# 1 Introduction

The problem of linguistic data and evidence is in the air. A tiny spark suffices to kindle interest in it. In fact, new publications on the nature of linguistic data and evidence appear and conferences are held almost continuously. Adherents of different linguistic schools conduct – sometimes extremely heated – debates with each other on this topic. The question of which data count as evidence is one of the most important topics in linguistics, which has a direct and profound impact on the *fundamentals* of this discipline. This is, of course, by no means accidental. It is easy to see that the evaluation of the workability of conflicting theories depends heavily on what kinds of data can be regarded as evidence that either supports or refutes (in some sense) their hypotheses.

In their volume *What counts as evidence in linguistics*?, Penke and Rosenbach formulate the problem as follows:

> While thirty years ago linguists were still debating whether linguistics ought to be an 'empirical science' . . . , today we can safely say that this issue has been settled by and large and that nowadays most linguists will probably agree that linguistics is indeed an empirical science. What is being discussed is . . . not *whether* empirical evidence may or should be used, but rather, *what type of* empirical evidence, and *how* it is to be used. (Penke & Rosenbach 2004a: 480; emphasis as in the original)

Thus, the current debate on data and evidence centres around the following *methodological problem*, which is of great significance for the everyday research practice in linguistics, and which we will treat as the *Central Problem* of the debate:

(CP) (a) What types of data/evidence can be used,
and
(b) how does data/evidence work in linguistic theories?

We aim to contribute to the current debate on data and evidence from a perspective which has not been developed so far, although it is, in our opinion, substantial.[1] The *basic idea* of the book is based on three considerations.

First, tackling and solving (CP) *presupposes* the clarification of a series of further problems: 'What is a datum?', 'What is evidence?', 'What is the

structure of data/evidence?', 'What is the function of data/evidence in linguistic theories?' etc. Most of the current literature does not raise these issues, while the few attempts which touch on them are fragmentary and controversial.

Second, due to the nature of linguistic theorising, there is a close but hidden relationship between the *structure and function of data and evidence* on the one hand and the *argumentation structure of theories* on the other. Thus, so as to understand the former, it is indispensable – among other things – to reveal this relationship.

Third, the main shortcoming of the literature dealing with (CP) is that mechanically adopted metatheoretical prejudices overlap with the naïve reflection of the linguists on their everyday research practice. Therefore, in order to propose an adequate solution to the problems in connection with data and evidence, it seems to be necessary to develop a novel, coherent and sophisticated *metatheoretical* framework.

On the basis of these three considerations, we reformulate (CP) as the *Main Problem* of this book:

(MP) (a) What is the *structure* of linguistic data/evidence
and
(b) which *functions* do they fulfil in linguistic theories
*if*
(c) (a) and (b) are tackled by an argumentation-theoretic model of linguistic theorising?

(MP) will be divided into subproblems, addressed by different parts of the book. The solutions of these subproblems will lead, step by step, to the solution of (MP).

In Part I, we will provide a critical analysis of the current literature and reveal the solutions to (CP) which different approaches to linguistic data and evidence seem to have proposed. Thereby, we will raise several questions which the approaches to be discussed have left open. The open questions, along with the fact that the proposed solutions to (CP) are contradictory, will motivate the raising of (MP)(a) and (b) as well as the characteristics of the metascientific model referred to in (MP)(c).

The task of Part II will be the elaboration of this model – which will be called *the p-model of plausible argumentation* – and the exemplification of its workability by way of linguistic examples.

Part III, entitled 'Data and evidence', shows how the two notions which the book centres on can be defined with the help of the p-model.

The function of Part IV is to exemplify the workability of the p-model. It presents a detailed case study which illustrates how the p-model can be applied

to the practice of linguistic theorising. In this way it elucidates that the p-model is indeed capable of bridging the gap between linguistic research practice and metatheoretical reflection on this practice.

Finally, in Part V, we will summarise those conclusions on the basis of which we will obtain a possible solution to (MP).

*Part I*

# The state of the art

# 2 The problem $(P)_I$

In accordance with the structure of the book as outlined in Chapter 1, we begin with an analysis of the current literature discussing (CP). This analysis is expected to answer two questions. First, we will survey which solutions have been proposed for (CP) by the different approaches in the current discussion on linguistic data and evidence. The second question is motivated by our assumption that the problem of linguistic data and evidence may be solved by the elaboration of an appropriate metatheoretical model. Therefore, we have to clarify which metascientific background assumptions the proposals put forward in the literature as putative solutions to (CP) rely on.

Accordingly, the first subproblem of (MP) can be formulated as follows:

$(P)_I$ (a) What solutions have been proposed to (CP) in the current literature?
(b) What metascientific background assumptions can be revealed in the particular views?

Part I aims at

(a) the presentation and systematisation of views belonging to a significant – and from the point of view of the development of linguistics *relevant and progressive* – trend of the current debate on linguistic data and evidence;
(b) the critical analysis of these views on the basis of this systematisation;
(c) the solution of $(P)_I$ on the basis of systematisation and critical analysis, and
(d) the identification of the questions which the analysed views have left open and which motivate raising (MP), as well as the elaboration of a new metatheoretical model.

We cannot go into a detailed and thorough investigation of the historical roots of the contemporary debate on data/evidence nor can we provide an exhaustive analysis of the empiricalness debates in different fields of linguistics. We will solely highlight that trend which immediately motivates the problems raised in a certain part[1] of the most recent literature.

Chapter 3 will be devoted to the pre-history of the current debate on linguistic data and evidence. We will reveal the most important stages in the emergence of the *standard view of linguistic data and evidence*. This view presupposes the

dichotomy of introspective and corpus data, as well as *the standard view of the analytical philosophy of science*.

In Chapter 4, on the basis of the latest literature, we will show that the *practice* of linguistic research – that is, *object*-scientific inquiry – has clearly departed from the standard view of linguistic data.

In Chapter 5, we will examine whether, and if so, in what way and to what extent, *meta*scientific reflection follows linguistic practice in this respect. We will scrutinise the solutions to (CP) proposed by the current literature on linguistic data and evidence. However new the trend of the current debate about data/evidence we mentioned may be, it is enormously complex and multifaceted. It comprises several sub-tendencies, some of which partially overlap, while others run in opposite directions. We will focus solely on those approaches that can be regarded as *progressive* in so far as they explicitly question, at least partly, the standard view of linguistic data and evidence. Accordingly, we will not go into the case studies published in the literature but will discuss the following state-of-the-art articles (occasionally referring to further relevant publications as well): Borsley (2005a), Geeraerts (2006), Kepser & Reis (2005a), Lehmann (2004), Penke & Rosenbach (2004a); we will also refer to Schütze (1996).

In our answer to the question in $(P)_I$(a) we point out that none of the proposed solutions to (CP) is satisfactory. They are *progressive* insofar as they try to go beyond the shortcomings of the standard view of linguistic data in a problem-sensitive way. Nevertheless, contrary to this intention, they remain fragmentary at several important points and also include elements of the standard view of linguistic data. Although they recognise the untenability of the standard view of the analytical philosophy of science, they *still contain its remnants*. Therefore, they do not provide us with the systematic and elaborated metascientific framework that is needed for the sophisticated and comprehensive treatment of the problems raised. In this way, we will obtain the answer to $(P)_I$(b).

Accordingly, Chapter 6 will summarise the double-facedness revealed in the writings examined, and will also draw some far-reaching conclusions.

# 3 Historical background

## 3.1 Overview

In the present chapter we provide a concise overview of the historical antecedents of the current debate on linguistic data/evidence. First, we will summarise those tenets which shaped the methodological foundations of mainstream linguistics in the twentieth century and call them 'the standard view of the analytical philosophy of science'. Second, our analysis will reveal that the two predominant camps in linguistics from the middle of the 1950s to the end of the 1990s, corpus linguists and generative linguists, heavily rely on methodological assumptions which are in perfect accord with the standard view of the analytical philosophy of science. We will call the set of these assumptions 'the standard view of linguistic data and evidence'. It is this view which has been dominant over the past decades, right up to the turn of the last millennium, and which new approaches to linguistic data and evidence seriously question.

## 3.2 The problem of evidence in the philosophy of science

In order to solve $(P)_I$, we have to refer to the well-known circumstance that *the standard view of the analytical philosophy of science* played a decisive role in the emergence of the norms against which scientific theories have been evaluated for several decades in the twentieth century. Basically, the standard view of the analytical philosophy of science involved two approaches: logical empiricism initiated by the Vienna Circle and Popper's critical rationalism.[1] The reason these approaches have been subsumed under the standard view is that despite their differences they shared similar assumptions on a number of important issues.

First of all, both of them accepted Reichenbach's (1938) distinction between the 'context of discovery' and the 'context of justification'. The 'context of discovery' covers the creative, cognitive, social, historical etc. aspects of scientific discovery, theory formation and problem solving, while 'the context of justification' involves the logical reconstruction and the evaluation of the results of the discovery process, i.e. scientific theories. Both trends focused solely on the

*justification* of scientific theories, excluding the process of discovery, theory formation and problem solving from their field of interest.

Another central idea of the standard view of the analytical philosophy of science was that *empirical* theories must consist of statements which can be tested on the basis of a special subset of data, namely, *evidence*.[2] According to this idea – to put it in a very simplified manner, as a first approximation – the intuition underlying the notion of evidence was the following:

(E) Evidence
- (a) is objective;
- (b) serves as a neutral arbiter among rival hypotheses/theories;
- (c) is expected to justify (verify, falsify, confirm) hypotheses/theories;
- (d) is immediately given;
- (e) is primary to the theory;
- (f) is reliable.

The properties mentioned in (E) motivate the central role of evidence; they can be regarded as the common intuitive core of the different approaches to evidence within the standard view of the analytical philosophy of science.[3] Nevertheless, the interpretation of these properties is highly problematic. Therefore, in twentieth-century philosophy of science the discussions concerning the nature of evidence aimed at their explication, and also attempted to decide which of them were really relevant. In the course of the controversies, each of (E)(a)–(f) has been seriously questioned and each of them has been explicated in a variety of ways often incompatible with each other.[4] In certain cases, however, suggestions were put forward that seem to have been widely accepted.[5] During the debates quandaries and paradoxes were continuously raised (see primarily Goodman 1983 [1955]; for a detailed discussion of Goodmans' and others' paradoxes see Stegmüller 1970).

According to *inductivists* such as Carnap, the function of evidence is the *verification* of the hypotheses. Later, still within inductivism, it was also explicated in terms of *confirmation*, according to which evidence strengthens the hypothesis, or makes it more probable, without, however, proving its truth.[6] Thus, confirmation includes verification as an extreme case. Popper's *hypothetico-deductive* view highlighted the *falsificatory* function of evidence in the sense that, although it cannot verify a hypothesis, it is capable of falsifying (i.e., refuting) it.[7] Although these three explications – each of which has several versions – were developed in different periods of the analytical philosophy of science and were the focus of heated debates, they are logically related:

If $E$ is evidence for some hypothesis $H$, then $E$ makes it more likely that $H$ is true: in such circumstances, $E$ *confirms* $H$. On the other hand, if $E$ is evidence against $H$, then $E$ makes it less likely that $H$ is true: $E$ *disconfirms* $H$. *Verification* is the limiting