『計量国語学』アーカイブ

ID	KK300707
種別	Guest Lecture
タイトル	ヨーロッパにおける計量言語学： 言語学に適用される科学哲学の諸原理について ［英語原文］
Title	Quantitative Linguistics in Europe: Principles of the Philosophy of Science as Applied to Linguistics [Original text in English]
著者	Reinhard KÖHLER
Author	Reinhard KÖHLER
掲載号	30巻7号
発行日	2016年12月20日
開始ページ	471
終了ページ	475
著作権者	計量国語学会
備考	KK300706の英語原文.

Mathematical Linguistics, Vol.30 No.7 (December 2016) pp.471-475.

Guest Lecture

Quantitative Linguistics in Europe: Principles of the Philosophy of Science as Applied to Linguistics

Reinhard Köhler (University of Trier, Germany)

Keywords: Quantitative Linguistics, Philosophy of Science, Europe, Synergetic Linguistics, self-organization in language

Abstract:
On the basis of the fundamental terms and concepts of the general philosophy of science, research strategies and methods of Quantitative Linguistics are described, as conducted in Europe since George Kingsley Zipf's seminal work. These general principles are in line with those practised in the natural sciences. Philosophical considerations as well as some decades of experience show that only specific methodological differences between studies on natural phenomena and on cultural ones such as language can be found. European approaches include empirical (inductive) and theoretical (deductive) strategies with a preference on the latter ones. The presentation starts with a description and explication of some terms and concepts which are fundamental for any kind of research, such as (scientific) problem, hypothesis, law, boundary condition, theory, and explanation – terms, which are often confused in linguistics. Examples are given of how these terms should be applied in linguistic work if the principles of linguistic research as practised in Europe are followed. In particular, the deductive strategy to find new laws is exemplified. The talk will also shed some light on the importance of laws not only for theoretical progress but also for practical purposes. Finally, a recent approaches is described which attempts to give a theoretical modelling background from which new law hypotheses can be derived and known ones can be systemised.

One of the fundamental requirements of scientific research is a clear and precise way of communication among scientists. The use of logics and mathematics serves this aim. An additional principle in science is the clarification (elucidation) and definition of each concept and of each term involved in scientific statements. The concepts and terms of science which have proved to be fundamental for each discipline have, therefore, been discussed in the Philosophy of Science. We will consider here the most basic terms used in any science.

Let us begin with the *scientific problem* because any scientific activity begins with a problem. Scientific problems differ from everyday problems by the following four properties: meaningfulness, well-formedness, well-formulatedness, and connections with previous scientific knowledge. Scientific problems can be differentiated into factual (empirical), conceptual, methodological, and valuational ones. A *scientific hypothesis* is an assumption about a possible way to solve a given problem, often in form of an assumed answer to the question in the problem. Hypotheses are the core of a science – not data as sometimes believed. A hypothesis in not a scientific hypothesis if it fails to be well-formed,

well-formulated and testable. There are several kinds of hypotheses, among them singular hypotheses, existential, statistical, bound and unbound universal ones. Only universal hypotheses which meet some additional conditions can become laws (see below). The way to formulate new hypotheses (and laws) thus starts from a problem and opens four kinds of epistemological approaches: analogy (find a similar problem which has already been solved and try to adopt the solution), inductive (starting from experience with appropriate data and formulate a generalization), intuitive (just have an idea and try), and deductive (start from a set of well-corroborated hypotheses (and laws if available), draw consequences from them which would answer the question at hand if correct).
According to M. Bunge, the famous Argentine philosopher of science, there are four levels of validation of hypotheses, which we can illustrate in the following way:

Experience/empirical findings	type of hypothesis	theoretical knowledge
none	intuitive, speculation, guessing	none
yes →	→ empirical generalization	none
none	deduced consequence ←	← yes
yes →	→ deduced consequence, empirically well corroborated, = plausible hypothesis ←	← yes

- Intuitive hypotheses without any empirical evidence and also without any theoretical justification may be nevertheless correct but there is only an infinitely small chance to find them true. - Empirical generalizations are grounded on some experience but they are still rather weak. For a very long time, science thought that all living things on earth need moderate temperature and oxygen to keep living. Only a few years ago, thermophile bacteria have been detected, which feel well in boiling water, and a kind of worms in the deep sea live on "black smokers", they get their energy from hot water above volcanic flaws. - Deductive hypotheses come with good reasons and are, therefore, more likely to turn out correct – but there is still a risk to find them wrong. - Only the fourth type, the universal, plausible, well-tested hypothesis has a perfect probability to be true. If such a hypothesis is anchored in a system of theoretical statements which are themselves at least of this type can it, when it has been tested again and again on more and more data from various sources, may be called a law. The structure of scientific research can be illustrated by means of a diagram after Bunge:

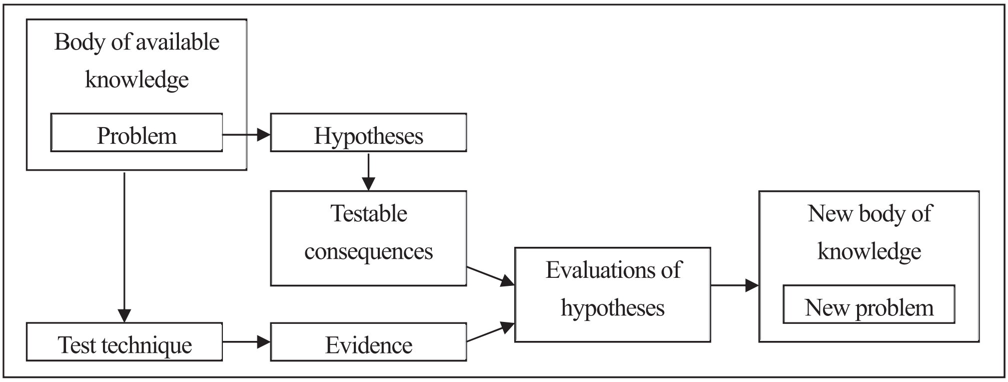

Figure taken from Bunge (1998a, p.11)

Science proceeds on three levels: observation (in linguistics the observation how people communicate with linguistic means), description (the level of dictionaries and grammars) and explanation (the level of laws and theories). We will therefore have a short look at the terms *law*, *theory*, and *explanation*.

As stated above, a *law* is a particularly well tested and corroborated universal hypothesis which follows from a theory or at least from a set of theoretically plausible statements. Laws are the only means to explain fact or phenomena or to predict events. Nothing can be explained without corresponding laws. They are expected to be valid everywhere in the universe and at all times.

A set of interconnected laws and accepted hypotheses is called a *theory*.

We must confess that in linguistics the presented terms are very often misused and cause some confusion. Individual claims and statements, methods, concepts, and descriptive approaches are called theories although all these elements do not contain any laws or hypotheses and fail to explain anything. All these elements will have to be explained themselves.

Now, what is a scientific explanation? It took many centuries until this concept was understood.

Not earlier than in 1965, Carl Hempel published a paper which showed that previous ideas about how logically correct and complete explanations can be achieved. The following scheme has become known as "Hempel-Oppenheim structure", showing the logical structure of the so-called deductive-nomological explanation. As this term indicates, explanation is understood as logical subsumption of a fact under laws by logical deduction.

$$\frac{\begin{array}{l} G_1, G_2, G_3, \ldots, G_n \\ R_1, R_2, R_3, \ldots, R_m \end{array}}{\text{Fact}}$$

The first two lines are called *explanans*, the last line is the *explanandum*. The $G_1, G_2, G_3, \ldots, G_n$ stand for the relevant laws, whereas the $R_1, R_2, R_3, \ldots, R_m$ represent boundary conditions. In linguistics, some laws are connected with boundary conditions such as that the object of analysis be an individual full text in a natural language – not a fragment and not a collection of texts. From such laws and boundary conditions facts can be deduced following the rules of logic reasoning.

Thus, the highest and most demanding goal of any science is the explanation of observed and described phenomena by means of a theory. In linguistics, two approaches to set up a theory are known. One of them is called *Synergetic Linguistics* (Köhler) and is based on a systems theoretic model, the other one, the *Unified Theory* (Wimmer and Altmann) consists of a complicated differential or difference equation. We will here describe Synergetic Linguistics because it is particularly well suitable to find and develop new laws by deduction from theoretical considerations, to test them, to combine them into a network of laws and hypotheses which are law candidates, and to explain the phenomena observed. As linguistic explanation is not likely to be possible by means of causal relations, synergetic linguistics aims at functional explanation (similar to biology). This type of explanation, however, is logically sound only under certain circumstances. A central axiom of synergetic linguistics is, therefore, that language is a self-organizing and self-regulating system (similar to an organism) – a special kind of dynamic system with particular properties.

Modelling in the framework of synergetic linguistics proceeds iteratively in refining phases, where each phase consists of six individual steps. In the first step, axioms are set up for the subsystem under consideration. There is one structural axiom which belongs to the synergetic approach itself: the axiom that language is a self-organising and self-regulating system. Other axioms take the form of system requirements. In synergetic terminology, these requirements are order parameters. They are not part of the system under consideration but are linked to it and have some influence on the behaviour of the system. In the terminology of the philosophy of science, they play the role of boundary conditions. These requirements can be subdivided into three kinds: (1) language-constituting requirements (among them the fundamental coding requirement, representing the necessity to provide expressions for given meanings, the application requirement, i.e. the need to use a given expression in order to express one of its meanings, the specification requirement, representing the need to form more specific expressions than the ones which are available at a given time, and the de-specification requirement for the cases where the available expressions are too specific for the current communicative purpose), (2) language-forming requirements (such as the economy requirement in its various manifestations) and (3) control-level requirements (the adaptation requirement, i.e. the need for a language to adapt itself to varying circumstances, and the opposite stability requirement). The second step is the determination of system levels, units, and variables which are of interest to the current investigation. In step three, relevant consequences, effects, and interrelations are determined. Here, the researcher sets up or systematises hypotheses about dependences of variables on other variables, e.g. with increasing polytextuality of a lexical item its polysemy increases monotonically, or, the higher the position of a syntactic construction (i.e. the more to the right hand side of its mother constituent) the less its information, etc. The fourth step consists of the search for functional equivalents and multi-functionalities. In language, there are not only 1 : 1 correspondences – many relations are of the 1 : n or m : n type. This fact plays an important role in the logics of functional explanation. Step five is the mathematical formulation of the hypotheses set up so far – a precondition for any rigorous test – and step 6 is the empirical test of these mathematically formulated hypotheses.

Laws are not only indispensible for the theoretic development of a scientific discipline but also of fundamental importance for practical problems. Even such seemingly simple tasks such as the comparison of two texts with respect to a specific property cannot be reliably solved without the background of at least one corresponding law. A researcher can define and use some measure and determine its value for the two (or more) texts. However, it is impossible to determine the significance of the differences between the measured values without knowledge of the lawful distribution of the given measure. The quest for laws serves both kinds of scientific activities: pure science and applied science.

References:

Bunge, Mario

1967 *Scientific Research I, II.* Berlin, Heidelberg, New York: Springer.

1998a *Philosophy of science. From problem to theory.* New Brunswick, London: Transaction Publishers. 3rd ed. 2005.

1998b *Philosophy of science. From explanation to justification.* New Brunswick, London: Transaction Publishers. 4th ed. 2007.

Hempel, Carl

1965 *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*. New York: Free Press.

Köhler, Reinhard

1986 *Zur linguistischen Synergetik. Struktur und Dynamik der Lexik*. Bochum: Brockmeyer.

1987 "Systems theoretical linguistics." In: *Theoretical Linguistics* 14/2-3; 241–257.

1990a "LinguistischeAnalyseebenen.Hierarchisierung und Erklärung im Modell der sprachlichen Selbstregulation." In: Hřebíček Ludék (ed.), *Glottometrika* 11. Bochum: Brockmeyer, 1–18.

2005b "Synergetic linguistics." In: Köhler, Reinhard; Altmann, Gabriel; Piotrowski, Rajmond G. (eds.), *Quantitative Linguistik. Ein internationales Handbuch. Quantitative Linguistics. An International Handbook*. Berlin, New York: de Gruyter, 760–774.

Wimmer, Gejza; Altmann, Gabriel

2005 "Unified derivation of some linguistic laws." In: Köhler, Reinhard; Altmann, Gabriel; Piotrowski, Rajmond G. (eds.), *Quantitative Linguistik. Ein internationales Handbuch. Quantitative Linguistics. An International Handbook*. Berlin, New York: de Gruyter, 791-807.

(Received on May 29, 2016)