nowadays very popular. Obviously, to this end mathematical explanations should be such that they cannot be recast as causal, but rather be *distinctively* mathematical.
- While the existence of mathematical explanations may constitute a problem for the causalist, it doesn't directly affect the pluralist, who is happy with there being various kinds of explanation.
- Still, the interesting issue remains of understanding how exactly mathematical explanation works and what makes it different from other kinds of explanation.

Mathematization III

- Bird (2013):

  Kuhn describes an immature science, in what he sometimes calls its 'pre-paradigm' period, as lacking consensus. Competing schools of thought possess differing procedures, theories, even metaphysical presuppositions. Consequently there is little opportunity for collective progress. Even localized progress by a particular school is made difficult, since much intellectual energy is put into arguing over the fundamentals with other schools instead of developing a research tradition.

  see: Bird, Alexander, "Thomas Kuhn", The Stanford Encyclopedia of Philosophy (Fall 2013 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2013/entries/thomas-kuhn/>.

Mathematization IV

- What is a mathematical explanation of an empirical phenomenon?

  $\rightarrow$ A *systematized* and *formal* explanation.

- How does this debate shape our understanding of scientific explanations in general?

- The mathematical methods and levels of mathematization differ in the different scientific disciplines. In some, methods such as differential calculus are highly efficient (physics, biology, economy, etc.), in others, methods such as set-theory, mathematical logic or graph theory are more suitable.

2. Bloomfield's Structural Linguistics

Bloomfield was THE main figure in the first half of twentieth-century linguistics:

- *An Introduction to the Study of Language* (1914).
- *A Set of Postulates for the Science of Language* (1926).
- **Language (1933)** ← The paradigm of structural linguistics.
- *Menomini Morphophonemics* (1939).

The central claims of Bloomfield's theory

- Language manifests through **acts of speech**. Such acts are concrete actions of utterance. Languages are always spoken in **linguistic communities**.
- Linguistic communities are groups of speakers. Their utterances are alike or at least partially alike. This is a necessary feature of such communities.
- Recurrent sound features are **forms**. These forms receive meaning through recurrent stimulus-response-features.

The central claims of Bloomfield's theory II

- There is no semantics grounded on mental principles. The only way to study linguistic meaning is by behaviorist psychology (scientistic view - in 1933).
- Not further divisible forms are *morphemes* (e.g. *-hood, -ness, etc.*).
- There exist *free* and *bound* forms. A minimal free form is a word. Non-minimal free forms are phrases and sentences. A maximal form is a sentence.

The central claims of Bloomfield's theory III

- The minimal free forms (the words) of a language are finite in its number. Bloomfield refers to the totality of words that are part of a natural language.
- A *phoneme* is a minimal unit of sound. The morphemes of a language can be divided into smaller, meaningless phonemes.
- The phonemes of a language change through time. Changes in the phonemes of a language (phonological alternation) arise by the arrangement of certain phonemes, which influence on each other.

The central claims of Bloomfield's theory IV

- Bloomfield's theory focuses on the linguistic disciplines of phonology and morphology.
- There is no strong semantic theory. For Bloomfield, Behaviorism solves all semantic questions (very well fundamented in 1933!).
- The *typical* structural linguist starts with phonology, develops morphology and then a syntax out of it.
- The notion of linguistic form is central in Bloomfield's theory.

Bloomfield on Syntax

- *Word* :

  (1926: 156) *A minimum free form is a word.*

- *Phrase* :

  (1926: 156) *A non-minimum free form is a phrase* (e.g. the noun phrase *The dog*).

- *Sentence* :

  (1926: 158) *A maximum construction in any utterance is a sentence.*

Bloomfield preceding Harris

Bloomfield already mentioned what was systematized later by Harris: That certain morpheme-classes correctly combined build up a grammatical sentence.

- *Construction* :

(1926: 158) *Each position in a construction can be filled only by certain forms.*

Example: $\mathit{The}_1\, \mathit{tree}_2\, \mathit{is}_3\, \mathit{old}_4$.

In the empty slot, only certain classes of morphemes (or forms) can occur. In this case, certain noun-classes: $\mathit{The}_1\, [\emptyset]_2\, \mathit{is}_3\, \mathit{old}_4$.

- Bloomfield on *Substitution*:

  (1933: 247) *A substitute is a linguistic form or grammatical feature which, under certain conventional circumstances, replaces any one of a class of linguistic forms . . .*

  . . . thus, in English, the substitute 'I' replaces any singular-number substantive expression . . .

  [. . .] the substitute replaces only forms of a certain class, which we may call the domain of the substitute; thus, the domain of the substitute 'I' is the English form-class of substantive expressions.

  $\rightarrow$ We are able to define the set of English form-classes of substantive expressions: $D = \{he, the\ car, Peter, ...\}$.

2. Harris' Transformational Theory: Progress in Linguistics

- Harris developed his *Discourse Analysis* in 1952. His first systematic exposal was published as *Co-Occurrence and Transformation in Linguistic Structure* in 1957.
- Harris primarily searched for a criteria of differentiation between sentences.
- Before him, the analysis of language within the paradigm of structural linguistics had been carried out from the most elementary level of phonemes up to the level of sentences.

- Harris wanted to go further than Bloomfield and searched for a systematization of sentence-types.

- He developed a method that enabled linguists to describe how sentences are built and changed from one mood-type into another, like active - passive. These changes he called *transformations*.

- He introduced a formalism to label different types of phrases: noun phrases (N), verbal phrases (V), prepositional phrases (P), $\leftrightarrow$, etc.

Kernel Sentences

- 1957, 335: *The kernel is the set of elementary sentences and combiners, such that all sentences of the language are obtained from one or more kernel sentences (with combiners) by means of one or more transformations.*

Example of an elementary sentence: *John saw Jill.*

- 1957, 339: *Our picture of a language, then, includes a finite number of actual kernel sentences, all cast in a small number of sentence structures built out of a few morpheme classes by means of a few constructional rules; a set of combining and introducing elements; and a set of elementary transformations . . . .*

Sentences like:

John saw Jill, The cat sits on the mat, The tree is old, etc..

- Example: The kernel sentence:

  John saw Jill, in active is obtained from the passive

  Jill was seen by John, through a transformation.

- Harris' formalism in this case: $N_1\,V\,N_2 \leftrightarrow N_2\,V^*\,N_1$.

Co-Occurrence

- For Harris, morphemes can be grouped into classes (e.g. the class of suffix-morphemes *-hood*, etc.). The members of a class have similar sets of co-occurrents (here: *neighbour-*, *false-*, *likeli-*, etc. Each class occurrs with specific other classes to make a sentence structure.

- Furthermore, certain sequences of classes build up sentences. These sequences are products of a small number of elementary class sequences (constructions) which are combined in certain ways.

- *TNPNV* (= The risk of crisis vanished) results from the more elementary class sequences *TNV* and *NPN*.

Chomsky's Mathematization

- Chomsky took Harris' ideas and developed them further. This starts with his ground-breaking 1955 *The Logical Structure of Linguistic Theory* and his 1957 *Syntactic Structures*.

- As in Harris, the mathematization consists of the application of logical tools, i.e. first-order logic, especially predicate calculus, set-theoretic tools and, this more abundantly than in Harris, graph-theoretic methods (e.g. syntax-trees).

Chomsky's Mathematization II

- Harris introduced transformations as relations between sentences such as “I actually saw the car that you sold last year" and kernel sentences, as "I actually saw the car".
- Chomsky went further and developed a formal theory of grammar, where transformations manipulated not just the *surface strings*, but the *parse tree* associated to them, making transformational grammar a system of *tree automata*.
- It was a further development towards the same direction.

Chomsky's Mathematization III

- *Parsing* is the process of recognizing an utterance by decomposing it to a set of symbols and analyzing each one as a part of the grammar of a language.
- A first step towards a complete description of the meaning of an utterance is to break it down part by part and look at its analyzed form.
- The analyzed form that comes out, is what was called *deep structure* later on by Chomsky.

Immediate Constituent Analysis - Phrase Structure Trees

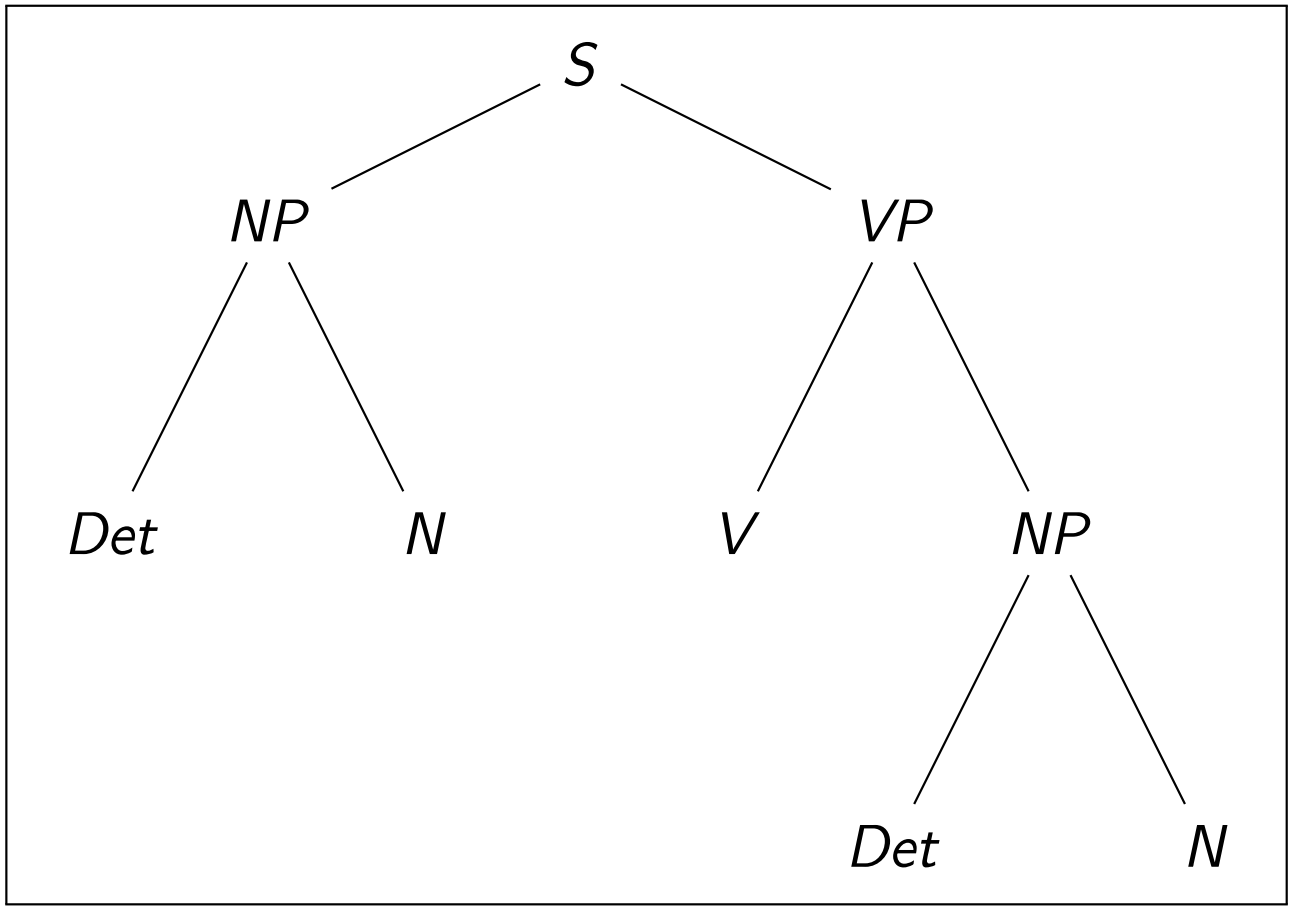

Figure: Analysis of the sentence: *The student reads a book*

Phrase-Structure-Rules - The Paradigm Case

Phrase structure rules determine how the constituents of a phrase are ordered. Consider the following example: The phrase structure rules for the english NP:

- Max (N)
- the crocodile (Det N)
- a big crocodile (Det Adj N)
- a crocodile in a lake (Det N PP)
- The general phrase structure rule for the english noun-phrase:

  NP $\rightarrow$ (Det) (Adj) N (PP)

A bit more on Chomsky

- The notions of *Deep Structure* and *Surface Structure* were introduced later (with Chomsky, 1965).
- Chomsky's proposals imply more controversial views on the relation between semantics and syntax.
- For Chomsky, every sentence had a deep structure and a surface structure. In the beginning, Chomsky even claimed that deep structure determined meaning.

- For the 1965 Chomsky, even interrogative sentences like

  Which car did John wash?

  had a deep structure, where their real meaning was

  John washed that car.

Results for the Philosophy of Science I

- Mathematization as a strong indicator of the maturity of a discipline.
- Bloomfield systematized an important part of syntactic theory.
- He already introduced that certain linguistic forms can be grouped into classes.
- Such form-classes cannot be substituted by any other form-class.

Results for the Philosophy of Science II

- Harris took Bloomfield's ideas on syntax and developed them.
- By introducing his formalism, Harris took a first step towards a mathematization.
- This made it possible to give linguistic theory a stronger explanatory power and certainly helped linguistics to become a *mature* science.

Results for the Philosophy of Science III

- Through the application of mathematical methods, whatever domain of knowledge can be precisely studied.
- It enables us to demonstrate objectively new results.
- One leaves the comfortable but nefarious castle of "unrevisability" and makes her work actually revisable, intersubjectively, by anyone who knows certain formal methods.
- Only through mathematization, a scientific theory is supposed to have *predictive power*. This can be measured, as well as the success of the explanations it provides.
- In this sense (very much in the spirit of Carnap), there is progress, not only in philosophy, but also in linguistics.

Thank you!

This work is funded by

Alexander von Humboldt
Stiftung/Foundation