*Acta Linguistica Hungarica, Vol. 50 (3–4), pp. 405–434 (2003)*

# SYNTACTIC MICROVARIATION AND METHODOLOGY: PROBLEMS AND PERSPECTIVES*

JUDIT GERVAIN

**Abstract**

Variation in empirical data has been a perseverant problem for theoretical linguistics, especially syntax. Data inconsistencies among authors allegedly analyzing the same phenomenon are ubiquitous in the syntactic literature (e.g., literature on focus-raising in Hungarian; É. Kiss 1987 vs. Lipták 1998), and partly result from the highly informal methodology of data collection. However, even if adequate controls are used to exclude potential biases, variation might remain. The general practice in syntactic research has been to ignore these "microvariations"—mainly in the lack of any systematic empirical method to detect them. The present paper shows that this practice leads to serious theoretical problems and proposes a new empirical method, cluster analysis, to discover, explore and systematize these variations. It also illustrates how this richer empirical basis gives rise to a more fine-grained theoretical analysis.

## 1. Introduction: questions of dark nature

What counts as empirical data in linguistics? What happens if "investigators proposing different analyses of the same phenomenon disagree about the status of various crucial data" (Levine 2001)? Forgotten after the early struggles between generative grammar and behaviorism, these foundational issues have recently received new attention within the generative paradigm.

The present paper is an attempt to contribute to this new line of research by proposing a potential solution to the methodological problem raised by variation in raw empirical data, as formulated in the initial question by Levine (2001). It is suggested that this methodological difficulty stems from the intuitive nature of methodological practice in generative grammar, and can therefore be remedied by placing research methods on empirically more sophisticated, experimental grounds.

* I am indebted to István Kenesei, Gábor Zemplén and Leonie Cornips for useful comments and discussions about earlier versions of the paper.

The article first investigates the current methodology of data collection and analysis in syntactic research, starting out from an overview of the epistemological assumptions that underlie this practice, then going on to the discussion of its empirical inadequacies. Second, it is claimed that experimentation should be introduced into syntactic research, and a new experimental method is put forth to remedy the specific methodological difficulties induced by variation in linguistic data. Hungarian focus-raising constructions serve as an illustration to demonstrate the workings and results of the new method.

## 2. Empirical methods in syntactic research: blunt tools

Generative syntax has mainly been interested in the characterization of native speaker (syntactic) competence (Chomsky 1957; 1995; 2000). However, the linguistic competence of the human mind is not directly available for observation; researchers in linguistics are obliged to resort to behavioral, i.e., performance data, and thus make inferences about the underlying competence.

What counts, from an epistemological point of view, as relevant performance data and valid inference thereof has been debated ever since the emergence of generative grammar (most importantly e.g., Quine 1953; 1972; Chomsky 1961; 2000; Hill 1961; Ziff 1964; Stich 1971; 1972; Chomsky–Katz 1974; Labov 1975; Tsohatzidis 2002). However, this philosophical aspect of the problem, encompassing questions like the naturalization of linguistics and the underdetermination of theory (i.e., description of competence) by empirical data (i.e., performance), will not be of concern here (for some of these issues, see Zemplén and Gervain forthcoming). In this respect, I will assume that the generative enterprise is basically on the right track when entertaining the possibility of deriving valid and interesting conclusions about the language faculty of the mind from behavioral linguistic data.

Rather, the issue I am raising is of methodological nature and centers around the more practical query of how to collect performance data bearing on the abstract linguistic competence in the most relevant, informative and empirically adequate fashion. What I am suggesting is that the current intuitive and introspective practice, long expelled from other behavioral sciences (Pléh 2000), is highly problematic and needs to be replaced by more strictly controlled experimental methods.

As a first step in examining the current methodology, I will briefly outline the underlying epistemological assumptions, then continue by demonstrating the inadequacies of such a practice.

### 2.1. Why trust native speaker judgments?

The language faculty and its contents, universal grammar (UG) are regarded by Chomsky (1972; 1995; 2000) as a biologically endowed, innate and thus universal property of the human species. Principles (universals of language structure) and parameters (language-specific switches) contained in UG define the logically possible space for all human languages. Every actually attested natural language falls, by definition, within this possible language space.

In the life of a single individual, UG is in state $L_0$ at birth. At this stage, all principles and parameters are in their default settings. It is linguistic input during early childhood that sets the parameters to fit the external stimuli.[1] At the end of this procedure, the individual's grammar gets into state $L_L$, which is the particular grammar of language L that the child was exposed to. Whatever the final state $L_L$, it will always fall within UG. This claim has two consequences: first, that all human languages are relevant for the study of UG; second, that every individual's mental grammar is relevant for the study of UG. Hence the intuitive-introspective methodology—as Pléh (2000, 75)[2] puts it: "the incontrovertible intuition".

This train of thought justifies one aspect of current methods, namely the use of individual judgments. At the same time, it also marks its limits. Since every individual $L_L$ is only a subset of the possibilities available in UG, analyses of individual $L_L$s have implications for the study of UG, but are not necessarily informative about other $L_L$s. For the analysis of a particular $L_L$ to be relevant with respect to another $L_L$, it has to be shown that the two $L_L$s are identical at least in the relevant aspects. In the case of speakers of the same language, overlap may be assumed as long as their judgments about a construction coincide. This is, of course,

[1] This is a rather simplified view of language acquisition. For more detailed descriptions of the parameter-setting theory, see Fodor (2001), Guasti (2002). For criticism of this approach and alternative hypotheses, see e.g., Elman (1993), Elman et al. (1996) and Gómez–Gerken (2000).

[2] See Pléh (2000, 75) for a very interesting analogy between Descartes and Chomsky with respect to mentalism and introspection.

not necessarily so, since data always underdetermine the models built on them (Duhem 1954; Quine 1975). However, when judgments diverge, or when no evidence is available to the contrary, it is necessarily not so, and the conclusion is that some of the parameters in the two individuals' $L_L$s are set differently. And while both are relevant for the study of UG, since they attest different possible settings of a parameter, they are not comparable to each other in any direct way.

Although the above discussion provides but a highly simplified account of some of the generative principles, it is apparent that reliance on native speaker intuition is a necessary consequences of these principles, while the informative, uncontrolled nature of data collection is not. Intuitive judgments can be obtained from informants through more objective, quantified and experimentally controlled procedures. As Chomsky (1957, 13–4), among many others, claimed:

> "The grammar of L[anguage] will thus be a device that generates all of the grammatical sequences of L and none of the ungrammatical ones. One way to test the adequacy of a grammar proposed for L is to determine whether or not the sequences that it generates are actually grammatical, i.e., acceptable by native speakers, etc. We can take certain steps towards providing a behavioral criterion for grammaticalness so that this test of adequacy can be carried out."

These early proposals notwithstanding, experimentation has never become the actual practice in generative syntax, which, with the evolution of syntactic theory over time and the increasing sophistication of data, has led to considerable methodological difficulties.

### 2.2. Empirical problems

There are two aspects of language structure that traditional methodology typically fails to grasp: gradedness and variation.

It has long been noted (for an overview, see Keller 2000) that strictly dichotomic distinctions of grammaticality (e.g., grammatical vs. ungrammatical, or acceptable vs. unacceptable) are unable to cover the full scale of intuitive judgments speakers formulate. Rather, at least for certain constructions, grammaticality is very often graded or even continuous. Although degrees of grammaticality are sometimes made use of in the literature, this practice is neither systematic nor uniform. Works do not generally define or specify how they measure gradedness. Moreover, the

number and the relative distance of the degrees employed vary from one publication to the other, rendering all principled comparisons impossible.

A second problem derives from the fact that judgments on certain, usually highly complex, syntactic constructions are often conflicting. Variation in the data comes from two main sources: methodologically uncontrolled data collection procedures and real speaker variation (Schütze 1996; Cowart 1997).

As shown above, the mapping from grammatical competence to actual linguistic behavior/performance is non-trivial and highly complex, involving several cognitive processes of different kinds (most of them non-syntactic). Evidence from psycholinguistics shows (Schütze 1996) that these factors bias judgments. World knowledge, pragmatic context, word length and frequency, length of utterance, just to mention a few of the potential biases, may all introduce random, and for syntactic purposes irrelevant, variability in linguistic data.

But even these factors filtered out, it may occur that a non-negligible amount of variation remains in the data set, reflecting existing syntactic heterogeneity, i.e., differences of parameter setting among speakers. Importantly, however, this variation is fundamentally different in nature from the above-mentioned random incongruity of the data. Unlike the latter, which is undesirable and should be controlled for, the former constitutes an important source of information with respect to the possible settings of (typically small, language-specific) parameters. These syntactic microvariations (van Oostendorp 2002) are therefore of primary relevance for syntactic theory.

Note that an informal, experimentally uncontrolled methodology fails to distinguish between irrelevant and relevant types of variation, just as it is unable to systematize the latter kind, even if the former is somehow filtered out.

This problem is well illustrated by the literature on focus-raising constructions in Hungarian. In these structures (É. Kiss 1987), an embedded focus constituent surfaces in the focus position of the matrix clause containing a bridge verb (*mond* 'say', *akar* 'want', *gondol* 'think', *szeretne* 'would like' etc.), as illustrated in (1b).[3]

(1) (a) Azt mondtad, (hogy) JÁNOS jön.
expl-acc said-2sg that János come-3sg
'You said that John was coming.'

[3] Small capitals indicate focus.

(b) JÁNOST$_i$ mondtad, hogy e$_i$ jön.
János-acc said-2sg that come-3sg

In her analysis of focus-raising, É. Kiss (1987, 141) claims that both (2a), with Accusative case on the raised constituent, and (2b), with Nominative case, are grammatical. In contrast, on the basis of a small-scale survey, Lipták (1998) argues that only (2a) is acceptable.

(2) (a) János KÉT **DOLGOT** szeretne, ha sikerülne.
János two thing-**acc** would-like-3sg if succeeded-3sg
'As for John, it is two things that he would like if they succeeded.'

(b) János KÉT **DOLOG** szeretné, ha sikerülne.
János two thing-**nom** would-like-3sg if succeeded-3sg

But what is one to make of Lipták's (1998) empirical claims? É. Kiss is a native speaker of Hungarian, thus, by definition, an authentic source of intuition. But so are Lipták and her informants. Since the two papers do not provide any information about the data collection procedure, it is impossible to decide whether the inconsistency between the two data sets results from random or real variation. Nevertheless, theoretical analysis of focus-raising cannot proceed without the clarification of this issue.

It can only be hypothesized that, the authors and their informants being trained linguists, the probability of pragmatic, lexical or other biases influencing their judgments is low. Thus, the present case is most probably an instance of real syntactic microvariation.

What follows from this with respect to the analyses the two authors propose? Does the second really falsify the first? Can they be compared at all? What are they the analyses of, in the first place? The individual grammars of the two linguists?

From a more general perspective, the informal nature of methodology undermines the reproducibility of the data, as well as the falsifiability of theories (Hoji 2002), both unwelcome in empirical sciences.

## 3. New methods in syntax

Recently, several authors have recognized the inherent flaws of informal data collection practices (Schütze 1996; Cowart 1997; McDaniel–Cowart 1999; Keller 2000; Sorace–Keller forthcoming). These works all capitalize on the importance of providing well-established empirical foundations for

syntactic theory by introducing experimental controls into its methodology. By eliminating random variation induced by irrelevant biasing factors, this move will guarantee clearer data. This, however, is only a first step. Specific methods are needed to address the questions of gradedness and speaker variation. While proposals to solve the first problem exist in the literature (Keller 2000; Sorace–Keller forthcoming), no solution has yet been suggested to remedy the second.

In the following section, I will first briefly describe some of the experimental tools that have been put forth. Secondly, I will delineate the principles of a new method that can be used to collect and analyze heterogeneous linguistic data in which variance results from syntactic microvariation.

### 3.1. Experimentation

As noted before, syntactic research has access to a speaker's mental linguistic competence only through his/her performance, that is, behavior. The relationship between an underlying competence and the corresponding surface behavior is rarely, if ever, one-to-one. Language is no exception. Linguistic competence, the core of the language faculty, communicates with the external world through peripheral performance systems like articulation, memory, the conceptual system, etc. (most recently Chomsky 2000; 2002). Consequently, when inferences are made about the underlying competence on the basis of observable surface behavior, it is best to exclude all possible sources of contamination. Since syntactic methodology lacks experiments and experimental controls, research is left unprotected against noise in the data induced by systems other than the core linguistic (grammatical) competence.

Recently, Schütze (1996), Cowart (1997), and McDaniel – Cowart (1999) have argued for the necessity of "experimental syntax", and have adopted the methodological principles of psycholinguistics and experimental psychology. I will discuss but a few of these below.

One of the central issues concerns habituation of informants to the specific structure of sample sentences, resulting in more indulgent judgments. This can be prevented by random ordering of sentences in a questionnaire, or by the use of 'filler sentences', avoiding too frequent repetition of lexical items and construction types etc.

Secondly, pragmatic, semantic and lexical factors are also known to influence judgments, and should thus be controlled for. For instance, all sample sentences should be matched for the frequency and length of the words contained in them. Differences in style, register or truth value may also bias speakers' judgments.

A third cause for concern is whether to include linguistically trained informants in the experiments. The general argument for using linguists as informants is that having more refined intuitions, they come closer to the concept of the 'idealized native speaker' than naïve informants. In other words, linguists tend to be less prone to bias. On the other hand, it is a standard requirement in experimental psychology that the subjects be naïve with respect to the experimental task, which linguists are not. To the present date, no conclusive evidence has been produced for either argument. However, the few psycholinguistic experiments that deal with this issue (for an overview, see Schütze 1996) have shown no significant difference between trained and naïve informants. (This conclusion will also be confirmed by the results of the present case study, see later.)

Fourth, the nature of the task, i.e., the type of judgment asked from the subjects (e.g., binary or scaled) determines not only what other works in the literature the results will be comparable with, but also the kind of statistical analyses that can be performed on the raw data (e.g., a binary grammaticality decision, producing categorical data, does not allow for certain statistical analyses). Therefore, due attention has to be paid to choosing the most suitable judgment paradigm and scale.

The aforementioned are but some of the factors that contribute to the noise in the data collection procedure. These and similar issues have to be carefully considered and controlled for in order to clarify the empirical bases of syntactic research.

### 3.2. Degrees of grammaticality

Even well controlled, non-biased data can be highly misleading, if they do not reflect important properties of the phenomenon investigated. Grammaticality rating scales constitute a case in point. As Cowart (1997) and Keller (2000) have argued, grammaticality may not be a discrete property, but a continuum, at least for certain constructions. As a consequence, the traditional binary grammaticality judgments, or even some of the graded scales, may sometimes conceal syntactically meaningful distinctions among constructions.

To overcome these difficulties, Keller (2000) and Sorace–Keller (forthcoming) adopted the method of magnitude estimation, routinely used in psychophysiology to measure perception, where stimuli are typically continuous. The major difference between traditional grammaticality judgments and the ones obtained through magnitude estimation is that the former are absolute, i.e., sentences are evaluated independently of each other, while the latter is relative; the grammaticality of each sample sentence is estimated in relation to a baseline sentence the grammaticality of which is used as the basic unit. Thus, the subject's task is first to assign a unit of grammaticality to the baseline sentence, presented at the beginning of the experimental session. The unit of grammaticality is visualized and quantified as a physical property of a geometrical object, for instance the length of a line. Then, subjects are asked to evaluate the grammaticality of sample sentences as proportional to that of the baseline sentence. They can do this by setting the magnitude of the physical property to a value that is a multiple or a fraction of the baseline. For instance, if a sentence is felt to be twice as grammatical as the baseline sentence, the subject can signal this by assigning the former a line twice as long as to the latter.

Keller (2000) emphasizes that the method, when applied in linguistics, is as reliable as in the original psychophysiological experiments, and its results nicely correlate with judgments obtained by other methods.

A major asset of this procedure is that it provides information about the grammaticality of test sentences with respect to each other, allowing insight into the hierarchy of syntactic constraints, rules etc.—data that can prove useful not only for an Optimality Theoretical approach.

### 3.3. Microvariation

Inconsistency can be introduced into the data not only by noise or graded grammaticality, but also by speaker variation. The empirical debate between É. Kiss (1987) and Lipták (1998) is most probably a case in point, since it is not unreasonable to suppose that, being professional linguists, these authors are not influenced by biasing factors. Rather, their judgments reflect existing syntactic variation.

There are two questions that need to be addressed at this point:

(i) What is the theoretical explanation for syntactic microvariation?

(ii) How can it be detected empirically?

They will be addressed in the remainder of this section.

### 3.3.1. Microvariation from a theoretical perspective

As Labov (1975, 16) put it, "the logic of linguistic inquiry has been to assume consensus rather than test it". But positing consensus is, of course, unfounded. The assumption that speakers of language L form a completely homogeneous community is empirically highly unrealistic and also theoretically unexpected in a model of the 'principles and parameters' type. Just as separate languages differ with respect to the setting of certain parameters, it is natural to believe that individual grammars may also do so. But importantly, in such a model, the differences are not just random; rather they are systematic along certain parameters. Therefore, the detection of syntactic variants, i.e., 'syntactic microvariation' contributes to a more fine-grained exploration of the set of parameters, thus a better understanding of UG.

The study of variation is relatively new in syntax, but not in the broader context of linguistic theory. The notion plays a key role in sociolinguistics. Therefore, it is worth making a short detour here to compare the two approaches. At first sight, they seem to have very little in common. Generative linguistics takes an internalist and individualist stance on language, conceiving of syntax as an autonomous, self-contained system. Sociolinguistics, on the other hand, aims at relating language structure, especially patterns of variation in phonology and morphology to external—social, geographical, historical etc.—factors. Moreover, syntax, unlike phonological and lexical phenomena, has traditionally been neglected in this discipline, as some of its inherent properties (high level of variation, low token frequency of construction types etc.) disfavor the reliable detection of variation patterns and their explanation in terms of external causes (Cornips–Corrigan forthcoming).

Nevertheless, there have recently been attempts in both fields to study syntactic variation in more principled ways. In sociolinguistics, new ideas have been put forth for the conceptualization of syntactic variation, partly adopting the principles and parameters model of generative grammar. Under this view (e.g., Cornips 1998), syntactic variants only relate to social factors indirectly; thus, at a linguistic level, they are considered to be differences in parameter setting, which, on a social level, are

attributable to external variation. Simultaneously, syntactic microvariation has been raising increasing attention in generative grammar (van Oostendorp 2002).

At the intersection of these new tendencies, there seems to emerge a new framework (Cornips–Corrigan forthcoming) that is able to capture both the internal and the external aspects of syntactic variation. This two-level framework first detects and describes structural variation and identifies the underlying parameters, then it makes an attempt to link (some of) the different settings of the parameters to external, social factors.

### 3.3.2. Detecting variation: clustering

Traditional parametric statistics is useful in analyzing raw experimental data when there is an a priori hypothesis to be tested, e.g., when the researcher has an assumption about the existence of certain variants and wants to verify this. However, without such expectations, classical statistics is of little help. When the aim is to detect systematic patterns (here, syntactic microvariation) in a data set, data mining techniques are used instead. Cluster analysis, the aim of which is to establish categories among the objects observed in an experiment, is the classificatory method that will be introduced and employed in the present paper.

Cluster analysis (Lance–Williams 1967; Everitt 1981; van Ooyen 2001) is a collection of 'heuristic' methods for the categorization of objects according to some similarity measure along one or, typically, several characteristics (variables). It is widely used, for instance, in microbiology to establish different strains of bacteria, or in biological taxonomy, to set up species, families, genera etc. of organisms. Cluster analysis has different types and techniques depending on the similarity measure and the classificatory criterion being used, but the underlying logic is the same—the stepwise, reiterative grouping of the two most similar objects or already formed clusters. Thus, agglomerative clustering starts out with as many clusters as there are objects, and by repeatedly putting together those two items (two objects, an object and a cluster, or two clusters) that are the most similar with respect to the measured characteristics, it gradually decreases the number of clusters until all objects eventually belong to one big category. (Divisive clustering proceeds in the other direction.) The result of the categorization is plotted on a dendrogram (Figure 1), from which the classes can be read off (here, for example, objects 1–3 form one cluster and objects 4–5 another).

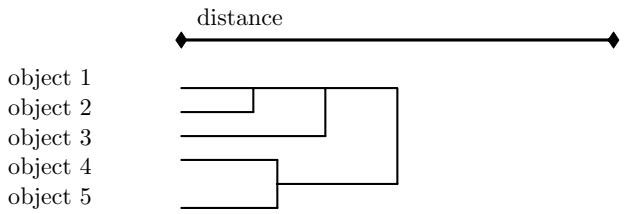


*Fig. 1*
A dendogram

(Dis)similarity is measured as a kind of 'distance' in the character space, an $n$-dimensional space defined by the $n$ number of properties observed in the experiment. For instance, if the height and body weight of individuals are measured, these two properties will define a two-dimensional space (the familiar Cartesian co-ordinate system), in which for each individual, height is represented as a value on one of the axes, body weight as a value on the other (Figure 2). The distance of two points (individuals, $I_1$ and $I_2$) in such an $n$-dimensional space can be calculated in a number of different ways, partly depending on the nature of the variables (binary, interval, count, etc.).

Without attempting to give an exhaustive list of these measures, I will simply introduce the most frequently used ones. First, the measure called Euclidean distance is determined in the narrowest sense of the word, i.e., it is the linear distance (e.g., measurable with a ruler) between two points. In our two-dimensional example of heights and body weights, it is the length of the line drawn between two points in the Cartesian co-ordinate system. This distance, the hypotenuse (c) of the right-angled triangle the two points define, can be calculated by the Pythagoras theorem (see Figure 2). A second possibility, known as the city block or Manhattan distance, is to add the two other sides of the triangle (a+b), which are adjacent to the right angle. Besides these two well-established measures, many others are reported in the literature (e.g., van Ooyen 2001).

Once the distance between the original objects is determined, the two closest, i.e., most similar can be fused to form a cluster. However, clusters are not 'natural objects' in the observed set. In their own rights, they have no values for the observed properties, thus no distances can be calculated. The values and distances have to be somehow obtained through a clustering algorithm from the original objects that are included in the cluster. Again, there are several different ways, and only some of

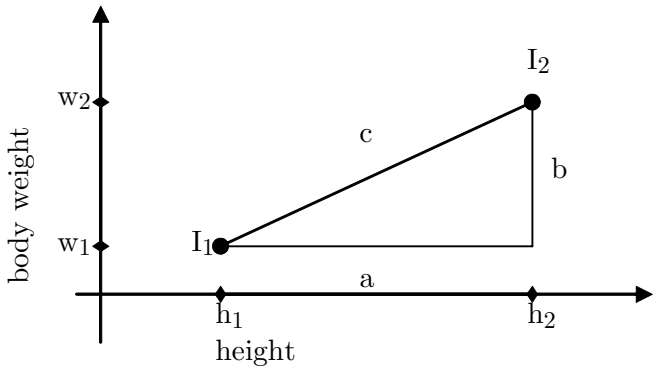


*Fig. 2*
Calculating Euclidean and city block distances
in a two-dimensional character space

the most popular ones will be enumerated here. The Unweighted Pair-Groups Method Average (UPGMA) defines the distance between a cluster, on the one hand, and an object or another cluster, on the other, as the average of the distances between each point in the cluster and the object or each point of the other cluster. The Furthest Neighbor method determines the distance between two clusters or a cluster and an object as the greatest distance between a member of the first and the second cluster (e.g., if cluster 1 = {A, B}, cluster 2 = {C, D}, distances: $d_{AC} = 2$, $d_{AD} = 6$, $d_{BC} = 5$, $d_{BD} = 10$, then this method will yield $d = 10$ as the distance between clusters 1 and 2), or the greatest distance between the object and a member of the cluster. Since distances are maximized, this method tends to produce well-defined, sharply distinguished clusters. The Nearest Neighbor method, on the other hand, chooses the smallest distance (using the previous example, the distance between the two clusters would be $d = 2$), and produces rather elongated clusters standing closely together, a reason for which the method is also known as chaining. A fourth technique, Ward's method fuses those items into a cluster the fusion of which augments within-cluster variance the least. This procedure also results in strongly demarcated groups, but has the disadvantage of lumping together the outliers, i.e., the aberrant points, which have nothing in common except that they are different from everything else.

A remark is in order here about the use and interpretation of cluster analysis. As mentioned at the beginning of this section, cluster analysis is a heuristic method. In other words, the resulting classification depends on a series of more or less subjective choices made by the user. Resorting to cluster analysis is the first of them. Any two objects can be likened according to some arbitrary criterion. Similarly, in any collection of objects, some classification can always be set up—the question is whether it 'makes any sense' or not, that is, if it corresponds to any 'natural' pattern in the set. To put it differently, cluster analysis cannot decide if there is regularity in the data set, rather it simply assumes it and tells us what it is like. Thus the burden of the decision is on the user. In syntax, the existence of some pattern in the varieties is 'guaranteed' by the principles and parameters framework—performing cluster analysis is therefore justified. (Note the strong theory-laddenness here.)

Secondly, as demonstrated in the short summary of the different types of analysis, the choice of the similarity measure and the clustering algorithm affects the outcome. Therefore, these decisions have to be adapted to and justified by the special needs and properties of the analysis to be carried out.

Furthermore, the interpretation of the final dendrogram is ambiguous. The tree can be split up into groupings in many different ways. Once again, it is up to the user to establish the final categories in such a fashion as to find a classification that fits the raw data in a meaningful way. Going back to the example in Figure 1, instead of the explanation that is given there (cluster 1 = {1,2,3}, cluster 2 = {4,5}), it could be argued that only objects 1 and 2 constitute cluster 1, and object 3 is only loosely related, thus an outlier. Although there is no unique solution, the distance measure scale might be suggestive. The greater the distance until the junction point, the more dissimilar the fused clusters or objects are.

In this section, three aspects of data collection and analysis have been discussed that contribute to the clarification of empirical issues. Experimental settings and controls help to filter out noise, while magnitude estimation and cluster analysis provide ways to exploit the data in terms of degrees of grammaticality and speaker variation. Recall that the existence of speaker variation is not contrary to the spirit of generativism; rather the opposite. Cluster analysis as a means to reveal syntactic microvariation can only work in a model of language that assumes variation to be systematic, like the principles and parameters framework. If suc-

cessful, this method should be able to account for such debates as the É. Kiss–Lipták polemic about focus-raising.

## 4. Focus-raising in Hungarian: a case study

The use and advantages of the new method will be illustrated below through the case of focus-raising. The example was chosen because there is every reason to believe that the empirical controversy is the reflection of true syntactic variation, not just experimental noise. This constitutes a good testing ground for the clustering method to the extent that its ability to accommodate both É. Kiss's and Lipták's position clearly shows its power in clarifying empirical issues.

### 4.1. Focus-raising: the empirical issues

Focus-raising, as defined above, is a construction in which the focused constituent of an embedded clause surfaces in the focus position of the matrix clause containing a bridge verb. Several properties of this structure are unequivocal. However, as cited earlier, the morphological case of the raised focused element is controversial. É. Kiss (1987) holds that when raised, a focused embedded subject can optionally keep its Nominative case or pick up Accusative from the matrix verb (see (2) repeated here as (3) for convenience).

(3) (a) János KÉT **DOLGOT** szeretne, ha sikerülne.
János two thing-**acc** would-like-3sg if succeeded-3sg
'As for John, it is two things that he would like if they succeeded.'

(b) János KÉT **DOLOG** szeretné, ha sikerülne.
János two thing-**nom** would-like-3sg if succeeded-3sg

Lipták (1998) has challenged this generalization on the basis of an informal small-scale survey (n = 12), arguing that only Accusative case is acceptable, i.e., (3a), while Nominative is ruled out (3b).

There is another property of subject focus-raising, not reported in the literature before, which appears to be subject to a substantial amount of speaker variation. This novel aspect, shown in (5), is best described as optional anti-agreement between the embedded verb and the raised subject, if it is of the type [$\text{NumP}_{\text{pl}}$ + $\text{NP}_{\text{sg}}$] (4). This phenomenon hinges

on a peculiarity of Hungarian DP morphology, namely that nouns preceded by plural quantifiers (*sok* 'many', *néhány* 'some' etc.) or numerals remain morphologically singular.

(4) Két fiú jön/*jönnek.
two boy-sg come-3sg/come-3pl
'Two boys are coming.'

(5) Két fiút mondtál, hogy jön/jönnek.
two boy-acc said-2sg that come-3sg/come-3pl
'You said that two boys were coming.'

Since variation and inconsistency was considerable in the data, an experiment was designed with the following three aims in mind:

(i) to settle the empirical controversy concerning the case of the raised subject;
(ii) to explore the new facts about the agreement between the focused subject and the embedded verb;

and

(iii) to determine whether there is any systematic relationship between the case facts and the agreement facts.

### 4.2. The experiment

#### 4.2.1. Material

In addition to case and agreement, four other factors were introduced in the construction of the test material to facilitate further studies. These additional factors will, however, not be evaluated or discussed here (but see Gervain 2002). The sample sentences were thus constructed along six factors:

(i) the nature of the raised operator (quantificational/non-quantificational)
(ii) the case of the raised operator (Nominative/Accusative)
(iii) the number agreement of the embedded verb (singular/plural)
(iv) the reading of the embedded verb (agentive/non-agentive, distributive/collective)
(v) island effects (yes/no)
(vi) contextual reference (possible/impossible).

Since some variables exclude each other or are mutually irrelevant, not all the possible combinations were surveyed. The used combinations yielded 53 test sentences.

Order of presentation effects were excluded by randomizing the sentences. Habituation and repetition were counterbalanced by the insertion of grammatical and ungrammatical filler sentences, 12 altogether. These were never included in the data used for statistical analysis. The complete questionnaire can be found in Appendix A.

#### 4.2.2. Subjects

Twenty-three informants participated in the survey. Contrary to Schütze's (1996) and Cowart's (1997) claims, subjects with linguistic background were also included (n = 10), since the psycholinguistic literature fails to show any significant difference between linguists and non-linguists. Indeed, an analysis of variance revealed no difference between the responses of linguist and non-linguist informants in the present experiment ($F(1, 21) = .846$, ns).

#### 4.2.3. Prodecure

A paper-and-pencil questionnaire was administered to informants through electronic mail or physically. Informants with and without linguistic training received different instructions. The latter were given detailed explanation and illustration of the notion of grammaticality. When asked for, further explanation was provided.

Subjects were asked to evaluate the sample sentences on a five-grade scale, ranging from totally unacceptable through three intermediate levels to fully acceptable. The five-grade scale was adopted, on the one hand, in order to allow comparison, as it is one of the most commonly used ratings in the literature; on the other hand, because there was no reason to believe that there would be important within-subject variations in the degree of grammaticality of the sentences (variation was only expected to be between-subject). This rating was treated as an interval measurement scale, i.e., one in which the points are at equal distances from each other. (Note that it is not obvious that such a rating should correspond to a real interval scale, rather than just a simple ordinal one. The extreme values, for instance, could be perceived by the subjects as being further away from the intermediate values than the latter ones from each other. Nevertheless, it can be shown (Schütze 1996; Cowart 1997) that working with interval scales does not distort the results.)

#### 4.2.4. Statistical analyses

As a first step, it has to be shown that there indeed is significant variation in the judgments. Therefore, data has to be tested for the homogeneity of variance to ascertain that speaker variation is indeed significant, at least for some of the test sentences. If this turns out to be the case, informants will be classified by cluster analysis into syntactic 'dialects' on the basis of systematic patterns in their responses. Then, within each dialect, analysis of variance tests will be performed to determine the role of the previously defined variables.

Several remarks are in order here. First, the availability of the last type of analysis crucially depends upon the assumption that the judgment scale was conceived of by informants as an equally paced interval scale, rather than an ordinal scale, because certain statistical analyses can only be performed on the former.[4] Second, I am not following Cowart (1997) in his categorical refusal of all types of individual data. For an appropriate characterization of the emerging clusters, raw data contained in the individual protocols are indispensable. Third, the decisions about the clustering measure and algorithm have to be made explicit at this point. Euclidean distance was used as the measure of (dis)similarity, as it is well suited for interval data. More importantly, Ward's method was chosen as the clustering algorithm, since its well-defined clusters and its within-cluster variance reducing property nicely mirror the concept of linguistic microvariants, which are clearly distinct from each other with respect to certain parameters, but are homogeneous inside. (In other words, linguistic variation is not gradual, but punctuated.)

### 4.3. Results: revealing two strategies of focus-raising

The test of variance (see Appendix B) shows that for 33 out of the 53 test sentences, variance in judgments was statistically significant ($p < .05$). In other words, there is variation in the data, and the principles and parameters framework tells us that it must be systematic, therefore it is meaningful to use cluster analysis to establish the syntactic 'dialects'.

A cluster analysis using Ward's method with Euclidean distances (and non-standardized data) was performed with the following result

[4] This assumption is reflected by the fact that the original scale (*, ???, ??, ?, unmarked) was recoded as a five-grade scale from −2 to 2 for the purposes of the statistical analyses.

(Figure 3). Three clusters were established with two outliers (subjects 4 and 11).[5] A major division lies between group 3 and groups 1–2. This main cut corresponds to the rejection or acceptance of focus-raising, respectively. Put differently, the grammars of some speakers of Hungarian (group 3) do not contain the option of focus-raising at all. The assumption was confirmed by an analysis of variance (ANOVA), which showed a highly significant ($F(1, 21) = 42,922$, $p < .0001$) difference between the mean judgment scores of groups 1–2, on the one hand, and group 3, on the other, for all sentence types instantiating operator-raising (all test sentences except *30, 35, 38, 41*).[6]

Informants who accept focus-raising can be further divided into two larger groups, 1 and 2 respectively. To establish whether these groups are indeed meaningfully distinct, the debated case assignment factor and the new (anti-)agreement feature were tested in a repeated measures ANOVA, together with group membership as a two-level (group 1/group 2) between-subject factor. A highly significant main effect of case ($F(1, 12) = 246.788$, $p < .0001$) was found (Figure 4).[7] Agreement and group membership showed no main effect ($F(1, 12) = 2.667$, ns; $F(1, 12) = .251$, ns, respectively). Significant interactions were obtained for case × group membership ($F(1, 12) = 24.473$, $p < .001$), for agreement × group membership ($F(1, 12) = 6.461$, $p < .05$) and for case × agreement ($F(1, 12) = 6.067$, $p < .05$). No triple interaction was attested ($F(1, 12) = 1.105$, ns).

[5] Informants 4 and 11 were excluded from both of the groups and from any further analysis. On closer investigation, their protocols appear incoherent. This may be due to experimental error, e.g. these speakers, both of them non-linguists, might have had difficulties understanding the notions of grammaticality or failed to reproduce the sentences with the appropriate focus intonation for themselves. Note, however, that the exclusion of these informants happens on purely methodological grounds. Importantly, this is quite different from the principled, theoretical refusal of individual protocols as suggested by Cowart (1997).

[6] Italicized sentence numbers refer to the sample sentences as they appear in Appendix A. Note that in the original questionnaire, sentences were not numbered. They appear here only for ease of reference.

[7] The judgment ratings reported in examples (5) and (6), and in Figure 4, might at first sight suggest that all of these constructions are marginal, and the difference between them lies only in the extent to which they are ungrammatical. Nevertheless, this is not the case. Bear in mind that values given in Figure 1 are means for all the six factors confounded, i.e., the focus-raising sentence types combined. There are types which are less grammatical than the values given here, while others are more grammatical. In fact, some of them are fully acceptable.

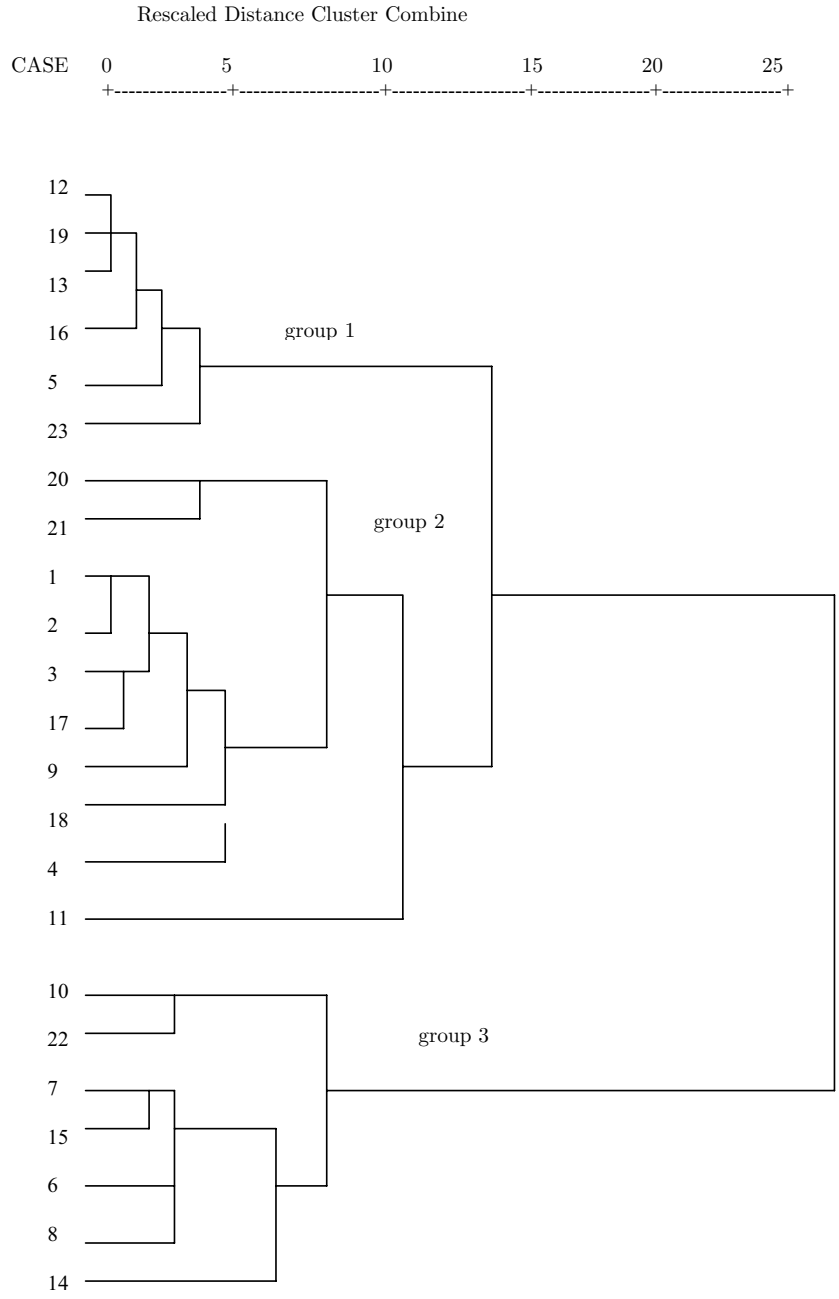


*Fig. 3*
Hierarchical cluster analysis dendrogram using Ward's Method

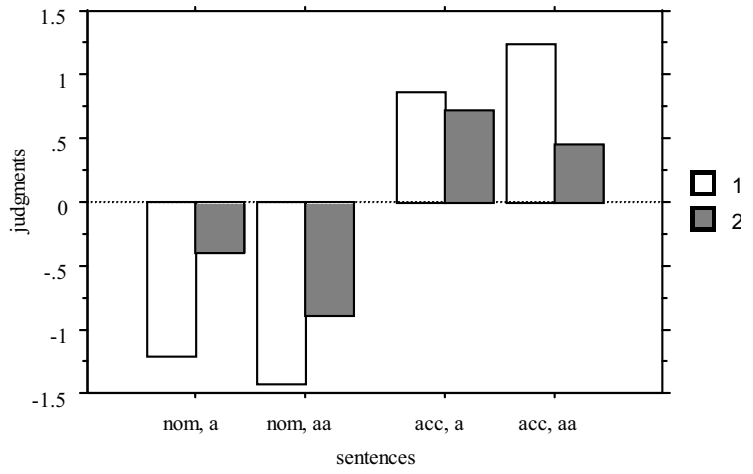


*Fig. 4*

Effects of case, agreement and group membership for groups 1 and 2.
(nom: Nominative, acc: Accusative, a: agreement, aa: anti-agreement)

These results show that what differentiates between the two groups is a complex pattern of case and agreement co-occurrences.[8] Group 1 speakers refuse Nominative case altogether, but they tolerate its co-occurrence with anti-agreement even less (6a,b). On the other hand, Accusative case on the raised operator is quite acceptable for them both with agreement and anti-agreement, and interestingly, it is actually preferable with anti-agreement (6c,d).

(6) (a) ???AZ ÖSSZES LÁNY mondtad, hogy jön.
the all girl-sg-nom said-2sg that come-3sg
'You said that all the girls were coming.'

(b) *AZ ÖSSZES LÁNY mondtad, hogy jönnek.
the all girl-sg-nom said-2sg that come-3pl

[8] The absence of main effect of group membership can be accounted for by the fact that the case group × membership and agreement × group membership interactions "explain away" the difference. In other words, the two groups do not behave radically differently with respect to focus-raising in general. Rather, they show different patterns with respect to the tested features of focus-raising. That the distinction between the two groups is nevertheless real is demonstrated by the fact that (i) group membership interacts with both variables and (ii) that these interactions are highly significant.

(c) ?AZ ÖSSZES LÁNYT mondtad, hogy jön.
the all girl-sg-acc said-2sg that come-3sg

(d) AZ ÖSSZES LÁNYT mondtad, hogy jönnek.
the all girl-sg-acc said-2sg that come-3pl

Group 2 informants do not consider Nominative case on the operator completely grammatical either; however, they are significantly more in favor of it than members of the other group, especially if it occurs with agreement (7a,b). These speakers accept Accusative case to a significantly lesser extent. Moreover, anti-agreement does not amend, but worsen acceptability (7c,d) for these speakers.

(7) (a) ??AZ ÖSSZES LÁNY mondtad, hogy jön.
the all girl-sg-nom said-2sg that come-3sg
'You said that all the girls were coming.'

(b) ???AZ ÖSSZES LÁNY mondtad, hogy jönnek.
the all girl-sg-nom said-2sg that come-3pl

(c) ?AZ ÖSSZES LÁNYT mondtad, hogy jön.
the all girl-sg-acc said-2sg that come-3sg

(d) ??AZ ÖSSZES LÁNYT mondtad, hogy jönnek.
the all girl-sg-acc said-2sg that come-3pl

On the whole, then, two different ways of constructing focus-raising have been detected: group 1 speakers refuse Nominative case (cf. Lipták 1998), especially with anti-agreement, while group 2 speakers do not accept anti-agreement, but tolerate the optionality of case (cf. É. Kiss 1987).

### 4.4. Discussion: deriving the two strategies

No detailed syntactic analysis will be undertaken here, since the orientation of the present paper is essentially methodological. It will only be shown that the two empirical variants found by the analysis correspond to two distinct syntactic strategies for the derivation of focus-raising.

#### 4.4.1. Movement and resumption

Previous analyses (e.g., É. Kiss 1987; Kenesei 1994; Lipták 1998) all treat focus-raising as a kind of movement—the focused element in the embedded clause raises into the matrix focus position. This scenario is plausible for the agreeing variant, i.e., group 2, but it is unsatisfactory for the anti-agreement option (group 1). Anti-agreement cannot be derived if the

DP starts out as the subject of the embedded verb, since anti-agreement is not tolerated at all in simple clauses. Therefore, I assume that the focused DP is base-generated in the position that the expletive occupies in the corresponding expletive construction (recall (1)), and it moves to the matrix subject position from there. As for the embedded subject position, it is filled by a dummy resumptive pronoun, which is coindexed with the focused DP. Since this latter is morphologically singular, but semantically plural, the resumptive pronoun can inherit either singular or plural features through this coindexation, and can thus trigger either singular or plural agreement on the verb. (For the technical details of this account, see Gervain 2002.)

This picture readily explains the correlation between agreement and case properties. In the movement variant, the focused DP is assigned Nominative case in the embedded clause and optionally Accusative in the matrix, so it may exhibit either (as to how this is technically possible, see Español Echevarría–Ralli 2000). Note that whichever case it exhibits, it incurs a mild violation of some constraint. If the DP keeps its original Nominative case, the Accusative of the matrix verb remains unassigned, while if the DP takes the Accusative, it will have double case, which is a possible, but rare and marked option in natural language. This is the reason why judgments by group 2 speakers never reach the level of perfect grammaticality (cf. (7)).

When the DP is base-generated as a matrix argument, as in the resumptive strategy, only Accusative case is available for it, hence the obligatory Accusative case with anti-agreement.

#### 4.4.2. A prediction: island constraints

Another prediction of the account that is nicely borne out by empirical data is the differential behavior of the two groups of speakers with respect to island constructions. As it is generally held since Chomsky (1981; 1982), (at least certain types of) resumptive elements are used as last resort mechanisms to overcome movement constraint violations. My analysis thus predicts that group 2 speakers, who employ the movement strategy, do not accept sentences violating movement constraints, while group 1 informants, for whom no movement takes place, do. To test this hypothesis, sentences with complex NP islands were included in the questionnaire (*2, 10, 19, 23, 26, 32*). As Figure 5 shows, the predictions are in complete conformity with the empirical results.

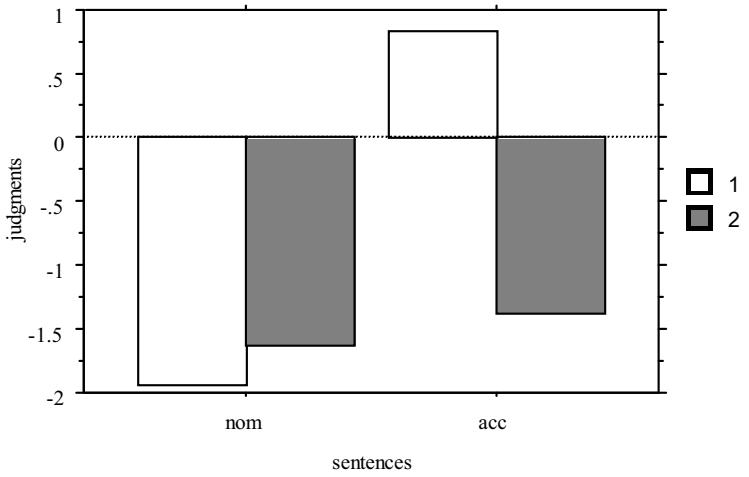


*Fig. 5*
Complex NP island effects in group 1 (resumption) and group 2 (movement)

## 5. Conclusion

The article has discussed the causes and consequences of empirical fuzziness in syntactic research. To overcome such flaws, a new methodology has been proposed which not only filters out inadequacies due to experimental noise and error, but is also able to systematize the remaining variation.

Through the example of focus-raising, it has been demonstrated that this novel method is adapted for the clarification of empirical issues, integrating conflicting data judgments into a meaningful framework. Concretely, for the case of focus-raising, the two cited authors represent different syntactic dialects. Neither of them can be rejected on empirical grounds, but their analyses are not commensurable either, reflecting different syntactic options.

Addressing the initial question of Levine (2001) then, what should happen is the experimental exploration of data, as a result of which a new, comprehensive empirical picture should emerge, accommodating conflicting judgments.

## Appendix A

| | |
|---|---|
| | 1.[9] A KÉT LEGJOBB BARÁTOD mondtad, hogy még sosem látták egymást. |
| | 2. AZ ELNÖKÖT mondtad, hogy hallottad a hírt, hogy megérkezett. |
| | 3. AZ ÖSSZES LÁNY mondtad, hogy jön. |
| | A LÁNYOK mondták, hogy későn jönnek.[10] |
| | 4. KÉT SZINÉSZNŐ tudod, hogy öngyilkosok lettek. |
| | A VENDÉGEK még nem látták egymást.[10] |
| | 5. PÉTERT mondtad, hogy meghívták. |
| | 6. VALAMELYIK FIÚ szeretnéd, hogy jöjjön. |
| | 7. AZ ÖSSZES LÁNYT mondtál, hogy jönnek. |
| | 8. KÉT SZOMSZÉDODAT hiszed, hogy tegnap szembetalálkoztak az utcán. |
| | A FIÚK voltak sokan.[10] |
| | 9. HÁROM SRÁC mondtad, hogy körbevették a lányt. |
| | 10. AZ ÖSSZES VENDÉGET mondtad, hogy hallottad a hírt, hogy megérkezett. |
| | 11. A KÉT LEGJOBB BARÁTODAT mondtad, hogy még sosem látta egymást. |
| | 12. HÁROM SRÁCOT mondtál, hogy körbevette a lányt. |
| | 13. KÉT FIÚT mondtál, hogy becsapták őket. |
| | 14. AZ ÖSSZES LÁNY mondtad, hogy jönnek. |
| | 15. A KÉT LEGJOBB BARÁTOD mondtad, hogy még sosem látta egymást. |
| | 16. KÉT FIÚT mondtál, hogy ő jön. |
| | 17. NÉGY SEBESÜLT jelentették, hogy meghaltak. |
| | 18. KÉT SZÍNÉSZNŐT tudsz, hogy öngyilkos lett. |
| | 19. AZ ÖSSZES VENDÉG mondtad, hogy hallottad a hírt, hogy megérkezett. |
| | Nem AZ ELNÖK érkezett meg.[10] |
| | 20. NÉGY SEBESÜLTET jelentettek, hogy meghaltak. |
| | 21. HÁROM SRÁCOT mondtad, hogy körbevette a lányt. |
| | 22. PÉTERT mondtad, hogy jön. |
| | 23. AZ ELNÖK mondtad, hogy hallottad a hírt, hogy megérkezett. |
| | 24. KÉT SZÍNÉSZNŐ tudod, hogy öngyilkos lett. |
| | 25. AZ ÖSSZES LÁNYT mondtad, hogy jönnek. |
| | 26. AZ ÖSSZES VENDÉGET mondtad, hogy hallottad a hírt, hogy megérkeztek. |
| | 27. HÁROM SRÁC mondtad, hogy körbevette a lányt. |
| | KÉT SZÍNÉSZNŐ érkezett meg.[10] |
| | SOK BARÁTOM jött el.[10] |
| | 28. KÉT SZOMSZÉDOD hiszed, hogy tegnap szembetalálkoztak az utcán. |
| | HÁROM SEBESÜLT halt meg.[10] |
| | 29. AZ ÖSSZES LÁNYT mondtad, hogy jön. |
| | 30. HAT MEGHÍVOTTRÓL tudom, hogy jönnek. |

| | |
|---|---|
| | 31. HÁROM SRÁCOT mondtál, hogy körbevették a lányt. |
| | 32. AZ ÖSSZES VENDÉG mondtad, hogy hallottad a hírt, hogy megérkeztek. |
| | 33. KÉT SZOMSZÉDODAT hiszel, hogy tegnap szembetalálkozott az utcán. |
| | PÉTERREL találkoztam szembe az utcán.[10] |
| | 34. KÉT SZOMSZÉDODAT hiszel, hogy tegnap szembetalálkoztak az utcán. |
| | 35. NÉGY BARÁTOMRÓL tudom, hogy hasonlít az apjára. |
| | 36. AZ ÖSSZES LÁNYT mondtál, hogy jön. |
| | 37. A KÉT LEGJOBB BARÁTODAT mondtad, hogy még sosem látták egymást. |
| | 38. NÉGY BARÁTOMRÓL tudom, hogy hasonlítanak az apjukra. |
| | ANNA hívta meg a fiúkat.[10] |
| | 39. KÉT FIÚ mondtad, hogy ők jönnek. |
| | 40. HÁROM SRÁCOT mondtad, hogy körbevették a lányt. |
| | 41. HAT MEGHÍVOTTRÓL tudom, hogy jön. |
| | 42. KÉT FIÚ mondtad, hogy ő jön. |
| | 43. PÉTER mondtad, hogy jön. |
| | Biztos, hogy PÉTER jön el.[10] |
| | 44. NÉGY SEBESÜLTET jelentettek, hogy meghalt. |
| | 45. KÉT SZÍNÉSZNŐT tudsz, hogy öngyilkosok lettek. |
| | 46. KÉT SZOMSZÉDOD hiszed, hogy tegnap szembetalálkozott az utcán. |
| | 47. VALAMELYIK FIÚ szeretnéd, hogy meghívják. |
| | 48. KÉT FIÚT mondtál, hogy ők jönnek. |
| | 49. NÉGY SEBESÜLT jelentették, hogy meghalt. |
| | ANNA találkozott az elnökkel.[10] |
| | 50. VALAMELYIK FIÚT szeretnéd, hogy jöjjön. |
| | 51. KÉT SZOMSZÉDODAT hiszed, hogy tegnap szembetalálkozott az utcán. |
| | KÉT CSOMAG sohasem érkezett meg.[10] |
| | 52. PÉTER mondtad, hogy meghívták. |
| | 53. VALAMELYIK FIÚT szeretnéd, hogy meghívják. |

[9] Numbers only appear here for the sake of easier reference. They were not contained in the original questionnaire.

[10] The sentence is a filler. It does not appear in the statistic analyses.

## Appendix B

**Test of Variance**

| | Variance | df | Chi 2 | p | 95% Inf. | 95% Sup. |
|---|---|---|---|---|---|---|
| S1 | 2,964 | 22 | 65,217 | <,0001 | 1,922 | 5,286 |
| S2 | 2,680 | 22 | 58,957 | <,0001 | 1,738 | 4,778 |
| S3 | 2,059 | 22 | 45,304 | ,0049 | 1,335 | 3,672 |
| S4 | 1,067 | 22 | 23,478 | ,7503 | ,692 | 1,903 |
| S5 | 1,340 | 22 | 29,478 | ,2634 | ,869 | 2,389 |
| S6 | 3,178 | 22 | 69,913 | <,0001 | 2,061 | 5,666 |
| S7 | 1,040 | 22 | 22,870 | ,8183 | ,674 | 1,854 |
| S8 | 2,605 | 22 | 57,304 | ,0001 | 1,689 | 4,645 |
| S9 | 1,431 | 22 | 31,478 | ,1735 | ,928 | 2,551 |
| S10 | 2,265 | 22 | 49,826 | ,0012 | 1,469 | 4,038 |
| S11 | 2,885 | 22 | 63,478 | <,0001 | 1,871 | 5,145 |
| S12 | 2,727 | 22 | 60,000 | <,0001 | 1,769 | 4,863 |
| S13 | 2,573 | 22 | 56,609 | ,0001 | 1,669 | 4,588 |
| S14 | 1,885 | 22 | 41,478 | ,0144 | 1,223 | 3,362 |
| S15 | 1,992 | 22 | 43,826 | ,0075 | 1,292 | 3,552 |
| S16 | 1,040 | 22 | 22,870 | ,8183 | ,674 | 1,854 |
| S17 | ,391 | 22 | 8,609 | ,0097 | ,254 | ,698 |
| S18 | 2,723 | 22 | 59,913 | <,0001 | 1,766 | 4,856 |
| S19 | ,723 | 22 | 15,913 | ,3596 | ,469 | 1,290 |
| S20 | 2,565 | 22 | 56,435 | ,0001 | 1,664 | 4,574 |
| S21 | 2,111 | 22 | 46,435 | ,0035 | 1,369 | 3,764 |
| S22 | ,747 | 22 | 16,435 | ,4125 | ,484 | 1,332 |
| S23 | ,723 | 22 | 15,913 | ,3596 | ,469 | 1,290 |
| S24 | 3,340 | 22 | 73,478 | <,0001 | 2,166 | 5,955 |
| S25 | 2,391 | 22 | 52,609 | ,0005 | 1,551 | 4,264 |
| S26 | 3,202 | 22 | 70,435 | <,0001 | 2,076 | 5,709 |
| S27 | 2,083 | 22 | 45,826 | ,0042 | 1,351 | 3,714 |
| S28 | 2,656 | 22 | 58,435 | <,0001 | 1,722 | 4,736 |
| S29 | ,897 | 22 | 19,739 | ,8013 | ,582 | 1,600 |
| S30 | ,043 | 22 | ,957 | <,0001 | ,028 | ,078 |
| S31 | 1,352 | 22 | 29,739 | ,2499 | ,877 | 2,410 |
| S32 | ,696 | 22 | 15,304 | ,3023 | ,451 | 1,240 |
| S33 | 1,613 | 22 | 35,478 | ,0690 | 1,046 | 2,876 |
| S34 | 1,443 | 22 | 31,739 | ,1639 | ,936 | 2,572 |
| S35 | 0,000 | 22 | 0,000 | • | • | • |
| S36 | 0,000 | 22 | 0,000 | • | • | • |
| S37 | 2,542 | 22 | 55,913 | ,0002 | 1,648 | 4,532 |
| S38 | ,696 | 22 | 15,304 | ,3023 | ,451 | 1,240 |
| S39 | ,747 | 22 | 16,435 | ,4125 | ,484 | 1,332 |
| S40 | ,391 | 22 | 8,609 | ,0097 | ,254 | ,698 |
| S41 | 0,000 | 22 | 0,000 | • | • | • |
| S42 | ,391 | 22 | 8,609 | ,0097 | ,254 | ,698 |
| S43 | 3,783 | 22 | 83,217 | <,0001 | 2,453 | 6,745 |
| S44 | 3,059 | 22 | 67,304 | <,0001 | 1,984 | 5,455 |
| S45 | 2,403 | 22 | 52,870 | ,0005 | 1,558 | 4,285 |
| S46 | 2,953 | 22 | 64,957 | <,0001 | 1,915 | 5,265 |
| S47 | ,767 | 22 | 16,870 | ,4590 | ,497 | 1,367 |
| S48 | 2,542 | 22 | 55,913 | ,0002 | 1,648 | 4,532 |
| S49 | 2,696 | 22 | 59,304 | <,0001 | 1,748 | 4,807 |
| S50 | 2,846 | 22 | 62,609 | <,0001 | 1,846 | 5,074 |
| S51 | 2,715 | 22 | 59,739 | <,0001 | 1,761 | 4,842 |
| S52 | 2,605 | 22 | 57,304 | ,0001 | 1,689 | 4,645 |
| S53 | 1,340 | 22 | 29,478 | ,2634 | ,869 | 2,389 |

## References

Chomsky, Noam 1957. Syntactic structures. Mouton, The Hague.

Chomsky, Noam 1961. Some methodological remarks on generative grammar. In: Word 17 : 219–39.

Chomsky, Noam 1972. Language and mind. Harcourt, New York.

Chomsky, Noam 1981. Lectures on government and binding. Foris, Dordrecht.

Chomsky, Noam 1982. Some concepts and consequences of the theory of government and binding. (Linguistic Inquiry Monographs 6). MIT Press, Cambridge MA.

Chomsky, Noam 1995. The Minimalist Program. MIT Press, Cambridge MA.

Chomsky, Noam 2000. New horizons in the study of language and mind. Cambridge University Press, Cambridge.

Chomsky, Noam 2002. On nature and language. Cambridge University Press, Cambridge.

Chomsky, Noam – Jerrold J. Katz 1974. What the linguist is talking about. In: The Journal of Philosophy 71 (12) : 347–67.

Cornips, Leonie 1998. Syntactic variation, parameters, and social distribution. In: Language Variation and Change 10 : 1–21.

Cornips, Leonie – Karen Corrigan forthcoming. Convergence and divergence in grammar. In: Peter Auer – Franks Hinskens – Paul Kerswill (eds): The convergence and divergence of dialects in contemporary societies. Cambridge University Press, Cambridge.

Cowart, Wayne 1997. Experimental syntax: applying objective methods to sentence judgments. Sage Publications, Thousand Oaks.

Duhem, Pierre 1954. The aim and structure of physical theory. Princeton University Press, Princeton.

É. Kiss, Katalin 1987. Configurationality in Hungarian. Kluwer, Dordrecht.

Elman, Jeff 1993. Learning and development of neural networks: the importance of starting small. In: Cognition 48 : 71–99.

Elman, Jeff – Elizabeth Bates – Mark Johnson – Annette Karmiloff-Smith – Domenico Parisi – Kim Plunkett 1996. Rethinking innateness. MIT Press, Cambridge MA.

Español Echevarría, Manuel – Angela Ralli 2000. Case mismatches in Greek: Evidence for the autonomy of morphology. In: Acta Linguistica Hungarica 47 : 179–203.

Everitt, Brian 1981. Cluster analysis. Heinemann-Halsted, London & New York.

Fodor, Jerry D. 2001. Parameters and the periphery: reflections on syntactic nuts. In: Journal of Linguistics 37 : 367–92.

Gervain, Judit 2002. Linguistic methodology and microvariation in language. MA thesis, University of Szeged, Szeged.

Gómez, Rebecca L. – LouAnn Gerken 2000. Infant artificial language learning and language acquisition. In: Trends in Cognitive Sciences 4 : 178–86.

Guasti, Maria Theresa 2002. Language acquisition. MIT Press, Cambridge MA.

Hill, Archibald A. 1961. Grammaticality. In: Word 17 : 1–10.

Hoji, Hajime 2002. Falsifiability and repeatability in generative grammar. Talk given at TSSS, UIL-OTS, University of Utrecht on the 16 April 2002.

Keller, Frank 2000. Gradience in grammar. Doctoral dissertation, University of Edinburgh, Edinburgh.

Kenesei, István 1994. Subordinate clauses. In: Ferenc Kiefer – Katalin É. Kiss (eds): The syntactic structure of Hungarian. (Syntax and Semantics 27), 275–354. Academic Press, New York.

Labov, William 1975. What is a linguistic fact? The Peter de Ridder Press, Lisse.

Lance, G. N. – W. T. Williams 1967. A general theory of classificatory sorting strategies. 1. Hierarchical systems. In: Computer Journal 9 : 373–80.

Levine, Robert 2001. The extraction riddle: just what are we missing? In: Journal of Linguistics 37 : 145–74.

Lipták, Anikó 1998. A magyar fókuszemelések egy minimalista elemzése [A minimalist analysis of focus raising in Hungarian]. In: László Büky – Márta Maleczki (eds): A mai magyar nyelv leírásának újabb módszerei III, 93–115. JATEPress, Szeged.

McDaniel, Dana – Wayne Cowart 1999. Experimental evidence for a minimalist account of English resumptive pronouns. In: Cognition 70 : B15–B24.

Oostendorp, Marc van 2002. Microvariation and grammatical theory. Lecture notes. (http://www.meertens.knaw.nl/medewerkers/marc.van.oostendorp/microvariation)

Ooyen, Arjen van 2001. Theoretical aspects of pattern analysis. In: L. Dijkshoorn – K. Dijkshoorn – J. Towner – M. Struelens (eds): New approaches for the generation and analysis of microbial fingerprints, 31–45. Elsevier, Amsterdam.

Pléh, Csaba 2000. A lélektan története [The history of psychology]. Osiris Kiadó, Budapest.

Quine, Willard 1953. From a logical point of view. Harvard University, Cambridge MA.

Quine, Willard 1972. Methodological reflections on current linguistic theory. In: Donald Davidson – Harman Gilbert (eds): Semantics of natural language, 442–54. Reidel, Dordrecht.

Quine, Willard 1975. On empirically equivalent systems of the world. In: Erkenntnis 9 : 313–28.

Schütze, Carlson 1996. The empirical base of linguistics. Grammaticality judgments and linguistic methodology. Chicago University Press, Chicago.

Sorace, Antonella – Frank Keller forthcoming. Gradience in linguistic data. In: Lingua.

Stich, Stephen P. 1971. What every speaker knows. In: The Philosophical Review 80 (4) : 476–96.

Stich, Stephen P. 1972. Grammar, psychology, and indeterminacy. In: The Journal of Philosophy 69 (22) : 799–818.

Tsohatzidis, Savas 2002. Grammars as objects of knowledge: the availability of dispositionalism. In: Language Sciences 24 : 97–106.

Zemplén, Gábor – Judit Gervain forthcoming. Az elefántcsonttorony átépítése [Rebuilding the ivory tower]. In: János Kelemen (ed.): Proceedings of the Conference Norma, kijelentés, cselekvés [Norms, statements, acts].

Ziff, Paul 1964. About ungrammaticalness. In: Mind 73 (290) : 204–14.

Address of the author: Judit Gervain
SISSA CNS
via Beirut 2–4
34014 Trieste
Italy
gervain@sissa.it