STUDIES IN FUNCTIONAL AND STRUCTURAL LINGUISTICS 64

Linguistic Theory and Empirical Evidence

EDITED BY
Bob de Jonge
Yishai Tobin

John Benjamins Publishing Company

Linguistic Theory and Empirical Evidence

Edited by

Bob de Jonge
University of Groningen

Yishai Tobin
Ben-Gurion University of the Negev

John Benjamins Publishing Company
Amsterdam / Philadelphia

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Library of Congress Cataloging-in-Publication Data

Linguistic Theory and Empirical Evidence / Edited by Bob de Jonge and Yishai Tobin.
p. cm. (Studies in Functional and Structural Linguistics, ISSN 0165-7712 ; v. 64)
Includes bibliographical references and index.
1. Linguistic analysis (Linguistics) 2. Corpora (Linguistics) I. Jonge, Bob de. II. Tobin, Y.
P126L53 2011
410.1--dc22 2011008170
ISBN 978 90 272 1574 1 (Hb ; alk. paper)
ISBN 978 90 272 8687 1 (Eb)

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

Linguistic Theory and Empirical Evidence

Studies in Functional and Structural Linguistics (SFSL)

Taking the broadest and most general definitions of the terms functional and structural, this series aims to present linguistic and interdisciplinary research that relates language structure — at any level of analysis from phonology to discourse — to broader functional considerations, whether cognitive, communicative, pragmatic or sociocultural. Preference will be given to studies that focus on data from actual discourse, whether speech, writing or other nonvocal medium.
The series was formerly known as *Linguistic & Literary Studies in Eastern Europe (LLSEE).*

For an overview of all books published in this series, please see
http://benjamins.com/catalog/sfsl

Volume 64

Linguistic Theory and Empirical Evidence
Edited by Bob de Jonge and Yishai Tobin

Table of contents

Introduction

Linguistic theory and empirical evidence

Bob de Jonge
University of Groningen

The relation between many linguistic, or otherwise psychological observations is unclear, not even demonstrable. A proposal to formalize (a 'generative grammar') the subconscious observations discovered by Freud and their influence on human behavior is eventually possible.

Freud has caused the greatest revolution in ages in Western thinking. He was certainly aware of this fact. 'But' he generously claims, 'my work is scientific'. In this confusion of tongues, it may be wished that linguistics never become a science.
Balk-Smit Duyzentkunst (1974: 16) [translation mine]

This introduction is meant to present the main theme of this book: the necessity of independent, objective demonstrations for linguistic analysis. It will compare and contrast different traditions in the demonstration of hypotheses within linguistics and try to define which kinds of demonstrations are considered to be independent and which are not. Finally, a short overview of the chapters will be given establishing their relationships to the general theme of the volume.

1. Introduction[1]

Intellectuals will not have any problem in stating that Sigmund Freud has been very important for (psychological) science as it is today. However, scientific methods have changed enormously over the last century. Whereas Freud depended largely on introspection, i.e. he based his hypotheses mainly on his own experiences and his own interpretations of them, nowadays social sciences generally reject this working method and use diverse kinds of observations and objective techniques to corroborate findings, a method typically used in the 'hard' sciences.

1. I am indebted to two anonymous readers of this volume for their comments on this article, and especially to Yishai Tobin for helpful comments and corrections in my English. All remaining flaws and errors are my own responsibility.

It is surprising that one of the oldest sciences, the study of languages and literatures, has advanced so relatively little in the technical procedures of the validation of hypotheses. However, Columbia School (CS) Linguistics does have an empirical tradition that includes independent validation. It is the purpose of this volume to continue and elaborate further this empirical tradition and to offer a variety of empirical studies for many and diverse languages to enrich CS research and linguistic analysis in general.

2. Validation in linguistics

Traditionally, linguists describe and explain a particular phenomenon in a given language by studying intuitively contrived or carefully selected examples in order to demonstrate their postulated hypotheses. Contini-Morava (1995:23–24) describes this widely used method in the following way:

> As evidence in support of a particular analysis, all sign-oriented linguists use the method of citing examples that illustrate a number of contextual interpretations of the signs being analyzed, and explaining how the various individual senses can be derived from the more abstract, general meanings that are being hypothesized, taken together with information from the linguistic and extralinguistic context [...].

Of course, this methodology has a major flaw: the analyst generally tries to illustrate his hypothesis by selecting those examples that fit the hypothesis best, and may therefore omit other examples that may possibly even contradict the hypothesis. So, it is uncertain if the examples are representative of the actual use and if the hypothesis is indeed the best means of explaining it. This is especially problematic from the point of view of other sciences. Diver (1995: 104) pertinently summarizes this shortcoming by stating that "The impression that the outsider gets, at the outset, is that this kind of judgment is hopelessly *ad hoc*."

Moreover, grammarians generally accept exceptions to the rules established by their own hypotheses as common rather than something that requires explanation. Even worse, they may even claim that: "The exception proves the rule!" Thus, expectations of scientific rigor in linguistics have been disappointingly low, as Diver (1980: 5) pointed out:

> If we no longer permit ourselves the luxury of rules that freely admit exceptions, procedures of verification must be raised well above the usual level to be found in linguistics.

It may be true that pure, genuine linguistic facts do not exist outside of theoretical frameworks (Reid 1995; Davis 2002, especially p. 75). However, are we justified to assume that the actual nature of the verification procedure also depends on the theoretical approach the linguist adopts?

Tobin (1990: 68–69) states this explicitly:

> Every linguistic analysis is the direct result of a specific set of theoretical and methodological assumptions which are directly related to how the linguist:
>
> 1. defines language;
> 2. defines a linguistic problem;
> 3. determines the source, kind and amount of data to be selected and analyzed;
> 4. chooses a methodology to select and analyze the data;
> 5. evaluates, compares and contrasts analyses in light of all of the above.

These five criteria basically serve to describe how and what the particular linguist views as the goal of linguistic research.

Davis (2002: 66) also asserts that this is the case. He states that:

> The appropriateness of any analytical technique must, of course, be evaluated in terms of the theoretical approach of which it is part. It is perfectly consistent with classical generative grammar to eschew quantitative data entirely (Chomsky 1957: 15–17). Grammaticality judgments on sentences as formal objects should not, in principle, depend at all on the discourse frequencies of those sentences, nor on their presence or absence in some corpus. Either the grammar generates a sentence's structure or it does not.

If we see what Chomsky has to say on this matter, we further note that he does mention analytical techniques, but rather discusses the construction of 'general laws' and 'hypothetical constructs' in relation to the observations:

> (Chomsky 2002: 49): Any scientific theory is based on a finite number of observations, and it seeks to relate the observed phenomena and to predict new phenomena by constructing general laws in terms of hypothetical constructs such as (in physics, for example) "mass" and "electron".

The question remains: What are the hypothetical constructs in linguistics? Are these hypothetical constructs the phonemes and the phrases, just as Chomsky suggests? Just like there are more constructs than "mass" and "electron" in physics, "phrase" and "phoneme" may very well not be the only ones in language, but traditional generative grammar based on a syntactic-oriented deep structure, appears

not to have even considered, for example, anything related to meaning as possible constructs, the only concern related to meaning being how to eliminate ambiguity in syntax (Chomsky 2002: 85–91).[2]

So, the question if there is more to grammatical theory than phonemes and phrases/sentences becomes relevant here. And if so, what kind of construct is important, which are the relevant observations necessary to determine them and how are these phenomena being observed and analyzed?

It is important to state that linguistics is an empirical science, also in the view of generative grammar (Chomsky 2002: 53). So, observations made in order to construct a possible grammar is done in a "given corpus of utterances" (for instance, Chomsky 2002: 51 *inter alia*). But, unlike observations made in other sciences, no conditions are required for the corpus of utterances, other than that they be grammatical in the observed language, or that other utterances are unacceptable i.e. unable to appear in such a corpus. The main criterion for judging grammaticality and/or acceptability is by means of the introspection of the analyst him/herself if the language in question is his or her native language, or judgments made by other mother tongue informants. And then – lo and behold – we are back at the beginning: again, we are dealing with a selection of examples *sans* independent data collection nor observation mechanisms. Further development of the formal generative theory (neither parameter theory (Chomsky 1981), nor the Minimalist Program (Chomsky 1998, 2008)) has fundamentally changed this methodology.

It is the intention of this volume to underline the importance of and need for objective and independent data collection in linguistic science, as in all empirical sciences. Contrary to Davis (2002: 66), we do not think that the underlying theory may excuse for or justify the absence of independent validation of the hypothesis.[3] Of course, this is not new, for Columbia School Linguistics has a well-established tradition of using independent data to test hypotheses. As Diver (1975: 14) already stated at an early stage of the development of the CS:

---

2. In the mid-seventies there was a movement led by James McCawley at the University of Chicago and Paul Postal to create a semantic-based deep structure in contrast to Chomsky's syntactic-based deep structure. The second editor of this volume was present at this syntactic-semantic deep structure debate held in a large auditorium with two entrances. The opposing sides sat together separated from each in each side of the auditorium and when the forum ended each side streamed out in the separate entrances/exits victoriously declaring that it had resoundedly vanquished the opposition.

3. There are, of course, other consequences that are related to the kind of hypotheses that are investigated and the constructs that are taken into consideration. However, we will leave this discussion for another occasion, concentrating here on the independent validation problem only.

> Although in the long run large numbers of apt examples undoubtedly provide the most subjectively satisfying confirmation of a hypothesis, it is understandable that we should have gradually moved to more objective techniques [...].

Diver (1979) in his seminal article "Phonology as Human Behavior" showed that the distribution of phonemes in English is not random, using quantitative techniques that objectively supported his hypotheses relating distinctive phonological features to other aspects of human behaviour. Diver's data are of a phonological nature, but he also hypothesized that meanings may be corroborated objectively:

> Contini-Morava (1995:23–24): The Columbia School differs from the other schools, however, in that it does not confine its validation to the individual sentence, but also considers the relation between grammatical meaning and the "macro-level" discourse. This attention to discourse has led to another major difference between the Columbia School and other sign-oriented schools, its extensive reliance on quantitative methods of validation. This usually involves comparing two or more signs with contrasting meanings with respect to a set of contextual features that are expected to show a statistical skewing in favor of one or the other meaning. Demonstrating that the predicted skewing exists is taken as support for the meaning hypothesis.

The premise presented here introduces a strategy that measures independently and indirectly a hypothesized meaning by controlling correlations between the investigated forms and contextual features. One should note that the indirectness is especially important and implies that it is not the phenomenon itself that is being observed and measured, but its significant co-occurrence with other, objectively observable contextual elements that are exclusively explained by means of the hypothesis that is being tested. However, as Diver (1975:15) already stated, there is a problem that needs further discussion:

> With indirect strategies [...] we must depend entirely on *objective procedures* to demonstrate that there are elements in the context that justify the use of the meaning in particular cases. [emphasis added]

In the following section we will discuss the problem of objectivity associated with observed data.

3. Characteristics of objectively observable testing

The purpose of quantitative testing in (any) linguistic analysis is to provide independent evidence for the proposed hypothesis. However, in order to be considered

independent evidence, objectivity is needed. The fact that objectivity in turn may be a problem was also observed twenty years later by Diver (1995: 104):

> One advantage of such a procedure would be that if it can be kept sufficiently objective, it might reduce some of the impression of *ad hoc* subjective judgment, although, to be sure, subjective judgment can never be entirely eliminated.

The question is, then, how these analytic tests can be kept 'sufficiently objective'. According to Diver, the only observable phenomena are the sound waves (1995: 48–49). Put in other words, in linguistics, the only objectively observable facts are the physical forms of language.

In Diver's view, all other observations, such as meaning, are the result of analysis and interpretation, and are therefore not objective. I think it can be sustained that as far as meaning is concerned, there is one phenomenon that represents them directly, i.e. the linguistic form of morphemes (and other overt linguistic signals), the minimal unit of meaning.[4] It is not meaning itself that can be observed, it is only its appearance, its form; but form can surely be observed objectively: all observations of linguistic forms can be repeated in a given corpus by different analysts using the same theoretical assumptions and methodolgical procedures, rendering the same results.

The same is not necessarily true for the testing method described above in which the co-occurrence of linguistic forms with other contextual elements is being observed. For example, in De Jonge 2000, an attempt was made to demonstrate the meaning of the simple past tense opposition between Spanish *indefinido* and *imperfecto*. The hypothesis was that *indefinido* indicates events in focus, and *imperfecto* describes supportive events. In one of the first tests, the distribution of both verb tenses was measured over actions that imply a movement (assumed to be more likely to be in focus) vs. the ones that do not (assumed to be more likely to be supportive, 2000: 241). In Table 1 we reproduce part of the relevant data:

Table 1. Distribution of *imperfecto* and *indefinido* over actions that imply a movement in the narrated scene and other actions, after De Jonge (2000: 241)

$X^2 = 96.2$ $p < .001$		Indefinido	Imperfecto
Totals of two stories by García Márquez	Movement	177/86%	30/14%
	Other	91/40%	138/60%

4. Of course, this line of reasoning only holds if we assume that there is a one-to-one relationship between form and meaning as explicitly advocated by Bolinger (1977).

As can be observed in Table 1, the expectations are confirmed by the data, since 86% of all *indefinido* forms can be characterized as a movement, and 60% of all *imperfecto* forms as a non-movement.

Although the absence or presence of movement associated with a verb might seem a fairly clear contextual feature that can indeed be measured, it is far from objective, for it depends basically on the interpretation of a particular context by the analyst. There is no objective way of telling that a verb like *to walk* implies a movement and *to see* not, other than relying on the generally accepted meanings. Should this case still be relatively easy to decide, things get more problematic when we think of verbs like *to laugh* or *to cry*. So, when another analyst wants to control this test by repeating it, s/he may not arrive at the same results as the original test.[5] But in order to test the hypothesis under study by means of the observation of contextual elements to justify the meaning, these contextual elements should be objectively observable as well.

Yet in spite of its subjectiveness, this kind of contextual testing is not useless, since it serves to further explain and elaborate the hypothesis, but it still cannot be taken as independent proof for the hypothesis. On the other hand, in De Jonge (2000), other tests were conducted where this objectivity is much clearer. For instance, when measuring the distribution of the above-mentioned verb tenses over main and subordinate clauses, it is not too complicated to define both clause types sufficiently in order to render the test repeatable by other researchers, with a guarantee of achieving the same results. The prediction is that, because events in focus are generally believed to occur more frequently in main clauses and supportive ones in subordinate clauses, *indefinido* should show a preference for main clauses and *imperfecto* for subordinate clauses. In Table 2 we give the results for this test, which are highly comparable to those of Table 1:

Table 2. Distribution of *imperfecto* and *indefinido* over over main clauses and subordinate clauses, after De Jonge (2000: 244)

$X^2 = 21.5$ $p < .001$		Indefinido	Imperfecto
Totals of two stories by García Márquez	Main	236/67%	118/33%
	subord.	32/39%	50/61%

The results of Table 2 clearly confirm the expectations: in main clauses, 67% *indefinido* forms are observed and 61% *imperfecto* in subordinate clauses. The measurements of Table 2 are of an objective nature, since the distinction between main

5. However, it is likely that the generally observed tendencies will be similar when this test is executed by different researchers.

and subordinate clauses, in principle, is a straightforward question and the results can be replicated by different researchers.[6] Other examples of such objective testing may be found, for example, in Huffman (1997:73, 75), where the distribution of French *lui* ('him, her, it') vs. a possessive adjective (*son/sa/ses* 'his, her, their') is measured with the absence vs. presence of other adjectives with body parts, as in *les larmes* ***lui*** *montèrent aux yeux* vs. *les larmes montèrent à* ***ses*** *propres yeux* 'tears rose (him) to (the)/his own eyes' (adapted from Huffman 1997:73). but also in various instances of Reid 1977 and Kirsner 1979 and many other CS studies. For instance, Reid (1977:321) measures the relation between *être* 'to be' vs. other verbs with *imparfait* vs. *passé simple*, two French verb forms with a similar opposition as the Spanish ones discussed above, in order to see if *être* occurs relatively more with *imparfait*. This is indeed observed and verified statistically, which is taken as evidence for the hypothesized LOW FOCUS meaning of the *imparfait*. Other objectively measured correlations are: negation/affirmation, (non) human subjects (1977:322) etc., all providing the expected results. Kirsner (1979:362), who discusses a difference in meaning between two Dutch demonstratives *deze* 'these' and *die* 'those', also takes (non-)humanness, singularity/plurality of the corresponding NP's, etc. as relevant factors to demonstrate the hypothesized difference in deixis.

Dreer (2007) is a more recent study where all kinds of objective correlations are measured. For instance, he measures correlations of French subjunctive vs. indicative mood in subordinate clauses with negation vs. affirmation in the main clause (2007:152), which shows an expected preference of subjunctive in subordinates for negations in main clauses, being the result of the compatibility of the invariant meaning ALTERNATIVE of the subjunctive and an alternative (affirmative) outcome, implied by negation.

In other cases, however, the measurements may not be as objective as the examples indicated above, but nevertheless are or seem to be perfectly controllable. For example, in Huffman (1997:66–8) the distribution of three pronoun types is measured over two verb groups. First, both a complete classification of the verbs into groups (rather than contexts) and a justification for this choice are given, making independent repetition of the test possible. However, there is no formal criterion in order to classify verbs into groups and although the justification in the cited study is crystal clear and the results of the test most convincing,

6. We are aware of the fact that grammars do not always agree on which conjunctions are coordinate and which subordinate. However, if the analyst describes carefully how s/he classifies the measured feature on a formal basis, the results must be the same when repeated by other analysts.

for real independent testing it would be preferable to have objectively observable characteristics in all testing procedures.[7]

The major problem with objective testing is that although the procedure is controllable, the interpretations of its results are not. So, the only thing that we can rely on is that the observation is objectively accurate, but the implications of what is observed will depend on the researcher:

> Diver (1995: 110): One danger in making predictions for quantitative procedures should be pointed out. What might be thought of as a perfectly "logical" prediction from the hypothesis may turn out not to be true at all. We can exercise our ingenuity to invent ways to take communicative advantage of a particular meaning, but the actual users of the language may have thought of other ways.

One would expect that the hypothesis that explains the predicted observation best will persevere until a better hypothesis appears (which ultimately depends on which criteria are chosen to determine which is the 'best' hypothesis). However, there is a way to increase the probability that a hypothesis is indeed a plausible explanation for the phenomenon under study. If a certain number of quantitative procedures is executed and all the individual observations may be explained by one and the same hypothesis, this obviously bespeaks its credibility.

The next question is, then, how many different tests and corresponding predictions are needed for the hypothesis to be acceptable. This depends on the nature of the predictions. For instance, in De Jonge 2000, another test was conducted in order to show the hypothesis stated above: it was predicted that there should be a correlation between animate subjects and the *indefinido* and inanimate subjects and the *imperfecto*, based on the idea that animates are more likely to be subjects of events in focus and inanimates are more likely to be the subjects of supportive events (De Jonge 2000: 245). Although the prediction is confirmed by the data and both this prediction and the one on the correlation between the verb tenses and main and subordinate clauses may be explained by the same hypothesis, the question remains whether both predictions are independent demonstrations of the same hypothesis. It can be argued that animate subjects are more prominent

7. Classifying verbs into type groups is most common in linguistic analysis, cf. for example the famous Vendler classification into verbs that express, respectively 'activity', 'accomplishment', 'achievement' and 'state' (Vendler 1957). Since there is no formal criterion to distinguish among these classes, it comes as no surprise that there is disagreement in the actual classification of individual verbs in different languages. Even a distinction between transitive and intransitive verbs cannot be done independently, and so it is perfectly logical, for example, to distinguish between verbs that have one complement and the ones that have two (or three), but see also García (1975: 84–91).

and dynamic than inanimates, and may therefore occur more frequently in main clauses and with verbs of motion than inanimates, so all the tests fundamentally are inter-dependent i.e. depend on one another and therefore cannot be considered as being independent and autonomous. Although this particular case does not seem to be too problematic, the idea that the interrelation of contextual factors is indeed an important element to be taken into account is shown by Mauder (2008), who convincingly demonstrates that most quantitative validation of the research done on Spanish oblique pronouns, and ultimately in linguistics in general, leaves much to be desired.

4. Conclusion: This volume

From all of the above, it may be clear that the purpose of this volume is to open a debate on the necessity of independent validation of hypotheses in linguistics, and how this validation should conducted, i.e. what forms it may have, and which observations count as objective and which do not. Each and every paper in this volume contains a section which aims at measuring certain linguistic phenomena in order to provide evidence for the presented hypotheses. Practically all papers reflect the theoretical approach of Columbia School Linguistics or have an approach most similar to it. The volume consists of two parts: a first part in which grammatical problems are treated, and a second part on Phonology as Human Behavior (PHB), the Columbia School version of phonology, which aims at 'discovering both the optimal characterization of the abstract sound unit, the phoneme, and an explanation for the nonrandom skewing of the sound units within language' (Tobin 1997: 31).

The first part opens with Dreer, who presents a textual analysis of the use of the French subjunctive and indicative. His paper shows how a postulated invariant meaning that motivates the distribution of each mood in French can be validated statistically. The tested invariant meaning for the subjunctive forms is ALTERNATIVE TO OCCURRENCE CONSIDERED, the used quantitative testing techniques are the bottom-to-top sign-to-text and the top-to-bottom text-to-sign approaches (cf. Tobin 1990, 1993, 1995), procedures that reveal the non-arbitrary relations of the meaning of the form and its distribution with certain elements in a given text (sign-to-text) and, *vice versa*, the message of a text as a determiner for the distribution in it of a given form (text-to-sign). The former approach predicts co-occurrence of subjunctive mood with certain characters; the latter relative predicts a preference for subjunctive mood in certain text fragments.

Even-Simkin & Tobin is the least quantitative paper of this volume. However, the observations made are strictly objective: it deals with the systematic Internal Vowel Alternation (IVA) in English nouns and verbs, which is generally