

Aaron Halpern, *On the Placement and Morphology of Clitics*, Stanford, CA, 1995. ISBN 1 881 52660 7, £17.95 (paperback). ISBN 1 881 52661 5, £35.00 (hardback).

The term *clitic* has a large thematic content. It has a history of use in various related, non-identical senses. Two core examples of clitics are unstressed pronouns in French, e.g. *le* in (1), and auxiliaries in Serbo-Croatian, e.g. *je* in (2).

- (1) *Pierre le voit.* (*Pierre him-sees*, i.e. *Pierre sees him*)
- (2) *Taj-je čovek svirao klavir.* (*That-AUX man played piano*)

Halpern's book is a contribution to a body of studies that assume that the phonology and distribution of "real" clitics can be accounted for entirely by phrasal rules. A result of this assumption is that some of the core cases of clitics have to be treated as something other than clitics. Zwicky (1987) observed that in phrases like (3) the morpheme pronounced "/s/" is realized once, but interpreted twice (haplology).

- (3) the dogs' kennel

The contexts where this haplology occurs are restricted to words ending in this morpheme. Since the realization of the possessive depends on individual morphemes in the context, it cannot be a clitic. Therefore, the English possessive is, according to Zwicky, inflectional. Miller (1992) uses a similar line of reasoning to show that French pronouns as in (1) are inflection.

At this point, the following two problems emerge:

- How do we get the right wordform in the right place?
- How do we restore a unified concept of inflection?

In the discussions by Zwicky and Miller, the first problem is central, whereas the second problem is largely ignored. In ten Hacken (1994), it is shown that phonological criteria like haplology, as used by Zwicky and Miller, are not suitable if we want to arrive at a unified concept of inflection. The problem is not only that certain phenomena are labelled inflection that are not generally considered as such, but also that in many cases the criteria proposed by Zwicky and by Miller do not lead to a decision whether something is inflection or not.

Both problems are solved by an alternative definition of inflection based on agreement in certain configurations. Among its consequences are the exclusion of the English possessive and of French weak pronouns from inflection. Like Zwicky and Miller, Halpern concentrates on the first problem, but his presentation of the argument for the inflectional status of the controversial items is much stronger than Zwicky's and Miller's, and presents a serious challenge to ten Hacken's system of definitions.

Halpern's book consists of three parts and a brief introduction. He starts with an interesting account of the clitics whose status is beyond dispute, before addressing the controversial cases where phonology and distribution seem to point in diverging directions. The final chapter is devoted to properties of the cluster of clitics and inflectional affixes grouped together in one syntactic position.

The discussion in the first part (Chapters 2–3) focuses on second position (2P) clitics, especially in Serbo-Croatian. In Serbo-Croatian, 2P clitics can appear either after the first word (2W), as in (2), or after the first daughter (2D), as in (4):

- (4) *Taj čovek-je svirao klavir.*

The mechanism Halpern develops combines syntactic and phonological rules. 2P clitics are generated to the left of the IP, but since they are specified as non-sentence-initial, some action is required to save the relevant sentences. One possibility is to front a constituent, so that the clitic can attach to its last word. This leads to 2D. 2W is the result of a phonological rule of *Prosodic Inversion*, applying as a last resort to avoid clause-initial clitics. Independent factors determine whether in a given language or construction the conditions for 2D or 2W can be

satisfied. In support of this mechanism, data from a variety of languages are addressed. Even though not all details are explained, my general impression is that of an appealing solution. Unfortunately, in order to account for the syntactic position of Serbo-Croatian 2P clitics, Halpern introduces a functional category Cleft with its projection CleftP between CP and IP. This move seems fairly ad hoc, because it is motivated rather by the lack of an alternative than by external evidence and it is unclear whether this projection appears in any other language.

In the second part of his book (Chapters 4–5), Halpern introduces the notion of *extended inflection* for clitics that do not behave as real clitics should. Extended inflection is accounted for by a system of trigger and marker features manipulated in a GPSG-like fashion. Miller's (1992) Edge Feature Principle is invoked to ensure correct percolation and triggering. This approach is compared to solutions within the framework of Sadock's (1991) Autolexical Syntax.

In Chapter 4, the English possessive is addressed. An example of phrasal attachment is (5):

(5) the mayor of Oakland's office

The phrase structure rule producing an NP like (5) introduces the feature POSS on the possessive NP. The Edge Feature Principle interprets POSS as a trigger feature for PM, the possessive marker, and passes on PM recursively to the rightmost daughter until a lexical node is reached. Then, in (5), *Oakland* PM is realized as *Oakland's*. Although it can be shown that the trigger-marker mechanism leads to the correct forms for English possessives, no decisive argument can be based on it, as Halpern admits (p. 143). Therefore, he presents additional data in Chapter 5, especially from the Balkan definite article.

In Bulgarian, Rumanian, Macedonian and Albanian, the definite article seems to be a 2P clitic in the NP, as can be seen for Bulgarian in (6):

- (6) (a) kniga-ta (the book);
 (b) xubava-ta kniga (the nice book);
 (c) moja-ta xubava kniga (my nice book).

Halpern argues that, contrary to first appearance, the Balkan definite article is extended inflection. On the one hand he gives cases where the distribution is slightly different from what the 2P mechanism of Chapters 2–3 would predict and cases of phonological realization subject to lexical idiosyncrasies. On the other hand he shows that a system of trigger and marker features parallel to the one used for the English possessive is able to account for (almost) all properties of the Balkan definite article.

A number of objections occurred to me, which make the argument less than fully convincing. First, the unexpected distribution of the article might also be taken as a sign that the 2P mechanism needs to be modified. Then, what is presented as facts of phonological realization can often be reinterpreted as facts of distribution, because the only alternation is between the regular and the zero realization. One of the stronger phonological arguments is the zero-realization of the Bulgarian article on certain kinship terms. A similar distribution can be found in Italian, however:

- (7) (a) mia sorella (my sister);
 (b) la mia casa (my house).

Presumably, this should be taken as evidence that the Italian article is inflectional as well, although intuitively this result is odd. I wonder, however, how this type of lexical idiosyncrasies in distribution is distinguished from selectional restrictions or collocations, which I assume do not involve inflection. If the lexicon is allowed to specify this type of distributional exceptions, why is it not allowed to do the same when the definite article is involved?

Another problem concerns the focus on exceptions. If idiosyncratic behaviour is the criterion to recognize inflection, what happens with fully regular cases? It seems that either inflection should be restricted to processes with at least one lexical exception or it is not

possible to classify such cases at all. Although Halpern does not explore this kind of consequences, his discussion of the Balkan article is highly interesting, because it describes an amazing phenomenon in great detail.

The final chapter is entitled "Cluster morphology". Despite the morphology in the title, the chapter concentrates on phonology. Halpern argues that in Sekani, an Athabascan language described by Hargus (1988), the cluster of affixes and clitics is a phonological unit combined with the stem in a compounding process. This account is contrasted with Hargus's lexical phonological one. The Sekani data are not presented in sufficient detail to make a critical appraisal possible, but the term compounding seems to me a bad misnomer. Besides, the discussion seems to identify inflection with affixation, concentrating on templates for clusters. While Halpern refers to Anderson (1992) in support for his thesis that clitics are similar to inflectional affixes, he fails to address Anderson's point that the IA-based model of templates is mistaken.

Before I arrive at a conclusion, I have to say a few words about the presentation of the text, because I found it prominent in a negative sense. As a reader of this book you are lucky if you are not distracted by an untidy layout, inconsistency in typographical conventions, missing or repeated words, typos, hyphenations like *cl-itics*, etc. Worse are the inconsistencies in references between the main text and the list of references and the inconsistent use of often unexplained abbreviations (e.g. *SC* and *SCr* in one sentence on p. 23). Above all, however, this book would have deserved an index in order to make the many interesting ideas accessible to readers who do not write down every single idea they might want to use later on.

Let me conclude by summing up the principal strengths and weaknesses of this book. A first weak point is the lack of a uniform theoretical background. Whereas the first part tends towards GB by referring to movement and functional categories, the second part is rather based on GPSG with its mechanisms for feature handling. This makes it difficult to see the book as a comprehensive account of clitics, but does not affect the value of individual ideas.

A second weak point is that the notion of inflection resulting from the treatment of the English possessive and other phenomena as extended inflection is far removed from what is generally understood as inflection. The link with agreement and paradigms argued for by ten Hacken (1994) is entirely lost when, as Halpern is forced to assume, *to's* in (8) is an inflected form of *to*:

(8) the man I was talking to's hat

This weakness is of course a direct consequence of Halpern's point of departure and the strength of his book is that his argument in favour of this view is more convincing than others I have seen. The general line of reasoning is:

1. Clitics are accounted for by phrasal rules only.
2. Apparent counterexamples are treated as extended inflection.
3. There are independent reasons why clitics and extended inflection are different.

To my taste, a stronger structural link between the discussion and this reasoning scheme would have been desirable, but at least it is presented in the afterword. The main strength of this book is the way interesting data in support of the third step are presented. They make the book essential reading for anyone specializing in clitics and morphology.

References

- Anderson, S. R. (1992) *A-Morphous Morphology*. Cambridge University Press.
 ten Hacken, P. (1994) *Defining Morphology: A Principled Approach to Determining the Boundaries of Compounding, Derivation, and Inflection*. Hildesheim: Olms.
 Hargus, S. (1988) *The Lexical Phonology of Sekani*. New York: Garland.
 Miller, P. H. (1992) *Clitics and Constituents in Phrase Structure Grammar*. New York: Garland.

- Sadock, J. M. (1991) *Autolexical Syntax: A Theory of Parallel Grammatical Representations*. University of Chicago Press.
 Zwicky, A. M. (1987) Suppressing the Zs. *Journal of Linguistics* 23:133–148.

PIUS TEN HACKEN

University of Basel, Switzerland

email: tenhacken@ubaclu.unibas.ch

Douglas Biber. *Dimensions of Register Variation: A Cross-linguistic Comparison*. Cambridge University Press, 1995. ISBN 0 521 47331 4, £37.50 (hardback). 444pp.

This book carries forward the research program described in Biber (1988), taking the methodology described there forward into cross-linguistic and diachronic territories. The review will address the research program as a whole: the whole is covered thoroughly in this book, though various parts have already been described in the 1988 book and elsewhere.

As the title implies, Biber's goal is to chart the various ways in which language varies. This is a large goal, and might seem too broad to make for a coherent book. But Biber brings to the question a methodology which both gives coherence to the enterprise, and which makes a number of major advances on earlier work of this kind.

In this review, I first describe the methodology and present a thumbnail version of the analysis for English, to give a flavour of what it produces; then, argue why researchers in language engineering should be interested; then, outline the structure of the book, and finally, comment on the book's results and conclusions.

1 The Multi-Dimensional (MD) methodology

- Gather a set of text samples to cover a wide range of language varieties;
- Enter them ("the corpus") into the computer;
- Identify a set of linguistic features which are likely to serve as discriminators for different varieties;
- Count the number of occurrences of each linguistic feature in each text sample;
- Perform a factor analysis (a statistical procedure) to identify which linguistic features tend to co-occur in texts. The output is a set of "dimensions", each of which carries a weighting for each of the linguistic features;
- Interpret each dimension, to identify what linguistic features, and what corresponding communicative functions, high-positive and high-negative values on the dimension correspond to.

For English, Biber identifies seven dimensions, numbered in decreasing order of significance (so dimension 1 accounts for the largest part of the non-randomness of the data, dimension 2, the next largest, etc.). The first he calls "Involved versus Informational Production". Texts getting high positive scores are typically spoken, typically conversations. Texts getting high negative scores are academic prose and official documents. The linguistic features with the highest positive weightings are "private" verbs (*assume, believe* etc.), *that*-deletion, contractions, present tense verbs, and second person pronouns. The linguistic features with the highest negative weightings are nouns, word length, prepositions, and type-token ratio. (There were a total of 67 linguistic features for English.) Dimension two was "Narrative versus Non-narrative Discourse", for which past-tense verbs and third-person pronouns are the high-positive features, and fiction texts got the high positive scores, and Dimension three was "Situation-dependent versus Elaborated Reference": broadcasts are the highest-scoring texts, and official documents the lowest-scoring, with the positive linguistic features being