Esa Itkonen

On explanation in linguistics

1. Introductory Remark

This paper purports to give an overview of the different ways that the term 'explanation' has been or is being used in the linguistic literature. It will be seen that in my opinion some of these uses are perfectly justified, others less so. As in all my previous publications, I have resisted here the utopian impulse which is all too common among the representatives of 'theoretical' linguistics: the present is thought to be full of promises that will be redeemed in the near future. As far as I can see, the reverse is true. If the present moment is experienced as less than satisfactory, it is so with respect to the past and not to the future. Is this view justified? Certainly no one who is ignorant of the history of linguistics has the competence to answer this question.

2. Rational (= Purposive) Explanation

Ever since Aristotle, the received view on human action (and action-explanation) is as follows. An action is performed in order to achieve a goal. The agent has a volitional or conative attitude towards a goal Y or, more simply, desires Y; and (s)he believes that an action X (which is at his/her disposal) is a means to achieve Y, i.e. that there exists a causal relation such that X is apt to bring about Y. (At this point it is irrelevant whether X is believed to be the only cause of this kind.) As a first approximation, it is this "volitional-epistemic complex" (G.H. von Wright) constituted by the goal-*cum*-belief, i.e. G & B, which brings about the action A; hence, (G & B) ⇨ A.

This needs to be spelled out more explicitly. First, we have to distinguish between the (goal-entertaining) conative attitude G and its object Y; hence, G:Y. Second, we need to distinguish between the belief B and its object X → Y; hence, B:(X → Y), where the simple arrow → expresses ordinary causation. Third, we need to express the idea that "who wants the end wants the means": if one genuinely wants to achieve Y by means of X, then the conative attitude G is necessarily transferred from Y to X, as expressed by the entailment sign ⊢. Fourth, we must express the transition from the mental to the spatio-temporal: it is by means of mental causation, expressed by the double arrow, that the observable spatio-temporal action X is brought about by the mental antecedents, or representations, discussed up to now. In sum, we get the following schema of **rational** explanation (RE):

RE: {[G:Y & B:(X → Y)] ⊢ G:X} ⇨X; and if all goes well, X → Y

RE, qua explication of the logical structure of actions, has the general structure α ⇨ β. To begin with, whatever stands to the left of ⇨ may be characterized as mental, whereas whatever stands to its right qualifies as spatio-temporal. RE contains three relations or transitions of different types: (i) ordinary causation → (first as mentally represented, and then in the space-time); (ii) entailment ⊢; (iii) mental causation.

RE says that if one wants to achieve Y and believes that X is conducive to bringing about Y, then one must want/intend to do X (or some action equivalent to it), and does X (unless prevented from doing so); and, in the successful scenario, X in fact brings about Y.

The notion of **reason** occurs in the following types of explanations: "Why did *a* kill *b*?" – "In order to have *b*'s money" or "because *a* wanted to have *b*'s money". These formulations are elliptical in the sense that they explicitly mention only the goal; but they implicitly also contain the information "and *a* believed that *a* would have *b*'s money by means of killing *b*". Thus, [....] is an explication of **reason**, just as {...} is an explication of **mental cause**. Reason-explanations are by definition means-end explanations.

RE represents the concept both of **action** and of **action-explanation**. This is possible because "intentional action is, on causal theory, defined by its causes" (Davidson 1973: 151). In the same vein, Hollis (1977: 131, 137) notes that a rational action is its own explanation. Woodfield (1976: 213) characterizes RE-like expressions as "hybrid" because X and Y stand both for mental entities and for spatiotemporal ones. But in reality the situation is even more complex. RE also involves some sort of **necessity**, as expressed by the entailment sign ⊢. The use of this sign may be illustrated by the following pair of sentences:

A) If John is looking at Mary, he is looking at something
B) If John is looking at something, he is looking at Mary

Here A expresses an entailment $p \vdash q$, whereas B does not. It is important to realize that the entailment sign stands for a (necessary) relation between concepts (or meanings), and not for a relation between mental events. If x fails to see the difference between A and B, or more specifically denies the (necessary) truth of A, this is a mental fact about x; but x is **wrong**, which is a conceptual (or normative) fact.

It is the purpose of the entailment sign in RE to express the idea that the **concepts** of goal and belief make it necessary that if someone desires Y and believes that Y will be brought about by X (which is at his/her disposal), then s/he **must** do X (or some equivalent action). But now the question arises: How is it possible for mental causation to exhibit conceptual necessity? As argued in Itkonen (1983: 49–53), the only coherent option is to assign to goals and beliefs an **ambiguous** status as units both of "world-3" and of "world-2" (to use Popperian terms). It is in their former capacity that they may have necessary (= 'conceptual') relations and be shared by several persons; and it is in their latter capacity that they may occur 'inside' individual persons and be involved in the processes of mental causation.

In sum, the three transitions contained in RE express three distinct **ontological** relations:

$\alpha \rightarrow \beta$ = between spatiotemporal entities
$\alpha \Rightarrow \beta$ = between mental vs. spatiotemporal entities
$\alpha \vdash \beta$ = between conceptual/normative entities

An action X **is** rational if (and only if) it happens to be an adequate means to bring about Y. It may only **seem** rational, namely when it is in fact an inadequate means to bring about Y. In the latter case the explanatory task consists in showing why an irrational action has come to

seem rational to the agent. The important thing is that all actions, both rational and irrational, are explained by RE:

> To explain an action as an action is to show that it is [or seems] rational. This involves showing that on the basis of the goals and beliefs of the person concerned the action was the means he believed to be the most likely to achieve the goal (Newton-Smith 1981: 24).

Man-made instruments are constituted by formal and 'functional' properties. The former are explained by RE in a straightforward way: If the agent has the goal of splitting wood, (s)he achieves it by **means** of making an instrument with requisite formal properties; and the same is true, mutatis mutandis, of each and every instrument.

It is a universal truth that the formation of instruments is governed by the principle of **economy** (or parsimony), which entails that this principle is crucially involved also in applying RE to **linguistic structure**. Another such principle, more specifically concerned with human systems of signification, or semiotic systems, is that of **iconicity** (cf. Haiman 1985, Itkonen 2004), which in turn exemplifies the more general notion of **analogy** (cf. Itkonen 2005a: 3.2).

Some wider implications need to be discussed, however briefly. As mentioned above, rationality has a 'Janus-like' nature, being both normative and psychological (as expressed by the two transitions $\alpha \vdash \beta$ and $\alpha \Rightarrow \beta$) (cf. Itkonen 1983: 177–181). As normative phenomena, rationality principles are not accessible to sense-perception (= observation) but to intuition (just as rules of correctness are). This is the source of (formal) models for rational behavior, or what Diesing (1972) calls 'synthetic' models. They may be reinterpreted as **causal** models simply by assuming that people have indeed internalized those rationality principles that have been chosen as the object of formalization. The notion of synthetic model is both explored and exemplified at some length in Itkonen (1983: 286–313).

The same (or at least a similar) idea has been expressed by Popper (1957) as follows:

> For in most social sciences, if not in all, there is an element of **rationality**. Admittedly, human being hardly ever act quite rationally..., but they act, none the less, more or less rationally and this makes it possible to construct comparatively simple models of their actions and interactions, and to use these models as approximations. The last point seems to me to indicate perhaps the **most important difference in their methods** [i.e. the methods of the natural and the social sciences] ... I refer to the possibility of adopting, in the social sciences, what may be called the method of logical or rational construction, or perhaps the 'zero method' (pp. 140–141; original emphasis).

> The 'zero method' of constructing rational models is not a psychological but rather a logical method (p. 158; discussed in Itkonen 2003b: Chap. 23).

In some of my publications, RE has been applied e.g. to the following cross-linguistic phenomena: marking of intransitive vs. transitive subjects, marking of the SG vs. PL distinction, grammaticalization, N-and-N constructions, converbs, zero in verb morphology.

To conclude this section, a set of objections against RE need to be addressed:

(i) The standard objection is to say that "not all actions are rational". This has already been shown to be a misunderstanding.

(ii) RE has been criticized on the alleged grounds that rationality needs conscious deliberation. This is just another misunderstanding: "There is no need for an agent to be aware

of the operation of his desire and belief when he acts purposively. He need not even know that he has them" (Woodfield 1976: 171). It follows that there is no obstacle, in principle, against applying RE to the behavior of (higher) animals.

(iii) RE qua causal explanation has been criticized on the alleged grounds that causality demands nomicity (which is lacking in RE). This turns out to be another prejudice without foundation: "I shall dogmatically assert the need for an account of **agent causality**..., according to which causality does not presuppose 'laws' of invariant connection (if anything, the reverse is the case) ..." (Giddens 1976: 84; original emphasis).

(iv) By definition, rationality entails the possibility of choice, both between different goals and between different means to achieve one and the same goal. Our notion of RE seems **elliptical**, insofar as it does not explicitly account for these different alternatives. It also seems **informal**, considering that very sophisticated models of rational choice have been developed over the decades in decision theory and in game theory. For instance, if the agent X has to choose between two alternative courses of action A and B, he should – ideally – base his choice on their respective 'expected utilities' (= EU). A (just like B) may result either in success (= A1) or in failure (= A2). Now, the EU of A can be computed by adding two products: 'the probability of A1 × the gain connected with A1' and 'the probability of A2 × the loss connected with A2'. X should choose A or B, depending on which one has the greater EU (cf. Benn & Mortimore 1976). – This criticism may be answered as follows. Because REs are generally post hoc, those who formulate them concentrate on finding out the most plausible or **coherent** (cf. Sect. 8) account of those goals and beliefs that the agents entertained in fact. Before zeroing in on the most likely candidates, they have discarded several alternatives. The fact that these may not be explicitly mentioned does not mean that they have not been taken into consideration. As for models based on expected utilities, they are just impracticable in most cases: "decision theorists concentrate on what they call risks, that is, numerically calculable probabilities. But such calculable risks are rarely found in real social situations" (Gibson 1976).

(v) RE qua causal explanation has been criticized because the notion of **mental** causation remains unclear. This is true, but there is no alternative; cf. the next point.

(vi) There are those who think that RE is **old-fashioned**. Instead, they recommend that such (socially conditioned) mental entities as desires and beliefs be straightforwardly **reduced** to 'brain-states', and ultimately to physics. Hilary Putnam once subscribed to this view, but has since then thought better of it:

> [We have] no idea of the nature of the theory in terms of which we are supposed to do the reducing (and only a very problematic idea of what theory we are supposed to reduce) (1999: 35).

> Saying 'Science may one day find a way to reduce consciousness (or reference, or whatever) to physics' is **here and now**, saying that science may someday do we-know-not-what we-know-not-how. And from the fact that these words may in the **future** come to have a sense we will understand it no more follows that they **now** express anything we understand than it follows from the fact that I may someday learn to play the violin that I can now play the violin (p. 173; original emphasis).

3. Functional Explanation

It is a fact that not just laypersons but also professional biologists standardly explain the existence of organs by means of the following kind of **functional** explanation (FE):

> FE: As part of Z, X has, and is explained by, the function Y if, and only if, X causes Y which is necessary for Z's survival (i.e. Y = effect/function)

For instance: In vertebrates, the heart (= X) causes the blood to circulate (= Y) through the organism (= Z), thus keeping keeping it alive; thus, Y explains why X is in Z.

The relation between RE and FE is not straightforward. On the one hand, RE and FE are similar insofar as they both exemplify the notion of **teleology**: "The consequences [= effects] of goal-directed [= teleological] behavior are involved in its own etiology: such behavior occurs **because** it has certain consequences" (Wright 1976: 20; original emphasis). Thus, both in RE and in FE the explanation of X involves some sort of reference to Y.

On the other hand, there are the following crucial differences between RE and FE: (i) The framework of **mental representations** (and the consequent ontological ambiguity), characteristic of RE, does not apply to FE. (ii) As a corollary, Y in RE need not occur in space-time at all whereas Y in FE is actually there. (iii) Causation is **non-nomic** in RE and nomic in FE.

It is a curious fact that, in spite of these clear-cut differences, "people still confuse functional explanations with purposive [= rational] explanations, just as Aristotle did" (Woodfield 1976: 212).

It needs to be added that nomicity is a matter of degree:

> Conscious actions qualify as teleological, representational, and non-nomic. Their non-nomicity means that although such internal causes as reasons may come up in response to various external causes, this does not happen in accordance with a regularity, but rather as a result of spontaneous causation. The more one descends the continuum of human behavior towards its 'lower' end occupied by fully unconscious and automatic subactions, the more the degree of nomicity increases (Itkonen 1983: 54).

FE, unlike RE, can arguably be **reduced** to efficient or mechanistic (= non-teleological) causation: the internal mechanism of Z is such as to cause X which in turn causes Y: hence, Z → X → Y. For instance, the functioning of the heart can be reinterpreted in this way, namely as a matter of efficient physiological causation. The same type of interpretation is, or seems to be, illustrated in concrete detail by the description of **homeostatic machines** (like thermostats). The following objections may, however, be raised against this attempt at reduction (cf. Itkonen 1983: 33–44):

(i) Even when the reduction can be carried out, teleolological explanations may remain convenient.
(ii) There are many contexts where the reduction cannot carried out, at least not yet.
(iii) It can be claimed that mechanistic explanations fail to capture the meaning of teleological ones.
(iv) Mechanistic descriptions of homeostatic machines are **defective** insofar as they leave out the **purpose** for which they have been designed, in the first place; i.e. such a machine is explained, ultimately, by applying (not FE but) RE to the person who has designed it.

(v) Even when considered in themselves, homeostatic machines exemplify representational causation (and thus remain in the vicinity of RE) insofar as "the 'desired' end-state is encoded in their internal structure" (Woodfield 1976: 193). On the other hand, the human body may be considered as a homeostatic system (rather than, literally, machine).

The following terminological inconsistency is to be noted. As I have shown in several publications, the huge majority of explanations proposed by representatives of the typological-'functional' school exemplify the notion of RE, and not of FE.

Is there, then, no use for FE in linguistics? The fate of FE in social sciences serves as a cautionary example. In the heyday of functionalism, as represented by Malinowski, Radcliff-Brown, Parsons, and Merton between 1930 and 1955, it was thought that a social institution can, and should, be explained by its "latent functions", i.e. the unintended consequences of the corresponding institutional behavior, insofar as these were taken to be necessary or at least beneficial for the "survival" of the society at large. Taken at the face value, this is of course FE pure and simple. But the validity of such explanations may well be doubted. In biology there is a general consensus concerning what is functional or dysfunctional for the organism. In anthropology and sociology, by contrast, there is no similar consensus, with the result that each scholar seems to apply his/her own version of FE. Hempel (1965c [1959]: 319–325) and Nagel (1961: 520–534), for instance, are quite critical in their assessment of functionalism, and Giddens (1976) comes to the same conclusion: Its many defects "undermine any attempt to remedy and rescue functionalism" (p. 20), "with its emphasis upon social 'adaptation' to an 'environment'" (p. 111).

Accordingly, the prospects for FE in linguistics may look grim. This view does not, however, agree with the fact that in the linguistic literature there are frequent suggestions to the effect that languages constitute "functioning wholes", with antithecal tendencies cancelling out one another. This is how Whitney (1979 [1875]) described the resulting state of equilibrium (capturing, in the process, the essence of **erosion**):

> The tendency to abbreviation for ease, for economy of effort in expression, is a universal and blind one; destruction lies everywhere in its path. ... But we may note for our consolation that [a speech community] does not lose what it once possessed in the way of inflectional apparatus without providing some other and on the whole equivalent means of expression (pp. 106–107).

Hermann Paul likewise envisaged each synchronic state of language as a compromise between sound change and analogy: "So sehen wir in der Sprachgeschichte ein ewiges Hin- und Herwogen zweier entgegengesetzter Strömungen" (1975 [1880]: 198).

The same overall view was propounded by so-called Natural Morphology, developed in the 1980's by Dressler, Meyerthaler, and Wurzel. To be sure, conflicts were now seen to exist not only between phonetics/phonology and morphology, but also within morphology: on several distinct dimensions there is a striving after 'naturalness' (or simplicity), but since these different tendencies cancel out one another, the end result is not, in general, an overall simplification, but rather a state of equilibrium.

So what should we conclude about the validity of FE within linguistics? Perhaps Giddens (1976) provides a clue. He makes (p. 121) a distinction between homeostatic systems and "equilibrium systems". The former are guided by "control centres by means of which input

and output are mutually assessed and coordinated". The latter, by contrast, react to local disturbances "blindly", or on an *ad hoc* basis. Giddens sees no objection against conceptualizing a given society as this type of equilibrium system; and it is quite plausible to conceptualize any given language in the same way. I submit that this is exactly what was intended by Whitney and the others mentioned above.

Thus we reach the conclusion that FU may be applied to language in some vague sense. But this very vagueness constitutes a problem. There seems to be no systematic or theoretical foundation for FU. What we have, instead, is just a set of disparate observations that either do or do not support the idea of a 'balance' between various linguistic subdomains. For instance, it might be thought that if a language has few vowels, it must have many consonants, and vice versa. But Maddiesen (2005a: 15) quashes this idea: cross-linguistically "absolutely no correlation was found between the number of vowels and the number of consonants". On the other hand, "complex tone systems are strongly correlated with the occurrence of moderate rather than complex syllable structure" (2005b: 59), – but **not** with simple syllable structure. It remains to be seen whether any coherent equilibrium theory, formulated in FU terms, will ever emerge out of such disparate observations. Certainly this possibility cannot be not precluded.

4. Evolutionary Explanation

Explanations given within the Darwinist framework may be formulated in more than one way. The following formulation of evolutionary explanation (EE) has been chosen for a number of reasons:

> EE: The feature X (ultimately produced by mutation/recombination) is functional (= adaptive) for the organism Z if, and only if, X enhances Z's chances of survival, i.e. of not being eliminated by Y (= natural selection)

First, the explanatory role of Y (combined with the terms 'functional' and 'survival') is meant to suggest a *prima facie* similarity with FE.

Second, the label 'Y' itself may also suggest a *prima facie* similarity with RE, in particular with the designing of homeostatic machines, but "with natural selection replacing [conscious] design" (Wright 1976: 69).

Third, and in opposition to the two preceding points, EE incorporates the thesis of Mayr (2001: Chap. 6): "Natural selection is elimination", i.e. instead of the good individuals being selected, the bad ones are eliminated; hence, no teleology. The same point is asserted even more strongly by Gould (1989: 228): "Natural selection is the cause of evolutionary change" (p. 228); but, contrary to the traditional wisdom, it need not be the case that natural selection favors those with some "mechanical superiority in anatomical design" (p. 288). Rather, especially in the light of the data from the Burgess Shale, natural selection seems to be "decimation [= elimination] by lottery" (pp. 244, 261, 262). Hence, chance, or utter lack of teleology, turns out to govern not just mutation and recombination, as traditionally assumed, but also natural selection to some extent.

The non-teleological character of EE is emphasized here in order to differentiate it from RE and FE. Otherwise it seems redundant. Surely no one contests the non-teleological nature of evolution anymore? Curiously enough, such a dissenting view has been voiced recently: "Biology has been an unabashedly functional-adaptive discipline ever since Aristotle, ... Put another way, biological design is driven by a **teleology**" (Givón 2009: 20; original emphasis).

T. Givón's merits in the field of typological linguistics are beyond question (cf. Itkonen 2008a). He is so deeply (and rightly!) convinced of the teleological nature of linguistic change, and of linguistic behavior in general, that he commits the mistake of seeing evolutionary change in the same light. At least this is how I interpret his claim.

The nature of EE may be further clarified by the following passage from Cohen (1986), which I have often quoted in my previous publications: "Hence no evolutionary change of any kind came about through the application of intelligence and knowledge to the solution of a problem. This was at the heart of Darwin's idea" (p. 203). On the other hand, linguistic behavior in its totality must be conceptualized as problem-solving: In the RE terms, achieving Y is the problem, and finding the (even approximately) right X is the solution. It follows that EE does not apply to linguistics: I just refuse to jump onto the now-fashionable "Darwinist bandwagon". Notice, however, that Darwin should not be blamed for the excesses of his over-zealous acolytes.

The fascination with Darwinism rests on the broad analogy between evolutionary change and linguistic change. But, as suggested e.g. by Mayr (2001: Chap. 4), this analogy can be extended to **geological** change as well (and perhaps even farther). It is certainly no accident that the term **erosion** plays a central role in the typological framework developed e.g. by Haiman (1985) and Heine & Kuteva (2007). Haiman (Chap. 3) even goes so far as to identify erosion as the consequence of **economy**, the principal causal force behind language change. (To continue in the same vein, we may also note the structural similarity between geological river deltas and biological/linguistic family trees.) But extending the analogy in this way is a sure way to water down whatever explanatory force EE may have been thought to have within linguistics.

At a more specific level, we should note the following clear-cut difference in value-judgments concerning the **origin** of biological vs. linguistic entities. In evolutionary theory it is customary to distinguish between two types of similarity (cf. Gould 1989: 213). **Homology** means similarity "due to simple inheritance of features present in common ancestors". By contrast, **analogy** means similarity "arising by separate evolution for the same function", as exemplified by "the wings of birds, bats, and pterosaurs". Now, there is a stark conflict between these two notions. Discovering homologies is the only true goal of evolutionary theory. By contrast, analogies are nothing but "pitfalls and dangers" and constitute "the most treacherous obstacle to the search for genealogy". Because anatomical structures may lose their original functions, there is no obvious temptation to connect homology with FE. By contrast, it may seem natural to apply FE to explain analogy (as in the wings of birds, bats, and pterosaurs), and an extra effort may be needed to show that what may seem as suitable material for (teleological) FE must be so reinterpreted as to become suitable material for (non-teleological) EE.

The same distinction is of course well-known in (typological) linguistics as well, under the labels of 'genetic' vs. 'typological' similarity; but the value-judgment is quite different.

The data of typological linguistics is in its entirety based on the analogy (*sic*) between the world's languages. Itkonen (2005a: 6–7) shows how the sentence meaning 'I do not see it' is expressed in the following genetically unrelated languages: German, French, Finnish, Swahili, West Greenlandic, Wari' (and Hua has later been added to this list). In spite of considerable typological variation, all six sentences are structurally similar (= analogous) insofar as each of them must have distinct forms for expressing the same four meanings (or 'functions'), i.e. one lexical meaning (= 'see') and three grammatical meanings (= NEG, AG.1SG, PAT.3SG.N). This is literally unity in diversity.

Now, the important thing is that typological linguistics (and linguistics *tout court*) would cease to exist, if analogy (as opposed to homology) were considered a mere nuisance. The difference between between linguistics and evolutionary theory is absolute, just as it is between RE and EE.

What should we say about the current attempts to replace RE by EE? It is just vacuous to claim, as is done to an increasing extent, that each and every linguistic change is "a response to adaptive pressures". Such formulations annihilate today's explanations without giving anything in return, apart from some (biological) metaphors.

In concluding this section, it is good to add that I am dealing with linguistics, as this term has traditionally been understood. This means that my time-scale is maximally 10'000 years; it does not go beyond the time of (reconstructed) Uralic or Indo-European proto-languages. Above, I have considered the validity, or otherwise, of EE within these limits. If the time-scale is changed so as to encompass e.g. the last 1'000'000 years, I have nothing to say.

In all my publications I have emphasized the importance of **normativity**, either in the sense of language-specific correctness or in the sense of general (means–end) rationality (cf. e.g. Itkonen 2008b). It goes without saying that 1'000'000 years ago there was no normativity and that since then it must have emerged in one way or another. What matters for me is that it has been there during the last 10'000 years, or the time-scale which is my exclusive concern:

> An *ought* cannot be derived from an *is*; normativity cannot be derived from descriptiveness. Yet the descriptive fact that we do have biologically instilled normativity boxes and operators (as, I conjecture, is the case) can be given a thoroughly naturalistic and non-normative evolutionary explanation (Nozick 2002: 271–272).

I fully agree. But, as admitted by Nozick, normativity **is** with us today, and this fact suffices to expose the weak point in all-out physicalism, i.e. the view according to which all entities of the universe are of physical nature. As argued by Putnam (1999), there is no reason to grant this assumption; but let us do so anyway. Now, in order to show that everything can be reduced to physics (including the thoughts of those who are engaged in the very act of reducing), it must be possible to describe everything in (what ultimately reduces to) the **language** of physics. But this language (just like any other language) is of normative nature, as shown by the fact that those who use it can behave either correctly or incorrectly, which is something that physical entities can**not** do. Therefore, even granting that everything is physical, any attempt to state this fact scientifically would *eo ipso* amount to self-contradiction. Even if you are right, you can never assert what you want to assert; but if you are wrong, I have at my disposal the full power of the philosophical and scientific language to assert whatever I wish.

5. Deterministic Explanation

Genuine determinism equals the idea that there are (physical) laws valid always and everywhere. From the late 1940's until the early 1970's the generally accepted explication of deterministic explanation (= DE) was provided by the so-called deductive-nomological (= D-N) model, also known as the covering-law model (cf. Hempel 1965e). In this model, general (= deterministic) laws are formulated as universal implications, and single (observable) events are explained by deducing them from the premises which contain at least one general law, in addition to some (observable) events, i.e. antecedent conditions (= AC), that obtain just prior to or simultaneously with the events to be explained. (More precisely, we should speak about sentences referring to laws and events, rather than about laws and events *tout court*.) The D-N explanation may be illustrated by the following figure:

	Law	∀x(Fx → Gx)
Explanans		
	AC	Fa
		_____
Explanandum		Ga

AC and Explanandum can often – but not always! – be identified with 'cause' and 'effect', respectively. In its standard form, the D-N model is elliptical insofar as the deductive inference from the Explanans to the Explanandum requires an intermediate stage not explicitly mentioned in the model: first, *Fa → Ga* is derived from *∀x(Fx → Gx)* by Universal Instantiation, and then Modus Ponens is applied to *Fa → Ga* and *Fa* to yield *Ga*.

As it stands, the D-N model seems to be confined to very simple instances of explanation, since both *F* and *G* stand for observational concepts. The model may, however, be further developed to cover increasingly complex instances of **theoretical** explanation as well. How this happens, will only be outlined here (but cf. Stegmüller 1974: 168–174). Let us assume the law *∀x[(Fx & Cx) → Gx]*, which combines the theoretical concept *F* and the observational concept *C* in such a way that in an observable situation, e.g. *Ca*, the unobservable cause *Fa* produces the observable effect *Ga*. Then we have the following D-N explanation:

∀x[(Fx & Cx) → Gx]
Fa & Ca

_____
Ga

The truth of *Ca* is observed but the truth of *Fa* can only be assumed. For this assumption to be plausible, we need an additional law such as *∀x(Fx → Hx)*, which connects the theoretical concept *F* with the observational concept *H*. Then if *Ha* is observed to be true, we may inductively (= tentatively) infer *Fa* to be true, which allows us to assume *Fa* as a premise.

Ever since the 1970's, the D-N model has been subjected to extensive criticism, summarized in Kitcher (1998) and Salmon (1998). As noted above, the idea of a **law** is crucial

here; but there is no guarantee that this is indeed what is expressed by the universal implication. In other words, "when [the D-N model] is viewed as providing a set of necessary and **sufficient** conditions for explanation, it is far too liberal. Many derivations which are intuitively nonexplanatory meet the conditions of the model" (Kitcher 1998: 279). It follows that the model should be constrained, especially by adding the **causal** point of view. Still, it must have a rational kernel because it has been "something of a philosophical commonplace ever since the days of Mill and Jevons" (von Wright 1971: 175, note 35). Philosophical ideas with such a pedigree cannot be entirely wrong.

Originally the D-N model was also meant to show how less general laws are explained by deducing them from more general ones. That this was never done in fact, constitutes a further weakness of the model (cf. Salmon 1998: 248). It will be suggested in Section 8 that laws might preferably be explained by means of analogy than by means of deduction.

In any case, we need the notion of DE. Therefore, whatever the defects of the D-N model, it will serve here as an explication of DE.

Next, we shall consider the viability of DE in linguistics. In the early stages of generativism Chomsky (1957) simultaneously endorsed the natural-science view of linguistics and rejected any use of statistical/probabilistic methods, thus committing himself to all-out determinism. As he saw it, just as a theory of physics contains "general laws", so "a grammar of English ... will contain certain grammatical rules (laws) ..." (p. 49), such as those given on p. 26: *Sentence → NP + VP, NP → T + N, VP → Verb + NP, T → the, N → man, ball,* etc., *Verb → hit, took,* etc.

Needless to say, the analogy between physics and English grammar was ludicrous. The discrete (= two-valued) nature of norms was mistaken for the deterministic nature of physical behavior. Such a mistake results from "a catastrophic failure to distinguish nomological investigations from normative ones", to use an apt formulation by Baker & Hacker (1984: 285).

It was one of the tasks of my 1974 dissertation to demonstrate in detail the fundamental difference between synchronic grammatical description and (e.g.) Newtonian mechanics. My views met with considerable opposition (see e.g. Dahl 1980 [1975]), which – I am happy to report – is no longer the case.

Today generativism still subscribes to determinism, but in a new and seemingly more acceptable form. This needs to be spelled out more explicitly.

Between 1965 and 2005 the main focus of generativism was on the innate Universal Grammar (= UG). The postulation of UG was claimed to be justified by the two-pronged 'poverty of the stimulus' argument: First, it seemed that the child could not acquire its native language without the aid of UG, because the input data it encounters is "degenerate", and "ungrammatical utterances do not come labelled as ungrammatical". Second, there seemed to be no methods of induction or analogy that could drive the language-acquisition process, which was – again – taken to mean that the child needs the aid of a very complex UG.

Every component of the 'poverty of the stimulus' argument can be, and has been, disproved: "The ungrammaticality of everyday speech appears to be a myth with no basis in actual fact" (Labov 1972: 203).

Ungrammatical utterances do come labelled as ungrammatical: "It would be possible to recognize that someone is [correcting a slip of the tongue] even without knowing his language" (Wittgenstein 1958, I, § 54).

The notion of analogy has been given an 'existence proof':

> The problem that has to be solved is defined by the three representative examples given above. Each of them illustrates the case where three sentences A, B, and C fit the pattern A:B = C:X, and where we intuitively feel that we can solve X, because its relation to C is the same as the relation that B bears to A. The problem is to find a systematic way to formalize this intuition. The solution is given in the Appendix. It derives its interest from the claim, made by Chomsky and his followers, that the problem is unsolvable (Itkonen 2005a: 93–94).

Finally, and in fully traditional terms, analogy has been acknowledged to be the driving force of language-acquisition: "young children make analogies across whole utterances" (Tomasello 2003: 144; for general discussion, cf. Itkonen 1996: 478–483; 2005a: 67–76, 89–98, Appendix).

Since 2005, generativism has changed its focus. Now the importance of UG has strongly diminished. Beginning with Chomsky (2005), the main burden of language-explanation has shifted onto the 'third factor', which is intended to subsume every potentially explanatory principle that is **not** specifically linguistic. As such, it includes not just evolutionary theory (= our EEs) but also the laws of physics (= our DEs): "If [the strong minimalist] thesis were true, language would be something like a snowflake, taking the form it does by virtue of natural law, in which case UG would be very limited" (Chomsky 2011: 26).

Physics is regarded as the 'basic science', which is needed anyway. Therefore physical explanations are thought to come "for free", simplifying the language-explanation in an equal measure. But there is a gap in this argument: "Invariant laws of nature ... set the channels in which organic design must evolve. But the channels are so broad relative to the details that fascinate us!", as Gould (1989: 289) puts it. In other words, and as noted by Mayr (2004: 50–51), the fact that physics and evolutionary theory are (and must be) compatible does not mean that the latter has been explained by the former. Johansson (2011: 13) makes the same point in the following terms:

> Similarly, the shape of a bird's wings and feathers do come from the physical laws of aerodynamics, sort of. But they do not come for free. ... The only role of the physical laws of aerodynamics in this process is to determine which shapes provide better flying abilities. The actual shaping has to be done through normal evolutionary and developmental processes.

As was noted above, the third factor is given a purely negative definition, comprising anything that is not linguistic. Therefore Johansson's (2011) conclusion seems inescapable: "'The' third factor is a vague catch-all category, mixing entities with totally different causal and epistemological status, rendering its analytical value highly dubious." What we are given are just physical metaphors (in the "language-as-snowflake" style), in addition to those evolutionary metaphors that have been given so far. In sum, the validity of, and the need for, DEs in linguistics remains to be demonstrated.

The concluding section of Itkonen (1996) bears the following title: "Chomskyan Linguistics Is an Explanans in Search of an Explanandum". This overall assessment of the generative enterprise seems even more justified in 2013 than it was in 1996:

> In the present context it is of no importance that Chomsky's theory of syntax has undergone several modifications. What is important, is the fact that while he has continued to analyze the syntax of English by means of self-invented sentences which his own linguistic intuition deems either correct or incorrect, the interpretation of, and the justification for, what he is doing has changed completely: from antimentalist distributional analysis he has moved first to mentalist syntax and then to biology [and now to physics].
>
> Once generative syntax had been invented, something had to be done with it, i.e. it had to be used to 'explain' something. With the passing of time, the explanandum has been conceived of in increasingly ambitious terms: having started with the arrangements of English morphemes, Chomsky has now arrived at theoretical biology [and finally at physics]. Seen in perspective, innatism and modularity [and the third factor] are not claims with empirical content. They are just excuses for Chomsky not to do anything different from what he has always done (p. 498).

Although it requires a wide historical perspective to clearly discern this pattern of successive self-redefinitions, it is quite interesting to note that Hymes & Fought (1981: 242) were already able to do so.

6. 'Pseudo-Deterministic' Explanations

Implicational universals have been called "the paradigm example of typological generalization" (Croft 2003: 54), and they are standardly formulated in accordance with the D-N schema. I claimed in Section 5 that the D-N schema does not automatically produce genuine DEs, and now I shall substantiate this claim. More precisely, I shall single out two topics, namely the confirmation and the explanatory force of D-N-styled implicational universals. For a more detailed discussion of these and other related issues, the reader is referred to Itkonen (2013b).

Universal implications that constitute the main premises of D-N explanations give rise to the so-called paradoxes of confirmation. Consider a sentence like 'All ravens are black', formalized as $\forall x(Rx \rightarrow Bx)$. Obviously, it is confirmed by (the occurrence of a black raven, expressed as) $Ra \,\&\, Ba$ and falsified by (the occurrence of a non-black raven, expressed as) $Rb \,\&\, \sim Bb$. This agrees with the common-sense view that, for an implication *if p then q* to be either confirmed or falsified, the antecedent *p* must be true. However, since $\forall x(Rx \rightarrow Bx)$ is, by contraposition, logically equivalent to $\forall x(\sim Bx \rightarrow \sim Rx)$ (= 'All non-black things are non-ravens'), it paradoxically follows that the sentence 'All ravens are black' is also confirmed by $\sim Bc \,\&\, \sim Rc$, i.e. by anything that is neither black nor a raven. Even more paradoxically (if possible), since $\forall x(Rx \rightarrow Bx)$ is also logically equivalent to $\forall x(\sim Rx \vee Bx)$, anything that is not a raven (expressed by $\sim Rd$) suffices to confirm the sentence 'All ravens are black', and so does anything that is black (expressed by Be). This follows from the fact that in propositional logic the ('material') implication *if p then q* is true when *p* is false (i.e. when either $\sim p \,\&\, q$ or $\sim p \,\&\, \sim q$ is true) or when *q* is true (i.e. when either $p \,\&\, q$ or $\sim p \,\&\, q$ is true).

Hempel (1965b [1945]), who originally pointed out the paradoxes, was willing to accept these consequences, and he argued that the feeling of paradox is a "psychological illusion".

For him, the important thing is that, from the strictly logical point of view, "there is no object which is not implicitly referred to by a hypothesis of this type" (p. 18). In other words, a sentence like 'All ravens are black' is not, logically speaking, not just about ravens but about **all** things in the universe. This is a startling result. In his 'Postscript', to be sure, Hempel (1965d [1964]) adds that, since he has been dealing only with the qualitative notion of confirmation, he can afford to admit that $\forall x(Rx \rightarrow Bx)$ is confirmed **strongly** by *Ra & Ba* and **very weakly** by *~Ra & ~Ba* (or *~Rd*, or *Be*). But this is unconvincing; or, as Brown (1977: 29) puts it, "scientific research is not conducted in this manner".

To start with, the problem with implicational universals (as currently understood) may be illustrated by means of Greenberg's (1966) Universal 3: 'All VSO languages are prepositional'. According to the standard interpretation of universal implications, this sentence is confirmed not just by the languages with VSO & PREP, but also by any not-VSO languages. But this means confusing confirmation with non-falsification and thus committing the Hempel-type fallacy. In the same vein, all claims about polysynthetic languages are automatically 'confirmed' by analytical languages (just because they are **not** polysynthetic). Even more paradoxically, all claims about sign languages are automatically 'confirmed' by the mere existence of oral languages.

Let it be added that the foregoing caveats do not apply to purported implicational universals that involve binary predicates, for instance: A) 'All languages with a non-zero morpheme for the singular have a non-zero morpheme for the plural'. In this case, contraposition yields an equally valid universal: B) 'All languages with a zero morpheme for the plural have a zero morpheme for the singular'. Interestingly, and contrary to formal logic, A and B are **not** equivalent, as shown by the fact that they are (directly) confirmed by opposite types of language, namely A) by Swahili and B) by Chinese.

What is the explanatory force of implicational universals? They typically correlate two (or more) predicates. But correlations are not explanatory in themselves. They do not tell us **why** the predicates are correlated. The why-question can be answered only by giving an account of the underlying causal mechanism. (Of course, this can sometimes be expressed in the correlation itself; cf. 'All pieces of metal expand when heated'). Since we are speaking of human beings, it has to be some sort of **mental** causation, which means that what we have to do is discover those REs which operate 'inside' the D-N schema (cf. Sect. 2). This is what is meant by 'pseudo-determinism'.

Let us have a closer look at how the SG vs. PL distinction is expressed. As far as the universals A/B are concerned, there are four logically possible cases: (i) SG ≠ PL (meaning that SG and PL are non-identical but of equal length), (ii) SG > PL, (iii) SG < PL, (iv) SG = PL. The occurrence vs. non-occurrence of these four cases is explained by **iconicity**. It is identified here as the mentally effective principle according to which "what is ontologically less vs. more is expressed by what is linguistically less vs. more". Thus, (iii), exemplified by English, qualifies as iconic or prototypical, whereas (ii), exemplified by no language, qualifies as **anti**-iconic. By contrast, (i), exemplified by Swahili, and (iv), exemplified by Chinese, qualify merely as **non**-iconic. It follows that while (iii) is 'strongly' explained, (i) and (iv) are explained 'more weakly', namely in the sense that at least they, unlike (ii), do not violate iconicity.

As for the 'if VSO, then PREP' universal, it should be seen as part of a larger pattern. It is often assumed that head vs. modifer constructions exhibit the following type of cross-linguistically valid 'harmony' or 'symmetry': either VO, ADP N, N A, N GEN or OV, N ADP, A N, GEN N. Whatever the truth-value of this claim, it should be clearly understood that we are dealing with the well-known notion of **analogy**, and not with such ad hoc notions as 'harmony' or 'symmetry'. The ubiquitous nature of analogy, documented in Itkonen (2005a), makes it imperative to postulate "an innate faculty of analogizing", as suggested by Anttila (1989 [1972]: 103). It follows that the universals discussed here would ultimately be given the same explanation, given that iconicity is a special instance of analogy (cf. Itkonen 2005a: 3.2).

There have been various attempts in the literature to explain implicational universals in terms of such (overlapping) principles as analogy, iconicity, economy, efficiency, and salience. In Itkonen (2013a) they are all subsumed under the umbrella term 'expressive needs', in deliberate reference to Coseriu's (1974 [1958]) corresponding term *Ausdrucksbedürfnisse*.

Deterministic (physical) causation operates always and everywhere, whereas mental causation does not. To be sure, it is to be expected that in recurrent types of situations people have acted and will act in the same or at least similar way, if this is the rational thing to do. But similarity as such is not explanatory. Rather, it is produced, and explained, by the functioning of (the referent of) RE, as explicated in Section 2:

> Of course, it is possible to state any number of **generalizations** about such linguistic changes as have been observed to occur. But generality is not the same thing as nomicity. The former is non-explanatory while the latter is non-existent in diachronic and/or typological linguistics (Itkonen 2011: 206–207).

In other words, generalizations are aggregates of individual (sub)actions each of which is explained by its own instance of mental causation, and **not** by being subsumed under some generalization. REs are always operative, either in the absence of generalizations or **within** generalizations. The idea of 'generality without nomicity' is illustrated by the heterogenous collection of 'grammaticalization paths' given in Heine & Kuteva (2002).

It is not without interest to note in this context that such a champion of EEs as Stephen Gould has mounted an all-out attack against the alleged intellectual superiority of DEs:

> But historical science is not worse, more restricted, or less capable of achieving firm conclusions because experiment, prediction, and subsumption under invariant laws of nature do not represent its usual working methods. The sciences of history use a **different mode of explanation**, rooted in the comparative and observational richness of our data (1989: 279; emphasis added).

EE and RE share the characteristic of being non-deterministic; otherwise they are different. But linguists ought to have the courage to follow Gould's example and practice their own type of explanation (i.e. RE) without any guilt feelings.

7. Statistical Explanation

It is interesting to note that it is possible to deny the notion of (physical) determinism, which was taken for granted in Section 5:

> It may in some logically consistent way always be maintained that the unfolding of the universe in time is a deterministic phenomenon, and we simply do not have the clue to the details, but such a view is, if not logically inconsistent, highly improbable and not supported by the evidence (Suppes 1984: 26).

So much is clear, in any case, that there is no determinism in sociology (including sociolinguistics). Something else is needed.

In propositional logic the truth-value of a complex sentence is determined by the truth-values of its simple sentences. In sociology (as well as in sociolinguistics) the sentences *a* and *b* plus their truth-values are replaced by the classes A and B plus the relations of inclusion between A and B. In practice, the symbol A means – ambiguously – both a class and a property, as when we speak of 'the class of those entities which have the property A'. The sentence-negation *∼a* is replaced by the complement class *A which stands – in principle – for those entities which are not A. It is often the case, however, that *A does not just mean the lack of A but the opposite of A. The simplest type of variable is a dichotomous variable x-a, represented here by A/*A.

The analogy between sentences and classes may be illustrated as follows. An equivalence $a \equiv b$ is true when either both *a* and *b* are true or when both *a* and *b* are false, and false otherwise (Fig. 1). Notice that the truth of *∼a* and *∼b* equals, respectively, the falsity of *a* and *b*. If we assume that there are 100 entities one half of which are A & B while the other half are *A & *B, then we have a 'class equivalence', or a perfect correlation, between the classes A and B (Fig. 2).

	a	∼a
b	True	False
∼b	False	True

Fig. 1. Equivalence

	A	*A
B	50	0
*B	0	50

Fig. 2. Class Equivalence = Perfect Correlation

Perfect correlation contrasts with the lack of correlation, on the one hand (Fig. 3), and with perfect negative correlation, on the other (Fig. 4).

	A	*A
B	25	25
*B	25	25

Fig. 3. No correlation

	A	*A
B	0	50
*B	50	0

Fig. 4. Perfect Negative Correlation

Sociological research typically deals with less than perfect correlations, i.e. with cases that exemplify the notion of 'weak equivalence', for instance with such numbers as those in Figure 5.

	A	*A		
B	35	20	|	55
*B	15	30	|	45
	50	50	|	100

Fig. 5. Weak Equivalence

AB will stand for the intersection of A and B, while X/Y will stand for 'X divided by Y'. Then in Fig. 5 the **relative frequency** of those who are both A and B among those who are A is AB/A = 35/50 = 0.7. Let us assume that we are investigating the dichtomous variables x-a (= obesity) and x-b = (cardiac trouble), represented in Figure 5 by the classes A/*A and B/*B, respectively. The relative frequency of 'cardiac fatties' in our data is, as noted, 0.7, while the relative frequency of 'non-cardiac fatties' is A*B/A = 15/50 = 0.3.

One should clearly understand the difference between the relations AB/A and AB/B. In both cases AB represents the class of those persons who are both fat and cardiac. But this class may be related in different ways to more inclusive classes, namely either to the class A or to the class B. AB/A represent the portion of cardiacs among fatties, i.e. 0.7, whereas AB/B represents the portion of fatties among cardiacs, i.e. 35/55 = 0.64. With AB/A, B is the attribute class while A is the reference class: we investigate the incidence of cardiac trouble among fat people. With AB/B, A is the attribute class while B is the reference class: we investigate the incidence of obesity among cardiacs.

The relative frequency 0.7 says that 70% of fatties are cardiacs. The same number can also be taken to express **conditional probability**: it says that a fatty will be a cardiac with the probability 0.7. The standard notation is *p(B|A) = 0.7* ("the probability from A to B is 0.7"). The implicit assumption is that the direction of causation goes from A to B: "If you do A, you are likely to have B." Here I have adopted the 'frequentist' interpretation of probability, according to which "probability is defined in terms of the limit of the relative frequency of the occurrence of an attribute in an infinite sequence of events" (Salmon 1966: 83). That is, it is assumed in our example that the observed relative frequency 0.7 represents the value to which the relative frequency of B converges in a potentially infinite class A. For simplicity, the problems relating to sampling are assumed to be solved.

As shown by Figures 2–5, correlations between two dichotomous variables x-a and x-b may vary greatly, which makes it mandatory to define some **measure** for the strength of correlations, called '**correlation** coefficient'. Applied to the data of Figure 5, the correlation coefficient *f-ab* is as follows: f-ab = AB/A − *AB/*A = 35/50 − 20/50 = 0.7 − 0.4 = 0.3. As an instance of 'explication', this definition will be justified in Section 9.

Next, we shall define the **causal** coefficient, simplifying somewhat the account given in Boudon (1974: Chaps 2–3). Looking at Figure 5, we might think at first that the (proportional) class AB/A consists of 35 persons who are cardiacs **because** they are fat: they have B because they have A. (I disregard here the inverse possibility that obesity might be caused by cardiac trouble.) On reflection, however, we realize that this cannot be right. The (proportional) class *AB/*A consists of 20 persons who are cardiacs although they are **not**

fat: they have B although they are not A. Therefore their cardiac trouble must be due to some **other cause** than obesity; this unknown cause will be designated by x-z. (We can also say that these persons have Z.) If we assume that we have a fair sample, i.e. that the classes AB/A and *AB/*A are **homogeneous**, except that the members of AB/A have A whereas the members of *AB/*A have *A, then we must assume that the proportion of those who have B due to x-z must be the same in AB/A and *AB/*A.

Ex hypothesi, there are two distinct causes, i.e. the variables x-a and x-z, operating inside AB/A. Accordingly, AB/A must be divided in two, and we must find the values (= causal coefficients) for x-a and x-z. The total value of AB/A is 0.7. The value of x-z must be the same as that of *AB/*A, i.e. 20/50 = 0.4. Now we get the causal coefficient *g-ab* for x-a by subtracting the value of x-z from the total value: g-ab = 0.7 − 0.4 = 0.3.
Thus, the correlation coefficient and the causal coefficient coincide in this maximally simple case. With three or more variables, however, these two values will in general be seen to diverge (cf. below).

Next, the notion of **statistical explanation** will be illustrated with the aid of a more realistic example. Boudon (1974: 71) assumes two sociological variables with the following interpretations: x-b = ‘good vs. bad first experience at work’, x-c = ‘good vs. bad integration of immigrants within the host society’:

	B	*B	
C	374	120	494
*C	206	300	506
	580	420	1000

Fig. 6

Next, a third variable is introduced to further analyze the same data, namely x-a = ‘good vs. bad family integration’:

	A			*A		
	B	*B		B	*B	
C	294	72	366	80	48	
*C	126	108	234	80	272	
	420	180	600	160	400	1000

Fig. 7

In Figure 6 the probability for C to occur is p(C) = 494/1000 = 0.49. The conditional probability, or the probability from B to C, is p(C|B) = 374/580 = 0.64. We are investigating here the occurrence of the attribute class C in the reference class B, and we observe that B is **statistically relevant** to C, because it increases the probability for C to occur: p(C) < p(C|B).

According to Salmon (1971), statistical explanation consists in making the probability of the attribute class increase by restricting the reference class, i.e. by making it more homogeneous with respect to the attribute class. This happens when the data of Figure 6 is reanalyzed by taking the class A into consideration, which produces Figure 7. Now we get the probability of C in the new reference class AB: p(C|A&B) = 294/420 = 0.7.

The former reference class B has been divided into two (sub)classes AB and *AB, of which the former is more homogeneous than the class B with respect to C (and thus constitutes the new reference class), whereas the latter is less homogeneous with respect to C (i.e. the probability of C in *AB is 80/160 = 0.5). Because of its opacity, the definition of statistical explanation (=SE) will be accompanied by an example based on Figures 6–7:

SE: The probability of the attribute class is made to increase by restricting the reference class

SE:	p(C|*A&B) = 0.5 <	p(C|B) = 0.64 <	p(C|A&B) = 0.7
	stage 2	stage 1	stage 2

This version of SE is formulated in terms ‘statistical relevance’, as measured by correlations and probabilities. The following significant correction has been added afterwards by Salmon (1984):

> It now seems to me that the statistical relationships specified in the S[tatistical]-R[elevance] model constitute the **statistical basis** for a bone fide scientific explanation, but that this basis must be supplemented by certain **causal factors** in order to constitute a satisfactory scientific explanation (p. 34; original emphasis).

Such factors may of course be represented by more or less ad hoc means intended to ‘depict’ the causal mechanism ‘behind’ the correlations. But the elicitation of causality is built-in into the very formalism of Boudon (1974). Let us show this with the aid of Figure 7. The membership of the classes involved are as follows:

A = 6000, *A = 400
B = 420 + 160 = 580, *B = 180 + 240 = 420
C = 366 + 128 = 494, *C = 234 + 272 = 506

The correlation coefficients between the independent or explanatory variables x-a and x-b and the dependent variable x-c are as follows:

f-ac = AC/A − *AC/*A = 366/600 − 128/400 = 0.29
f-bc = BC/B − *BC/*B = (294 + 80) / (420 + 160) − (72 + 48) / (180 + 240) = 0.36

We also observe a correlation between the independent variables x-a and x-b:

f-ab = AB/A = 420/600 − 160/400 = 0.3

Of course, the nature of the variables (= independent or dependent?) cannot be seens from the mere numbers (e.g. of Fig. 7) and their correlations, but results from an **interpretation** of the data.

How do we compute the causal coefficients in the case of three variables? Let us consider the (proportional) class ABC/AB, which contains 294 members. On the basis of what has been said above, we know that it must be divided into three subclasses: those who have C because they have A; those who have C because they have B; those who have C because they have Z. This tripartite structure can be represented as follows:

ABC/AB = g-ac + g-bc + g-zc

The value g-zc can be computed straight away:it is represented by the class whose members have C due to other causes than A or B (i.e. because they have Z). Because the occurrence of both *A and *B is required, it must be the class *A*BC/*A*B. From Figure 7 we get the numerical value for this class:

g-zc = *A*BC/*A*B = 48/240 = 0.2

Now that we know the value g-zc, we can compute the value g-ac. It cannot be the case that the members of A*BC/A*B have C because they have B (since they, instead, have *B), which means that they must have C either because they have A or because they have Z:

A*BC/A*B = g-ac + g-zc
g-ac = A*BC/A*B − g-zc
g-ac = 72/180 − 48/240 = 0.4 − 0.2 = 0.2

In the same way we can compute the value g-bc. The members of *ABC/*AB have C, but its cause cannot be A (since they have *A). Therefore the cause must be either B or Z:

*ABC/A*B = g-bc + g-zc
g-bc = *ABC/A*B − g-zc
g-bc = 80/160 − 48/240 = 0.5 − 0.2 = 0.3

The correlation coefficients and the causal coefficients differ from one another in the following way:

f-ac = 0.29 ≠ g-ac = 0.2
f-bc = 0.36 ≠ g-bc = 0.3

This is due to the fact that, as was noted above, there is a correlation with the strength 0.3 between the independent variables x-a and x-b. This situation may be depicted with the aid of the following **causal model**:

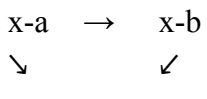

Fig. 8

The causal coefficients g-ac and g-bc express the **direct** influence of x-a on x-c, on the one hand, and of x-b on x-c, on the other. The correlation coefficients f-ac and f-bc also contain the **indirect** influence which is due to the correlation between x-a and x-b: in addition to directly influencing x-c, they also influence x-c 'through each other'. It has to be emphasized once again that the unidirectional arrows of Figure 8 result from an interpretation that 'transcends' the mere numbers contained in this Figure. Ordinary common sense is often enough; we know, for instance, that what happens before cannot be caused by what will happen afterwards. Taken as such such, not only correlation coefficients like f-ac but also causal coefficients like g-ac lack any kind of directionality.

The default assumption is that the causes x-a and x-b influence the effect x-c always with the same strength, regardless of whether they occur alone or together. This means that when they occur together, the common strength of x-a and x-b is the sum of the individual strengths of x-a and x-b. It is precisely this case that is represented by the numbers of Figure 7:

ABC/AB = g-ac + g-bc + g-zc
0.7 = 0.2 + 0.3 + 0.2

This default case is called **additive** causality. It is also possible, however, that when x-a and x-b occur together, their common strength is either greater or smaller than their sum. In such a case there is an (either positive or negative) 'interaction' between x-a and x-b, which means that we have to do with **interactional** causality (cf. Itkonen 1983: 19–21). Interactional structures with three variables are formalized in Boudon (1974: Chap. 5). They are summarized in Itkonen (2003b: 187–188). At the next stage, the formalization is extended so as to cover 'general' (= additive and/or interactional) structures with an arbitrary number of variables (Boudon 1974: Chap. 9), which is mathematically quite demanding.

The significance of the Salmon/Boudon-type analysis is immediately evident to anybody familiar with the 'variationist paradigm' that William Labov introduced into sociolinguistics in the early 1970's. One of the phenomena that has been the most intensely studied within the variationist framework is the loss vs. maintenance of the word-final *t/d* in spoken English (cf. Labov 1972: 216–226). This (dichotomous) dependent variable (= x-a) has been been explained on the basis of the following three independent variables, of which the first is phonological, the second morphological, and the third sociological in character: the word that follows *t/d* either begins or does not begin with a consonant (= x-b); *t/d* either does not or does express the grammatical meaning 'past tense' (= x-c); *t/d* either does or does not occur in the lower social class (= x-d).

Following Boudon (1974), I computed in 1977 the causal coefficients for the variables x-b, x-c, and x-d as well as for the interaction between x-b and x-c (cf. Itkonen 1980: 360–363). There is also a second-level interaction between this interaction and the variable x-d, but the

technology that was available to me at the time did not permit the computing the corresponding value. Two things should be noted concerning the analysis of the loss vs. maintenance of *t/d*.

First, the data is explained by SE, by means of a causal model similar to Figure 8, except that there are three independent variables and two (either first-level or second-level) interactions (op.cit., p. 362).

Second, the operation of each independent variable is further eplained by its own RE: it is rational (or energy-saving) to avoid consonant clusters (= x-b), especially if they have no grammatical function (= x-c); and it is rational – ceteris paribus – not to deviate from one's native dialect (= x-d).

Just as at the end of Section 6, we have now arrived at the idea of REs operating within generalizations or regularities. The same idea has been expressed by Harré & Secord (1972: 133) as follows:

> While the statistical method is ... a reasonable way of trying to discover and extend the critical description of social behavior, it is impossible to use it as the method for discovering the **generative mechanism** at work in social life ... The processes that are productive of social behavior occur in individual people, and it is in individual people that they must be studied (emphasis added).

On the other hand, it goes without saying that there is a huge number of legitimate questions which can be answered only by taking into account large quantities of social (including linguistic) data, and in all such situations the statistical approach becomes a necessity. Moreover, the discovery of causality is the overriding purpose also in this type of research: "Nous avons vu que l'explication sociologique consiste exclusivement à établir des rapports de causalité ..." (Durkheim 1973 [1895]: 124).

The ensuing conundrum of **statistical causality** (and explanation) can be summed up as follows (cf. Itkonen 1983: 24–31). The typical starting point is constituted by a situation like the one given in Figure 6. The components BC = 374 and *B*C = 300 are (or seem) unproblematic: The occurrence of the positive value B of the variable x-b favors the occurrence of the positive value C of the variable x-c (or vice versa), and the occurrence of the negative value *B of the variable x-b favors the occurrence of the negative value *C of the variable x-c (or vice versa); and it is natural – at least initially – to provide this correlation with an interpretation such that, for instance, B has the tendency to **cause** C. But the components B*C = 206 and *BC = 120 are problematical. With B*C the problem is that C fails to occur although B occurs. With *BC the problem is that C occurs – 'spontaneously', as it were – although B fails to occur. It is the purpose of statistical explanation to **reduce** the numbers that represent the problematic cases B*C and *BC, and this happens by taking **new** (independent) variables into consideration, in addition to x-b. Precisely this process is illustrated by the transition from Figure 6 to Figure 7. But whatever the number of new variables, one never manages to eliminate statistical variation altogether and thus to achieve deterministic explanation (or what appears to be such).

The same general problem has been formulated by Suppes (1984) as follows:

> Because one is not endorsing determinism as a necessary way of life for biological and social scientists, it does not mean that the first identification of a probabilistic cause brings a scientific

investigation to an end. It is a difficult and delicate matter to determine when no further causes can be identified. I am not offering any algorithms for making this determination.

Statistical variation may be experienced as either incomprehensible or annoying, but one should learn to overcome such feelings. First, if – *per impossibile* – we could inspect separately each of those thousands (or even millions) of individual cases that constitute a social regularity, we would understand and explain each of them much better, namely on the basis of its own individual RE. Second, in social matters some latitude should always be left for chance and/or free will.

8. Coherentist Explanation

It is customary to distinguish between (at least) three 'theories of truth', based on the following notions: correspondence, coherence, consensus. Depending on one's predilections, one or another of these three may come to be emphasized. Certainly any complete account of truth must take all of them into consideration (cf. Itkonen 1983: 110–129).

A strong version of the coherence theory of truth produces what might be called 'coherentist explanation' (= CE), defined as follows:

CE: X1 as part of Y is explained by Y if, and only if, Y is a coherent whole constituted by X1, X2, X3, etc.

Sections 2–7 have dealt with the following types of explanation: RE, FE, EE, DE, SE. These five types are different from each other (even if, for instance, SE may 'include' RE). By contrast, CE is shared by all the other types. Therefore one might regard CE as redundant. It will be seen, however, that CE is apt to highlight one generally neglected aspect of explanation in linguistics.

As shown by Rescher (1973), it is exceedingly difficult to define what it means, exactly, for two or more sentences S1, S2, etc. to 'cohere'. So much is clear that something more is required than mere compatibility (= lack of inconsistency) between S1, S2, etc. There must be **inferential** relations between S1, S2, etc., and the typical inferences are either **deductive** or **inductive** in character (cf. Bonjour 2002: 202–204). Deduction is well-understood whereas induction is not. Deductive coherence can be characterized as 'vertical', in two senses. First, it can be the transition from axioms to theorems, as in systems of logic in general. Second, it can be the transition from general laws to particular events, as in DE. (As noted before, DE has not succeeded in showing how less general laws are deduced from more general ones.) Inductive support between sentences has turned out to be impossible to formalize.

Philosophy of science has its own history, and increasing coherence – known under such labels as **incorporation** or **unification** – is the key to understanding how scientific progress has been conceptualized since the middle of the 19th century. Losee (1980) offers a revealing account of how this idea has developed:

"William Whewell compared [in 1847] the evolutionary development of a science to the confluence of tributaries to form a river. ... He cited Newton's theory of gravitational attraction as the paradigm of this growth by incorporation. Newton's theory subsumed

Kepler's laws, Galileo's Law of Free Fall ..." (p. 125). "He spoke of this incorporation as a 'consilience of inductions' ..." (p. 127).

"N.R. Campbell emphasized [in 1952] that ... the analogy is an essential part of a theory, because it is only in terms of the analogy that a theory can be said to explain a set of laws" (p. 138). "Campbell maintained that ... laws can be explained only by their incorporation in theories" (p. 139).

In 1961 Ernest Nagel distinguished between two types of reduction: the 'homogeneous' type is exemplified by "the 'absorption' of Galileo's law of falling bodies into Newtonian mechanics", while the 'heterogeneous' type is exemplified by "the reduction of classical thermodynamics to statistical mechanics" (p. 185). "Successful reduction is incorporation. One theory is absorbed into a second theory which has a broader scope" (p. 186).

More recently, Edmund Wilson (1998) has revived the 'consilience of inductions' in his meta-analysis of sociobiology. Kitcher (1998) too defends the notion of 'explanation/progress by unification'. He gives the following quote from Newton: "I wish we could derive the rest of the phenomena of Nature by the **same kind** of reasoning from mechanical principles ..."; and he notes that such passages "inspired Newton's successors to try to complete the unification of science by finding further force laws **analogous** to the law of universal gravitation" (p. 283; emphasis added). "The unifying power of Newton's work consisted in in its demonstration that one **pattern** of argument could be used again and again in the derivation of a wide range of accepted sentences" (*ibidem*; original emphasis).

It is crucially important that what Kitcher is dealing with "is not a pattern of the kind which interests logicians" (p. 285). By this claim, Kitcher distances himself from the Hempel-type tradition in the philosophy of science. Because pattern similarity, or **structural similarity**, is the very definition of analogy, Kitcher is eo ipso pleading for the importance of **analogy** in scientific progress; it certainly has more substance than deduction, if taken literally, can ever have (cf. Itkonen 2005a: 15–19, 176–186). It should also be noted that Kitcher sees (p. 299) his notion of explanation-as-unification confirmed by Wilson's sociobiology.

The content of the preceding paragraph may be summarized in the following slogan: **Analogy creates coherence**. Linguists of all people should understand this truth. It was the basic insight of structuralism that analogy (or 'proportional opposition') is the centripetal force that keeps any well-structured system together and establishes the identity of each unit in it (cf. Trubetzkoy 1958 [1939]: 60–66). This general idea may be illustrated by means of the (partial) phonological system given in Figure 9.

p	t	k
b	d	g
m	n	ɳ
f	s	χ

Fig. 9

What follows is a quotation from Itkonen (2005a: 76):

> In this type of system the phoneme /k/ is defined by four distinctive features, which are elicited by contrasting /k/ with its closest neighbours: it is a voiceless (as opposed to /g/), non-nasal (as opposed to /ɳ/), and velar (as opposed to /t/) occlusive (as opposed to /χ/). Phonological oppositions of this kind

are called **proportional**. They may be made more explicit by showing the precise place of /k/ in each of the four 'chains' of oppositions:

voiceless	p:b	=	t:d	=	**k**:g	
non-nasal	p:m	=	t:n	=	**k**:ŋ	
velar	p:t:**k**	=	b:d:g	=	m:n:ŋ =	f:s:χ
occlusive	p:f	=	t:s	=	**k**:χ	

Thus, the distinctive features of /k/ are identical with the **differences** between /k/ and its neighbours. 'Difference' between X and Y is just another term for the **relation** between X and Y. Now as can be seen from what precedes, the differences/relations between /k/ and its neighbours are the **same** as the differences/relations between other pairs (or triplets) of phonemes in each of the four chains of oppositions. Hence, by definition, there is an **analogy** between these pairs (or triplets).

We have here a perfect example of CE: the identity of each phoneme is **explained** by showing its place within a coherent whole.

It is good to note that analogy also functions at a level which is, if possible, even more fundamental:

It may take some mental effort to realize that the very concept of **structure**, as it applies within a single language, is based on analogy. This is illustrated e.g. by the structure of nouns or verbs in **any** language, inflectional or not. Consider the verbal inflection in Latin: *am-**o*** ('I love), *ama-**s*** ('you-SG love'), *ama-**t*** ('[s]he loves'), etc. Why are the grammatical meanings uniformly expressed by suffixes? Why is there not a suffixal ~ prefixal ~ infixal variation like e.g. *am-**o***, ***s**-ama*, *a-**t**-ma*? This question may sound silly, but it has to be answered; and the answer is, of course, analogy ... (op.cit., p. 8).

Again, we have a perfect example of CE.

Kitcher-type explanation-as-unification exemplifies **vertical** coherence because, dealing with natural sciences, it has to presuppose the basic division between general law and particular fact. Rescher's (1979) 'coherentist inductivism', with its "dialectical process of cyclical structure", represents the standpoint of **horizontal** coherence: "Rather than proceeding linearly, by fresh deductions from novel premisses, one may be in a position to cycle round and round the same given family of prospects and possibilities, sorting out, refitting, refining until a more sophisticatedly developed and more deeply elaborated resolution is ultimately arrived at" (p. 75). It is obvious at once that Rescher is here describing the so-called **hermeneutic cycle**. Its applications are better known in the context of human(istic) sciences, but clearly it is valid more generally. It goes without saying that the explanatory analogy of our linguistic examples exemplifies horizontal coherence.

Within anthropology and/or sociology, the explanatory force of horizontal coherence has been codified as **pattern explanation** (cf. Diesing 1972: 137–141, 235–243; Itkonen 1983: 35–38, 205–206), with obvious connections with FE (cf. Sect. 3). But the connection with RE (cf. Sect. 2) is just as evident:

The cogency of a teleological explanation rests on its ability to discover a **coherent pattern** in the behavior of an agent. Coherence here includes the idea of rationality both in the sense that the action to

> be explained must be reasonable in the light of assigned desires and beliefs, but also in the sense that the assigned desires and beliefs must fit with one another (Davidson 1975: 11; emphasis added).

As might be expected, there is also a connection with EE. Gould (1989) is clearly outlining some version of (horizontal) pattern explanation, when he makes the following remarks: "The sciences of history use a different more of explanation [i.e. different from DE] ..." (p. 279). "We search for repeated pattern, ... We know that evolution must underlie the order of life because no other explanation can coordinate the **disparate data** of embryology, biogeography, the fossil record, vestigial organs, taxonomic relationships, and so on" (p. 282; emphasis added). The closest analogue within linguistics is **etymological** research, which also has to coordinate disparate data, namely such as relate not just to language history but also to social history, archeology, and anthropology (cf. Anttila 1989).

Let us ask the following question: Can we also explain grammatical descriptions (as distinguished from what they are about)? Some people may think that this is an illegitimate type of question. I think it is perfectly legitimate. The type of explanation that first comes to mind is CE: One grammar G1 is explained (or 'explained') by showing how it fits into the general typological-descriptive framework which also accommodates grammars G2, G3, G4, and so on. Itkonen (2005b) shows in a preliminary fashion how this is done.

To sum up: CE is being practiced all the time by everybody, but its existence is seldom acknowledged explicitly. The level of methodological self-understanding should be raised accordingly.

9. Explication (= Conceptual Analysis)

In its pretheoretical use the word 'explanation' means something like 'increasing the degree of understanding'. The prototypical instance is getting to know **why** X happened. RE, EE, DE, and SE try to answer this question, each in its own way and applied to its own type of data The question answered by FE is slightly different: why is X 'there'? And when CE purports to tell us why X (e.g. a phoneme) is what it is, the question to be answered becomes indistinguishable from asking: **what** is X? Prototypically, however, this question is answered by conceptual analysis, also called **explication**.

Explication is the principal method of philosophy. Chapter 11 of Itkonen (1978), taking its cue from Pap (1958), is in its entirety devoted to 'explicating explication'. Here it is enough to say that explication means analyzing some pretheoretical, intuitively known concept in such a way that a set of 'criteria of adequacy' are satisfied by the outcome of the analysis.

To give an example, let us return to defining a measure for the **strength of correlation** (cf. Sect. 7). To start with, it is reasonable (although by no means necessary) to select the criteria of adequacy in the following way. The sought-after value should be some number between 1 and −1 in such a way that the perfect (positive) correlation between A and B (= Fig. 2) gets the value 1 whereas the perfect lack of correlation (= Fig. 3) gets the value 0; and if the perfect correlation obtains, not between A and B, but between *A and B, i.e. if there is the perfect negative correlation between A and B (= Fig. 4), then it should get the value −1.

In Figures 2 and 4 AB/A gets the respective values 50/50 = 1 and 0/50 = 0, whereas *AB/B gets the inverse values 0/50 = 0 and 50/50 = 1. If the former values are subtracted

from the latter ones, we get – just as we wished – 1 as the value of the perfect positive correlation and −1 as the value of the perfect negative correlation:

Figure 2: AB/A − *AB/B = 1 − 0 = 1
Figure 4: AB/A − *AB/B = 0 − 1 = −1

Moreover, in Figure 3 the numbers AB/A and *AB/B are the same, i.e. 25/50 = 0.5, which means that if one is subtracted from the other, we get 0 as the value of the perfect lack of correlation, just as we wished:

Figure 3: AB/A − *AB/B = 0.5 − 0.5 = 0

The proposed measure satisfies all three criteria of adequacy, and therefore it can (at least tentatively) be accepted as the definition of the strength of correlation, called 'correlation coefficient':

AB/A − *AB/B = f-ab

In sociology, interesting (positive) correlations always remain between $0 < 1$, i.e. they exemplify weak equivalences, i.e. less strict counterparts of $a \equiv b$. Of course there are such counterparts of the implication $a \rightarrow b$ as well, but the more equivalences and implications are weakened, the more they come to resemble one another.

To give another example of explication, let us consider the axiomatization of propositional **modal logic** (where *Lp* = 'It is necessarily true that p', *Mp* = 'It is possibly true that p', $p \vdash q$ = '*p* entails *q*'). What follows is based Itkonen (2003a: Chap. X).

There is no universal consensus on which modal formulae are or are not valid. "Nevertheless, there are certain conditions [= criteria of adequacy] which it seems **intuitively** reasonable to demand that a system should fulfil if it is capable of interpretation as a modal system" (Hughes & Cresswell 1972: 25; emphasis added). Here the criteria of adequacy are identical with a set of formulae which are definitely valid and must therefore be 'theses' (i.e. axioms, definitions, or theorems) of the system. Once this has been achieved, explication turns out to equal a **transition from intuitive to formal validity**. The following formulae must be among the theses:

(i) $Lp \equiv \sim M \sim p$, $Mp \equiv \sim L \sim p$ ('What must true, cannot be false', 'What can be true, is not necessarily false')
(ii) $Lp \rightarrow p$, $p \rightarrow Mp$ ('What must be true, is true', 'What is true, can be true')
(iii) $(p \vdash q) \equiv \sim M(p \ \& \sim q)$ ('It is not possible that what is entailed is false while what entails is true')
(iv) $[Lp \ \& (p \vdash q)] \rightarrow Lq$ ('Whatever a necessarily true formula entails must be necessarily true')

Hughes & Cresswell (1972) then proceed to "construct a number of axiomatic modal systems, each of which satisfies all of these requirements [= (i) – (iv)] but which differ from

one another in the presence or absence as theses of some of the less obviously valid formulae" (p. 25). As noted by Itkonen (2003a: 86), the authors apply here "the so-called **clear-case principle**, which is well-known also in the metatheory of linguistics: one concentrates on that type of data which is definitely known to be correct, and lets the grammar decide the less-than-clear cases".

Apart from the distinction between 'valid formula' and 'correct/grammatical sentence', there is an exact analogy between formal-logical axiomatizations and generative grammars, insofar they all are intent on generating all and only entities of the requisite kind (cf. Itkonen 1978: Chap. 10, 2003b: Chap. 17). The axiomatic ideal is inseparable from the striving after **simplicity**, as argued by Bloomfield, Hjelmslev, Harris, Chomsky, and Katz, among others (cf. Itkonen 2013a). Notice that axiomatics is an ideal which goes **against** the natural inclination of human thinking.

10. Conclusion

As documented in Itkonen (1991) and summarized in Itkonen (2013a), the history of Western linguistics is characterized by a tension between two distinct and even opposite descriptive goals: either axiomatics or causality; either presenting the data in the maximally simple, axiomatic form or discovering the (non-axiomatic) causal/functional mechanism that produces linguistic behavior. The former type of linguistic research, barely mentioned here in Section 9, has been analyzed in Itkonen (1978) and (2003b). The latter type of linguistic research was treated here especially in Section 2, and has been analyzed more in depth in Itkonen (1983), (2011), and (2013b).

References

Anttila, Raimo (1989 [1972]): *Historical and comparative linguistics*. Amsterdam: Benjamins.

— (1989): *Pattern explanation: The survival of the fit*. Diachronica.

Baker, G.P. & Hacker, P.M.S (1984): *Language, sense, and nonsense*. Oxford: Blackwell.

Benn, S.J. & Mortimore (eds) (1976): *Rationality and the social sciences*. London: Routledge.

— (1976): Technical models of rational choice. In Benn & Mortimore (eds.).

Bonjour, Lawrence (2002): *Epistemology: Classical problems and contemporary responses*. Oxford: Rowman & Littlefield.

Boudon, Raymond (1974): *The logic of sociological explanation*. Harmondsworth: Penguin Books.

Brown, Harold I. (1977): *Perception, theory, and commitment*. Chicago: Precedent Publishing Co.

Chomsky, Noam (1957): *Syntactic structures*. The Hague: Mouton.

— (2005): Three factors in language design. Linguistic Inquiry.

— (2011): Opening remarks.In Piattelli-Palmarini et al. (eds.): Of minds and language: A dialogue with Noam Chomsky in the Basque country. Oxford UP.
Cohen, L. Jonathan (1986): *The dialogue of reason*. Oxford: Clarendon Press.
Coseriu, Eugenio (1974 [1958]): *Synchronie, Diachronie und Geschichte*. München: Wilhelm Fink Verlag.
Croft, William (2003): *Typology and universals*. Cambridge UP.
Dahl, Östen (1980 [1975]): *Is linguistics empirical? A critique of Esa Itkonen's Linguistics and Metascience*. In Perry (ed).
Davidson, Donald (1973): "Freedom to act", in: T. Honderich (ed): *Essays on freedom of action*. London: Routledge.
— (1975): "Thought and talk", in: S. Guttenplan (ed): *Mind and language*. London: Routledge.
Diesing, Paul (1972): *Patterns of discovery in the social sciences*. London: Routledge.
Durkheim, Emile (1973 [1895]): *Les régles de la méthode sociologique*. Paris: Presses Universitaires de la France.
Gibson, Q. (1976): Arguing from rationality. In Benn & Mortimore (eds).
Giddens, Anthony (1976): *New rules of sociological method*. London: Hutchinson.
Givón, T. (2005): *Context as other minds.* Amsterdam: Benjamins.
— (2009): *The genesis of syntactic complexity*. Amsterdam: Benjamins.
Gould, Stephen Jay (1989): *Woderful life: The Burgess Shale and the nature of history*. New York: Norton & Company.
Greenberg, Joseph H. (1966): "Some universals of grammar with particular reference to the order of meaningful elements", in: J. Greenberg (ed): *Universals of grammar*. Cambridge, MA: The MIT Press.
Haiman, John (1985): *Natural syntax: Iconicity and erosion*. Cambridge UP.
Harré, Rom & P. Secord (1972): *The explanation of social behavior*. Oxford: Blackwell.
Haspelmath, Martin et al. (eds) (2005): *The world atlas of language structures*. Cambridge UP.
Heine, Bernd & Tania Kuteva (2002): *World lexicon of grammaticalization*. Cambridge UP.
— (2007): *The genesis of grammar*. Oxford UP.
Hempel, Carl G. (1965a): *Aspects of scientific explanation and other essays in the philosophy of science.* New York: The Free Press.
— (1965b [1945]): *Studies in the logic of confirmation*. In Hempel.
— (1965c [1959]): *The logic of functional analysis*. In Hempel.
— (1965d [1964]): *Postscript on 'Confirmation'*. In Hempel.
— (1965e): *Aspects of scientific explanation*. In Hempel.
Hollis, Martin (1977): *Models of man: Philosophical thoughts on social action*. Cambridge UP.
Hymes, Dell & John Fought (1981): *American structuralism*. The Hague: Mouton.
Itkonen, Esa (1974): *Linguistics and metascience*. Studia Philosophica Turkuensia II.
— (1978): *Grammatical theory and metascience*. Amsterdam: Benjamins.
— (1983): *Causality in linguistic theory*. London: Croom Helm.
— (1991): *Universal history of linguistics: India, China, Arabia, Europe*. Amsterdam: Benjamins.

— (1996): *Concerning the generative paradigm*. Journal of Pragmatics.
— (2003a): *Methods of formalization beside and inside both autonomous and non-autonomous linguistics*. University of Turku: Publicatins in General Linguistics 6.
— (2003b): *What is language? A study in the philosophy of linguistics*. University of Turku: Publications in General Linguistics 8.
— (2004): *T*ypological explanation and iconicity. Logos & Language.
— (2005a): *Analogy as structure and process: Approaches in linguistics, cognitive psychology and philosophy of science*. Amsterdam: Benjamins.
— (2005b): *Ten non-European languages: An aid to the typologist*. University of Turku: Publications in General Linguistics 9.
— (2008a): Review of Givón (2005). Language.
— (2008b): The central role of normativity in language and linguistics. J. Zlatev et al. (eds): The shared mind. Amsterdam: Benjamins.
— (2011): On Coseriu's legacy. Energeia. [Reprinted in Papers on typological linguistics. University of Turku: Publications in General Linguistics 15]
— (2013a): "Philosophy of linguistics", in K. Allan (ed): *The oxford handbook of the history of linguistics*. Oxford UP.
— (2013b): "Functional explanation and its uses", in S. Bischoff & C. Jany (eds): *Functional approaches to language*. Berlin: Mouton deGruyter.
Johansson, Sverker (2011): *Biolinguistics or physicolinguistics? Is the third factor helpful or harmful in explaining language?* Biolinguistics.
Kitcher, Philip (1998): Explanatory unification. In Klemke et al. (eds).
Klemke, E.D. et al. (eds): *Introductory readings in the philosophy of science*. New York: Prometheus Books.
Labov, William (1972): *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
Losee, John (1980): *A historical introduction to the philosophy of science*. Oxford UP.
Maddiesen, Ian (2005a)
— 2005a.
Mayr, Ernst (2001): What evolution is.
— (2004): *What makes biology unique?* Cambridge UP.
Nagel, Ernest (1961): *The structure of science*. New York: Harcourt.
Newton-Smith, W.H. (1981): *The rationality of science*. London: Routledge.
Nozick, Robert (2001): *Invariances: The structure of the objective world*. Cambridge, MA: Harvard UP.
Pap, Arthur (1958): *Semantics and necessary truth*. New Haven: Yale UP.
Paul, Hermann (1975 [1880]): *Prinzipien der Sprachgeschichte*. Tübingen: Niemeyer.
Perry, Thomas A. (ed) (1980): *Evidence and argumentation in linguistics*. Berlin: de Gruyter.
Popper, Karl (1957): *The povery of historicism*. London: Routledge.
Putnam, Hilary (1999): *The threefold cord: Mind, body, and the world*. New York: Columbia University Press.
Rescher, Nicolas (1973): *The coherence theory of truth*. Oxford: Clarendon.
— (1979): *Cognitive systematization*. Oxford: Blackwell.

Salmon, Wesley (1966): *The foundations of scientific inference*. University of Pittsburgh Press.
— (1971): *Statistical explanation and statistical relevance*. University of Pittsburgh Press.
— (1984): *Scientific explanation and the causal structure of the world*. Princeton UP.
— (1998): *Scientific explanation: How we got from there to here*. In Klemke et al. (eds).
Stegmüller, Wolfgang (1974): *Probleme und Resultate der Wissenschaftstheorie und Analytischen Phlosophie*, Band I: Wissenschaftliche Erklärung und Begründung. Berlin: Springer-Verlag.
Suppes, Patrick (1984): *Probabilistic metaphysics*. Oxford: Blackwell.
Tomasello, Michael (2003): *Constructing a language*. Cambridge, MA: The Harvard UP.
Trubetzkoy, N.S. (1958 [1939]): *Grundzüge der Phonologie*. Göttingen: Vandenhoeck & Ruprecht.
von Wright, Georg Henrik (1971): *Explanation and understanding*. London: Routledge.
Whitney, William Dwight (1979 [1875]): *The life and growth of language*. New York: Dover.
Wilson, Edmund O. (1998): *Consilience: The unity of knowledge*. New York: Random House.
Wittgenstein, Ludwig (1958): *Philosophical investigations*. Oxford: Blackwell.
Woodfield, Andrew (1976.): *Teleology*. Cambridge UP.
Wright, Larry (1976): *Teleological explanations*. Berkeley: University of California Press.