A Comparison of HPSG and LFG

Valia Kordoni, Frank Richter

ESSLLI 2000,
Birmingham,
8–18 August 2000

I

Pollard and Sag 1994

Methodological Background

Let us begin by making explicit some methodological assumptions. In any mathematical theory about an empirical domain, the phenomena of interest are *modelled* by mathematical structures, certain aspects of which are conventionally understood as corresponding to observables of the domain. The theory itself does not talk directly about the empirical phenomena; instead, it talks about, or is *interpreted by*, the modelling structures. Thus the predictive power of the theory arises from the conventional correspondence between the model and the empirical domain. (Pollard and Sag, 1994, p. 6)

In our view, a linguistic theory should bear exactly the same relation to the empirical domain of natural language, namely, the universe of possible linguistic objects, as a mathematical theory of celestial mechanics bears to the possible motions of n-body systems. Thus we insist on being explicit as to what sorts of constructs are assumed (i.e. what ontological categories of linguistic objects we suppose to populate the empirical domain) and on being mathematically rigorous as to what structures are used to model them. Moreover, we

require that the theory itself actually count as a theory in the technical sense of precisely characterizing those modelling structures that are regarded as admissible or well-formed (i.e. corresponding to those imaginable linguistic objects that are actually predicted to be possible ones). This does not mean that the empirical hypotheses must be rendered in a formal logic as long as their content can be made clear and unambiguous in natural language (the same holds true in mathematical physics), but in principle they must be capable of being so rendered. (Pollard and Sag, 1994, p. 7)

In HPSG, the modelling domain—the analog of the physicist's flows—is a system of *sorted feature structures* . . . , that are intended to stand in a one-to-one relation with the types of natural language expressions and their subparts. The role of linguistic theory is to give a precise specification of which feature structures are to be considered admissible; the types of linguistic entities that correspond to the admissible feature structures constitute the predictions of the theory. (Pollard and Sag, 1994, p. 8)

Since our principle goal in this volume is to propose analyses of linguistic phenomena and make them intelligible to the linguistic community, we eschew extreme formalization here and thereby avoid the many technical design decisions associated with the problem of choosing a feature logic; our rules and principles, in the form of feature structure constraints, will be expressed—clearly and unambiguously, we hope—variously in English or in a quasi-formal description language that we turn to directly. We doubt that the relative informality of analyses such as those presented here is likely to impede understanding; if anything, probably the reverse is true. Nevertheless we consider formalization of the theory an important goal, indeed a necessary one for the proof of computational properties . . . (Pollard and Sag, 1994, p. 9)

In our concern for processing issues . . . we have accepted the conventional wisdom that linguistic theory must account for linguistic knowledge . . . but not necessarily for processes by which that knowledge is brought to bear in the case of individual linguistic tokens. Indeed, we take it to be the central goal of linguistic theory to characterize what it is that every linguistically mature human being knows by virtue of being a linguistic creature, namely, universal grammar. And a theory of a particular language—a grammar—characterizes what linguistic knowledge (beyond universal grammar) is

shared by the community of speakers of that language. Indeed, from the linguist's point of view, that is what language is.

But what does language consist of? One thing that it certainly does not consist of is individual linguistic events or utterance tokens, for knowledge of these is not what is shared among the members of a linguistic community. Instead, what is known in common, that makes communication possible, is the system of linguistic types. For example, the type of the sentence *I'm sleepy* is part of that system, but no individual token of it is. (Pollard and Sag, 1994, p. 14)

Modelling types of conceivable linguistic entities as rooted labelled graphs of a special kind—totally well-typed, sort-resolved feature structures—we formulate universal grammar and grammars of particular languages as systems of constraints on those feature structures. Only those feature structures that satisfy the constraints are taken to model (types of) grammatically well-formed linguistic entities. The distinction between the system of constraints and the collection of linguistic entities that satisfies it can be viewed as corresponding both to Chomsky's (1986a) distinction between *I-language* and *E-language* and to Saussure's ((1916) 1959) distinction

between *langue* and *parole*. Though only the latter is directly observable, only the former can be embodied as a mental computational system shared by the members of a linguistic community. (Pollard and Sag, 1994, pp. 57–58)

Some Properties of the Modelling Structures

- The first formal point to be noted is that the feature structures employed in HPSG are *sorted*. This means simply that each node is labelled with a *sort symbol* that tells what type of object the structure is modelling; that is, there is one sort symbol for each basic type (ontological category) of construct . . . The (finite) set of all sort symbols is assumed to be partially ordered . . . (Pollard and Sag, 1994, pp. 17–18)

- Second, we require feature structures to be *well-typed* . . . This means that what attribute (or, equivalently, feature) labels can appear in a feature structure are [*sic*] determined by its sort; this fact is the reflection within the model of the fact that what attributes (or, equivalently, features or components of structure) an empirical object has depends on its ontological category. (Pollard and Sag, 1994, p. 18)

- Third, feature structures that serve as models of linguistic entities are required to satisfy further criteria of completeness. Roughly what this means is that they are total (not merely partial) models of the objects that they represent. More precisely, they are required to be both *totally well-typed* . . . and *sort-resolved*. A feature structure is *totally* well-typed in case it is well-typed, and, moreover, for each node, every feature that is appropriate for the sort assigned to that node is actually present . . . A feature structure is *sort-resolved* provided every node is assigned a sort label that is maximal (i.e. most specific) in the sort ordering. (Pollard and Sag, 1994, p. 18)

Methodological Assumptions: Summary and Outlook

Principle among these [methodological assumptions] are

- a concern with analyses of particular phenomena that are worked out in detail;
- insistence on mathematical precision (at least in principle) with respect to the grammar formalism itself and its intended interpretation;
- and commitment to the criterion of empirical consequence, that is, the requirement that given a grammar and (a model of) a candidate linguistic expression, it is a determinate matter of fact whether or not the entity satisfies the grammar.

This last criterion is fundamental to the scientific method and at the same time provides a point of departure for meeting the more ambitious criterion of decidability . . . (after Pollard and Sag, 1994, p. 59)

II

Pollard 1999

Methodological Assumptions

Grammars exist in the real world, more specifically the part of the world inside language knower's minds; utterance tokens are events, also in the real world but external (at least in part) to minds. Formal linguists create idealizations of both these things.

- A formal grammar, a certain kind of logical theory, is a mathematical idealization of the mental grammar;
- and utterance tokens which are judged grammatical by native language knowers are idealized as certain mathematical structures, in such a way that distinct utterances which are linguistically indistinguishable, e.g., two utterances of *Poor John ran away*, are idealized as isomorphic structures.

What I think ties all these things together is that, given an appropriate scheme of semantic interpretation for formal grammars, and assuming it is a good grammar, the mathematical structures which idealize the grammatical utterance tokens will be certain kinds of *models* of the grammar. (Pollard, 1999, p. 281)

On the other hand, generative grammarians are accustomed to thinking of a formal grammar as *generating* a set of structures, called the *strong generative capacity* (SGC) of the grammar. This set is usually considered to have the following properties: (1) no two members are structurally isomorphic; and (2) if the grammar is making good predictions, then utterance tokens whose structures are isomorphic to members of the SGC will be judged grammatical. (Pollard, 1999, p. 282)

Now very roughly speaking, the relationship between an HPSG grammar and its SGC is analogous to the relationship between a first-order theory and its models. There are, however, some subtle differences . . . (Pollard, 1999, p. 284)

It might be expected that a model of the grammar . . . could be employed as the SGC of the grammar. Such a notion of the SGC has the virtue of mathematical simplicity, but it suffers from a number of defects. First, there can be many models of a given grammar, and these need not all be isomorphic. Second, some models are 'too small', in the sense that the natural language whose grammar we are trying to write might have some grammatical expressions that do not correspond

to any object of the model. And third, models may differ from each other in ways that make no linguistic difference, e.g., the number of objects in the model which correspond to the sentence *Poor John ran away*. There is a clear intuitive sense in which this should not matter linguistically; we want a notion of SGC which abstracts away from how many 'instances' or 'tokens' of a given expression there are, and instead is sensitive only [to] isomorphism classes. (Pollard, 1999, p. 288)

Important differences to Pollard and Sag 1994:

- No ontological commitment to types; 'type' is replaced with 'isomorphism classes' in the relevant contexts.

- Technically, the conventional feature structures of Pollard and Sag 1994 are replaced with certain canonical models. Although Pollard keeps referring to them as 'feature structures' (as can be justified by certain mathematical correspondences), they are not traditional feature structures.

III

King 1999

Criticism

Against Object Types

Almost all linguistic tokens are partly empirical. Linguistic communication would be almost impossible otherwise. But to what extent are linguistic types empirical? Is it an empirical matter whether a linguistic type is part of a natural language? . . .

The only possible basis for claiming that a linguistic type is part of a natural language is that the linguistic knowledge of an ordinary user of the natural language distinguishes that particular class of tokens. However, it is far from certain that linguistic knowledge *does* distinguish particular classes of linguistic tokens. And even if it does then the extent to which linguistic types are empirical cannot exceed the extent to which linguistic knowledge is empirical. Now, linguistic knowledge is certainly empirical in part . . . Nonetheless, the observation of linguistic knowledge itself is currently very limited and usually indirect. Thus, (Pollard and Sag, 1994) literally characterises truth in such a manner that the veracity of a HPSG grammar of a natural language is immune—possibly in principle, and certainly in current practice—to experimental assessment. HPSGians must either content themselves with writing untestable theories and hoping for the appropriate advances in cognitive technology, or characterise truth anew in a fashion that is susceptible to currently available means of empirical appraisal. (King, 1999, pp. 307–308)

In light of the wholly metempirical nature of linguistic types [the principle of only postulating constructs that correspond to observables in the empirical domain] entails that a characterisation of truth should avoid linguistic types unless to do so compromises the simplicity and elegance of HPSG grammars. Indeed, parsimony suggests that feature structures also be avoided where possible. (King, 1999, p. 310)

Against the Presentation of the Formalism in (Pollard and Sag, 1994)

Secondly, the (Pollard and Sag, 1994) characterisation is poorly defined. Consider feature structures. (Pollard and Sag, 1994) is clear about some aspects of feature structures, but not others. We are unequivocally told that feature structures are both totally well-typed and sort resolved. But whether, to use the terminology of (Moshier, 1988), a feature structure is *concrete*—an individual graph—or *abstract*—an isomorphism class

of graphs—is unclear.. . . Also, (Pollard and Sag, 1994) nowhere states whether feature structures are necessarily finite or possibly infinite. Finite feature structures certainly predominate in the feature-structure literature that (Pollard and Sag, 1994) cites, yet (Carpenter, 1992) clearly explicates the tensions that arise if feature structures are both totally well-typed and finite. . . (King, 1999, pp. 310–311)

Finally, King points out that Pollard and Sag (1994) do not define a formal description language, and that the literature they refer to does not provide the classical description language that the grammatical principles of Pollard and Sag presuppose. (King, 1999, pp. 311–312)

A final Concern

My third—and I believe insuperable—objection leads me to abandon any attempt to repair and extend the (Pollard and Sag, 1994) characterisation. A definition must be not only well-defined but also accurate, in the sense that the definition must faithfully and demonstrably capture the intuitive concept it is intended to express.. . . Presumably, the grammarian has intuitions about linguistic types and constructs their grammar in accord with those intuitions. However, the grammar is formally interpreted via the (Pollard and Sag, 1994) characterisation of truth. Thus, the grammar reflects the intuitions of the grammarian only to the extent that (i) the intuitions of the grammarian are commensurate with those that underly the (Pollard and Sag, 1994) characterisation, and (ii) the (Pollard and Sag, 1994) formal machinery faithfully expresses those underlying intuitions. My gravest concern over the (Pollard and Sag, 1994) characterisation is that (Pollard and Sag, 1994) (and (Pollard, 1999)) barely mentions what those underlying intuitions are, far less demonstrates that the formal machinery faithfully expresses them. Nor can we rely on a consensual intuition shared generally by linguists. The fundamental

ontological notion of the (Pollard and Sag, 1994) characterisation, the linguistic type, is an amorphous, artificial notion that linguists can and do construe in a variety of ways to suit their particular purposes. (King, 1999, p. 313)

King's Proposal

Motivated by his considerations of ontological simplicity, King suggests a construal of a natural language as a system of linguistic tokens. This construal is then used in a new, formal characterization of linguistic truth.

A token can be an object such as my computer, an event such as my typing these words at my computer keyboard, a state such as my being at work, a location such as the space occupied by my computer, or a complex such as the contents of the file system of my computer. . . The intuition here is that a grammar of a natural language should account for the linguistic knowledge of mature users of the natural language, but not for the biological, physical and psychological limitations the universe imposes upon the exercise of that knowledge. I share this intuition, but maintain that a scientifically worthwhile account of linguistic knowledge can presently only be given largely in terms of linguistic behaviour. In order to reconcile this apparent antithesis, I endorse the notion of nonactual tokens—tokens that could exist in some universe but happen not to in ours—and assume that the linguistic knowledge of each mature user of a natural language determines which

tokens—actual and nonactual—are tokens of the natural language. A grammar of a natural language is then true only if the grammar delimits the collection comprising all and only those tokens—actual and nonactual—that the linguistic knowledge of mature users of the natural language would deem tokens of the natural language. (King, 1999, p. 316)

References

Carpenter, Bob 1992. *The Logic of Typed Feature Structures*. Cambridge University Press. Cambridge, Massachusetts, USA.

King, Paul J. 1999. Towards Truth in Head-driven Phrase Structure Grammar. In Valia Kordoni (ed), *Tübingen Studies in Head-Driven Phrase Structure Grammar*, (= *Arbeitspapiere des SFB 340, Nr. 132, Volume 2*), 301–352. Eberhard-Karls-Universität Tübingen.

Moshier, Michael Andrew 1988. *Extensions to Unification Grammar for the Description of Programming Languages*. PhD thesis, University of Michigan.

Pollard, Carl and Sag, Ivan A. 1994. *Head-Driven Phrase Structure Grammar*. University of Chicago Press.

Pollard, Carl J. 1999. Strong generative capacity in HPSG. In Gert Webelhuth, Jean-Pierre Koenig, and Andreas Kathol (eds), *Lexical and Constructional Aspects of Linguistic Explanation*, 281–297. CSLI Publications.