

Introduction

Derivation has not attracted sustained interest in typologically oriented research. At most, information on derivation in hand-books of typology concerns specific phenomena in comparatively few languages.

(Laca 2001: 1214)

The aim of this book is to identify typological differences in the word-formation of a number of languages of the world, as well as any associations which may exist between them. The purpose is to provide an overview of how word-formation is organized in different languages. From the point of view of languages, its focus is on the use of common resources in languages which belong to different genetic groups, insofar as this may be indicative of the import of those resources. From the point of view of word-formation processes and categories, its focus is on their relation as co-existing devices for the formation of new vocabulary.

The book is intended to provide a preliminary survey of what is usually called *derivational morphology*.¹ This is because the most significant achievements in the study of morphological typology and morphological universals tend to rely on cross-linguistic research into inflectional categories and properties and on their description, but not on derivation. Inflection has also been the subject of various typological classifications of languages since the times of Friedrich Schlegel (1808). Interestingly, while Edward Sapir (1921), the author of one of the most widespread morphological classifications of languages, meant to apply his typological categories to both inflection and derivation, the use of his classification and terminology, like *isolating*, *agglutinative*, *fusional*, *symbolic*, *analytic*, *synthetic* or *polysynthetic*, has been mostly biased towards inflection.

While the field of word-formation universals and word-formation typology is not an untilled area, the motto of this 'Introduction' and more recent assessments of the present state of knowledge in word-formation typology suggest that 'the result is rather miserable: no homogeneous picture either regarding the derivational categories investigated or the morphological techniques involved seems to emerge' (Gaeta 2005: 168).

One of the main causes of this state of affairs is that representative word-formation data of a substantial number and variety of languages are

¹ The terms *word-formation*, *derivation* and *derivational morphology* are used as synonyms in this book.

extremely difficult to gather. This is largely because of the lack of relevant descriptions of this kind. In turn, this is due to the almost total absence of attention paid in typological and morphological research to inflectional characteristics of languages outside the best-known European languages and perhaps some others. This is a frequent limitation of the field which has consequences of some importance. Thus, in a paradigmatic study on this issue (L. Bauer 2000), the choice of the languages examined was determined by the very low availability of detailed descriptions of derivation in the limited grammars of languages that were accessible. Bauer argues that this is due to the little attention that some grammarians pay to derivational morphology. This is meaningful and illustrative of the state of affairs in itself, but the consequences are even more telling: ‘it is frequently unclear to the reader of a description ... what is inflection and what is derivation; writers of descriptions (particularly descriptions of lesser-known languages) may not have all information to answer questions which can be answered for other languages – accordingly descriptions are not strictly comparable’ (L. Bauer 2000: 38–9).

This gap in typological research is not new nor has it been filled by renewed interest in typological studies over the past decade. While Plank and Filimonova’s (2000) ongoing data bank of linguistic universals (*The Universals Archive*, University of Konstanz) also covers morphological universals, the proportion between the total number of universals (about 2,000) and the number of inflectional morphology-related universals in the *Archive* on the one hand (about 170) and those pertaining to word-formation on the other (about sixty) speaks for itself. Significantly, in a 300-page volume published in 2007 on the occasion of the tenth anniversary of the existence of the journal *Linguistic Typology*, one can hardly find any specific reference to word-formation. It is therefore small wonder that Baerman and Corbett (2007) argue in favour of the typologists’ increased attention to morphological questions.

Antecedents

While the overall picture of the field in question is not very encouraging, the realm of derivationally relevant typology is not *terra incognita*. The general interest in word-formation over the past forty years and the growing awareness of the importance of this kind of study have set the appropriate conditions for a step forward in research into word-formation universals and typology.

Of the works of the past three decades, specifically since 1978, Volume 3 of Greenberg’s *Universals of Human Language* (1978) deals primarily with ‘word structure’. Reference should also be made to some general handbooks in the field, particularly since some of them have (usually brief) chapters/sections of morphological/derivational relevance (e.g. Whaley 1997 and Croft

2003). Two editions of *Language Typology and Syntactic Description* (1985, 2007) by Shopen start with a chapter entitled 'Typological Distinctions in Word Formation' written by Anderson (1985) and Aikhenvald (2007) respectively for each edition. The two chapters cover the same topic and clearly reflect the immense progress in the theory of word-formation over the twenty-two years which separate them. In the same volume, Comrie and Thompson (1985) discuss lexical nominalization processes in the languages of the world and Comrie (1985) maps the formation of causative verbs. Malkiel's (1978) 'Derivational Categories' and Moravcsik's (1978) 'Reduplicative Constructions' go significantly deeper into word-formation. Some of these works, like Malkiel's, had no pretence to be representative of the world languages. In Malkiel's case, this is because it limited itself to a selection of Indo-European and Semitic languages. Moravcsik, like Wiltshire and Marantz (2000), also shows a bias, this time towards one derivational process (reduplication, apparently because it is indispensable for the description of inflectional categories, especially for plural formation).

A number of authors have also focused on specific word-formation processes: Ultan (1975) undertook an important cross-linguistic analysis of infixation, Mithun (1984) studied noun incorporation, a frequent topic probably due to the controversy between the syntactic and lexicalist conceptions of noun incorporation, and L. Bauer (1996 and 1997) wrote a cross-linguistic analysis of evaluative morphology. Other similar works, like Anderson (1985) and Bybee (1985), stand out for their cross-linguistic approach to word-formation. The latter used a sample of fifty languages to examine the relation between inflectional and derivational morphology. More recently, Wälchli (2005) contributed to cross-linguistic research into co-compounds. All these are highly valuable and seminal for their scope of languages and their import, but most of them study individual word-formation processes without regard to other processes and/or categories in the languages covered.

A different type of antecedent is the research on the description of individual languages or families of languages, some of them endangered ones, as in specific series by publishers like Mouton de Gruyter and Lincom. Many of these monographs include sections about the description of word-formation processes and are an important source of data for research into word-formation universals and typology. Thus, Lüdtke (1996) wrote on word-formation tendencies in Romance languages and Werner on word-formation in the Yeniseian languages (1998). The scope of languages was wider in Kroeber (1910), who wrote on noun incorporation in American languages, Harrison (1973) on reduplication in Micronesian languages and Brill (2005) on special types of prefixes in New Caledonian and other Austronesian languages.

Typological research may focus on a specific fine-grained variable, as advocated for by Bickel (2007), or it may examine what is common and what is not between languages at a more complex level, e.g. at the level of compounding or at the level of affixation. Regarding the latter, it may admittedly

be difficult to find associations between broadly defined linguistic structures (like agglutinative or incorporating languages) and/or whole languages. Bakker (2004) provides evidence of this: his analysis of *The World Atlas of Language Structures* (Haspelmath, Dryer, Gil and Comrie 2005) concludes that less than 1 per cent of its logically possible associations are statistically relevant, and only a part of them is linguistically relevant.² However, this does not necessarily mean that there is no point in searching for associations. Typological research at all levels of language structure complexity is important and the results of each such research are complementary and contribute to a better understanding of languages of the world.

Research in the field of word-formation universals and typology also benefits from general theoretical frameworks, like Dressler's theory of word-formation within Natural Morphology (1981, 1982, 1985, 1987, 1988, 1997, 2005; and Dressler, Mayerthaler, Panagl and Wurzel 1987). Its focus on semi-otic principles and extra-linguistic conditioning of word-formation processes establishes crucial theoretical foundations for cross-linguistic research and identifies a number of universal tendencies and implicational universals in the field of word-formation.

As can be seen, typological research may take different directions and may pursue a variety of objectives: from Greenbergian large-scale research data from various languages to Chomskyan research within the tradition of Universal Grammar aiming at the generalization of a theoretical analysis of a single language. The following sections describe the approach taken in this book.

Word-Formation in the World's Languages: purpose, method and scope

Purpose

Croft (2003: 1) maintains that typology is the study of patterns that occur systematically across languages and that 'the characteristic feature of linguistic typology is cross-linguistic comparison' (Croft 2003: 13). This is also a point of departure for our research.

Nichols (2007: 261) lays emphasis on the fact that a small-sample approach (labelled *Basic Principles*) and a large-sample approach (*Phenomena Survey*) to typological research are not opposing or competing approaches: what is needed, Nichols argues, is that one methodology be able to use the results of the other. Moravcsik (2007) also highlights important parallelisms between cross-linguistic and single-language research approaches.

² Bakker, D. 2004. 'LINFER and the WAL database', paper presented at the Workshop on Interpreting Typological Distributions, December 2004, Leipzig.

For obvious reasons, all the targets set by the two approaches cannot be encompassed within a book like this. This book makes use of the *Phenomena Survey* in view of the absence of any similar typological research in the field of word-formation, but it also uses the available achievements in the area of *Basic Principles*, e.g. in the discussion of the interrelation between inflectional morphology and derivational morphology, word-formation and compounding, noun incorporation, reduplication, compounding, conversion and other topics. While the description presented here is intended to be as theory-neutral as possible, as recommended by many typologists, we show that typological research necessarily depends on how various linguistic phenomena are viewed. Plurality, incorporation, exocentric compounding and conversion are cases in point, to name just some of them.

Even such a brief review as this of the antecedents shows that this field is in no way a *tabula rasa*. What we find to be missing is comprehensive cross-linguistic research on various word-formation processes and categories (semasiological aspect) as well as the major cognitive categories (onomasiological aspect) in their interrelation by means of a representative sample of languages. Therefore, this book focuses on the identification of possible associations as a contribution to a typological classification in this area. At the same time, picking up on L. Bauer's (2000: 38–9) remark above, the book reviews wider distinctions, like the separation between inflectional and derivational morphology. This traditional separation seems to us to be justified in respect of the situation in the field of word-formation and, at a more general level, of the effort of a number of morphologists to understand and explain the relation between inflectional and derivational morphology. The traditional morphological classification has been called into question by a number of typologists, but it still has relevance, especially because no better large-scale morphological classification of languages has been accepted yet. In any case, there is no reason to exclude large-scale classifications of languages from the scope of typological research. Therefore, one of the goals of this book has also been to contribute towards the degree of similarities/differences between languages in terms of the classical morphological classification. The similarities would give support to the idea of a single morphology, while the differences would be an argument for the split morphology view.

For the purposes cited above, we rely on questionnaire-based research as well as drawing on grammars of individual languages. This method has limitations, like the availability of experts willing to fill out questionnaires for their languages and the very design of the questionnaire, which must be extensive and rigorous but still not exceed certain limits. Despite these and other limitations, questionnaires dispose of the most serious problem connected with relying on grammars of languages, i.e. 'the great diversity in individual coverage, focus of interest, and analytic approach [which] do

not make for the descriptive homogeneity which would be ideally suited to a study of this kind' (Ullmann 1978: 529).

Overall, this book deals with the following areas of linguistic typology:

- (a) associations or patterns, aiming to identify cross-linguistic interrelatedness between:
 - (i) various word-formation processes and/or categories,
 - (ii) word-formation processes/categories, on the one hand, and the genetic classification of languages present in the sample (language family) on the other,
 - (iii) word-formation processes/categories, on the one hand, and the morphological classification of languages present in the sample (inflectional type) on the other, and
 - (iv) word-formation processes/categories, on the one hand, and the prototypical word order of languages present in the sample on the other.
- (b) variation in a number of word-formation processes and categories, aiming to show the existing cross-linguistic and language-specific diversity of phenomena falling within one broadly defined category. This diversity is best describable in terms of a scale ranging from prototypical to peripheral manifestations of a particular word-formation process, category or feature. Fuzziness appears as a natural consequence of the scalar nature of linguistic facts, frequently reflected in vague boundaries between word-formation processes and categories on the one hand, and between word-formation *per se* and other levels of language on the other (e.g. inflection, syntax, phonology). What is thus characteristic for word-formation is not discrete variables, but continuous variables, as in Wälchli's (2005: 24) terminology.
- (c) descriptive complementarity, by evaluation of the above mentioned issues from the semasiological (traditional form-oriented) and the onomasiological (cognitively grounded) perspective.

Method

This book uses the terminology common to research on word-formation even if it admittedly is English-centred. Unfortunately, linguistics in general is a long way from using unified terminology. The interpretation of various terms may depend on a particular theoretical framework, on the specific conception of a linguist/informant covering the field of derivational morphology/word-formation in a particular language and also on the nature of languages covered.

Certain linguistic phenomena may also manifest themselves differently from the perspective of different languages or language types. This has given

rise to different interpretations of certain points covered by our questionnaire and, consequently, to further consultations with the informants.³

Using as a reference Song's (2007: 9) five steps of typological research (identification of a phenomenon to be investigated, generation of a language sample, creation of a typological classification, formulation of a typological generalization and explanation of the typological generalization), the method used here relies on the following foundations.

Regarding the first of these steps, the identification of the subject of study, typological studies of the Greenbergian tradition, which is also applied in this study, range from the examination of specific linguistic phenomena up to broadly outlined research into the typology and universals of a particular (sub)discipline. The focus of this book, word-formation, is defined broadly, largely because no comprehensive research where various word-formation processes and word-formation categories are interrelated has been implemented yet, but also for a number of other reasons. Word-formation established itself as an independent linguistic discipline with its own field of research and its own specific methods in the 1960s as the result of the publication of three major works: Lees (1960), Marchand (1960) and Dokulil (1962). Each developed the foundations of word-formation theory in a different paradigm pre-determined by the author's theoretical frameworks: the transformationalist account on Chomsky's transformational-generative grammar, the structuralist tradition of de Saussure and Coseriu, and the onomasiological approach in the tradition of the Prague School of Linguistics, respectively. The influence of these backgrounds favoured the rapid dissemination of theories developed in these frameworks and guaranteed the diversity and complementarity of theoretical approaches. Still, as an independent discipline, derivational morphology faced a number of theoretical issues to be dealt with, ranging from the description of word-formation almost exclusively in the most widespread Indo-European languages through a range of theoretical problems of diverse complexity (e.g. productivity, the nature of word-formation rules, different types of affixes, various word-formation processes, lexicalization, etc.). Not surprisingly, word-formation typology has been ignored for a long time. It must also be taken into consideration that a sound typological study is pre-conditioned by the availability of data. Therefore, some fifty years after the above landmarks in word-formation, the description of non-Indo-European languages

³ Cf. Haspelmath's view that there are no pre-established universal categories. Rather, 'the categories of language structure are language-particular' (2007: 121). This position rests on a fairly long tradition of the negation of the existence of universal grammatical relations and categories (e.g. Boas 1911; Lazard 1992; Dryer 1997; Croft 2001). Consequently, Haspelmath argues that linguistic typology must be 'substance-based, because substance (unlike categories) is universal' (2007: 126), and this implies for morphology that it must rest on semantics. For an opposite view, cf. Newmeyer (2007). This requirement is reflected in the onomasiological section of this book (cf. chapter 6).

is rather limited and the discussion in morphological typology has confined itself mainly to inflectional categories and processes, which were considered more or less representative for the whole of morphology. With this in mind, for a better knowledge of the intricacies of the study of word-formation and of its features, it is necessary to draw the reader's attention to those works which attempt to identify the scope of word-formation in relation to the other subfield of morphology, i.e. inflectional morphology (especially, Anderson 1982, 1992; Scalise 1988; Dressler 1989; Booij 1994, 1995, 2006; Plank 1994; van Marle 1995). The discussion about the scope of word-formation in chapter 1 is embedded in this framework.

As to the second of Song's steps, the generation of a language sample, comprehensive descriptions of the word-formation systems of the languages of the world are just not available. Sampling for typological research is here, perhaps more than in any other respect, a compromise that tries to reconcile the minimum standards of quality and quantity of sampling for objective results, with the very limited availability of descriptive sources, of experts capable of providing relevant and reliable data and of speakers of some of the endangered languages to consult the issues not described by grammars, to name only some of the limitations. As mentioned above, the method used here is based on questionnaires.

Third, regarding the creation of a typological classification, various sorts of typological classifications can be inferred from the data presented here. They are primarily determined by the specific method of analysis, semasiological or onomasiological. In particular, the typological classification pertains to the preferences for formal ways of expression of cognitive categories and for the semantic scope of the individual formal means of expression of genetically, morphologically and/or geographically related languages.

Concerning the fourth step, the formulation of a typological generalization, these generalizations may be of various natures and degrees of validity, ranging from existential typologies through statistical and implicational universals up to absolute universals (Moravcsik 2007: 28ff.). While statistical universals, extrapolated from samples, show what is possible in languages, absolute universals identify what is necessary. What is, however, important in this context and what any reader of a typological work like this should realize is that 'both probabilistic and absolute generalizations ... always remain hypothetical due to the unavoidable gap between the language sample that they are based on and the set of all human languages that they make a claim about' (Moravcsik 2007: 36). In fact, it has been pointed out that 'large databases almost invariably reveal exceptions to universals, and this, together with a substantial increase of newly described languages ... has practically done away with notions of absolute universals and impossibilities' (Bickel 2007: 245). Our analysis of results is therefore conservative to be on the safe side. In chapters 3 to 5 we identify cross-linguistic data for various word-formation categories, processes and relations. These data are

used in chapter 7 for the identification of associations by means of statistical methods (cf. 7.2).

Finally, explanations in word-formation can be sought at various levels. In our view, the universal principles of Natural Morphology and its three sub-theories (universal preferences, typological adequacy and system-dependent naturalness), are at the heart of typological explanations. The remarks contained here usually combine various explanations, e.g. about the non-existence of noun incorporation in Slavic languages and its high productivity in certain Amerindian languages, which can hardly be explained by any universal functional principle and is probably due to genetic and geographic factors. Each individual observation seems to result from the interplay of a number of functional, structural, genetic, geographic and language-specific factors.

Scope

It has already been mentioned that data of word-formation characteristics of languages are not easy to obtain in view of the lack of relevant descriptions due to almost exclusive attention of typological research to inflectional characteristics of languages outside the best-known European (and not many other) languages. Therefore, this book relies for the most part on questionnaire-based data, supported by specific bibliographical references whenever necessary to meet Nichols' (2007) requirement for a large sample survey type of typological investigation.⁴

Questionnaires have advantages and disadvantages. Their success as data collection tools depends on the availability of qualified informants, but this in turn imposes limits on the design of a well-structured cross-linguistic sample of languages. To paraphrase L. Bauer's (1997: 534) characterization of his sample on evaluative morphology, the sample used here is partly a sample of convenience, even if it is intended to comply with requirements like a certain representativeness and balance. Also, the range of the parameters within the questionnaire must be reduced to a psychological optimum to ensure the continuous cooperation of informants throughout the process of data collection and consultation.

Even so, questionnaires ensure the coherence and comparability of the data obtained, the use of common criteria for their collection and the possibility to ask morphologists and/or field researchers for further evidence or clarification. In fact, useful clarification-oriented consultations with some of our informants were frequent. These, we believe, are factors that, with a set of high-quality informants, compensate for many drawbacks. Specifically,

⁴ Data on seven languages rely exclusively on publications for want of an informant. In these cases, the decision to accept exclusively other-than-questionnaire data was based on the availability of a comprehensive description of word-formation. These languages are Diola-Fogny, Hindi, Jaqaru, Kwakw'ala, Movima, Pipil and Tzotzil.

Cambridge University Press

978-0-521-76534-3 - Word-Formation in the World's Languages: A Typological Survey

Pavol Štekauer, Salvador Valera and Lívia Körtvélyessy

Excerpt

[More information](#)

and since ‘the representativeness of samples is defined genealogically and/or areally’ (Plank 2007: 48), we use a sample of fifty-five languages representative of twenty-eight language families and forty-five language genera distributed over all continents and examined for almost seventy word-formation parameters.

Two language samples have been used: the basic sample and the study sample. The basic sample encompasses seventy languages and is used only as an illustration of the diverse manifestations of individual word-formation phenomena. The structure of the basic sample and its areal distribution are presented in Table 0.1 and Figure 0.1.

Table 0.1. *The basic sample by genetic criterion and by geographic distribution*

Genetic criterion ⁵	
Language families	28
Language genera	45
Geographic distribution	
Africa	15
America	13
Eurasia	30
South East Asia and Oceania	12

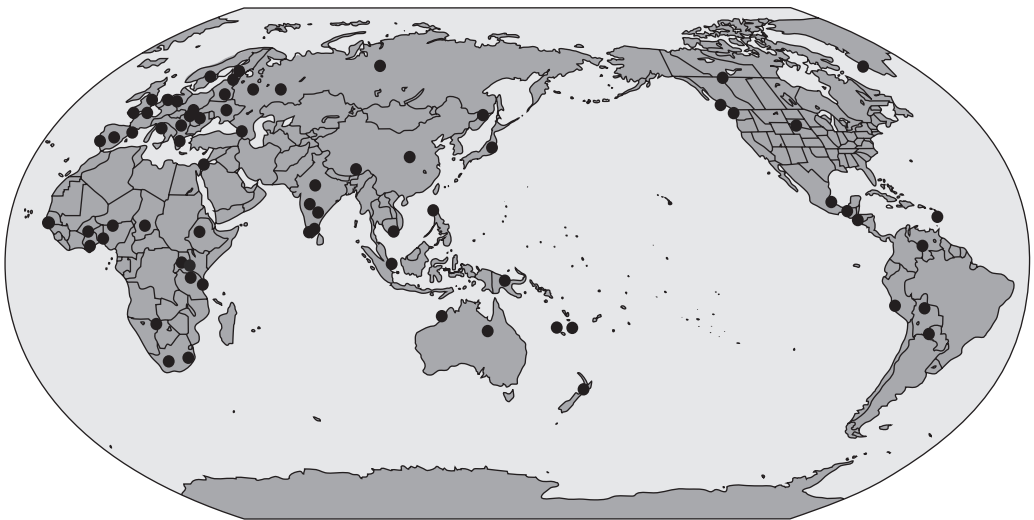


Figure 0.1. *Geographic distribution of the basic sample languages (seventy languages) as in Haspelmath, Dryer, Gil and Comrie (2005)*

⁵ Classified primarily according to Haspelmath, Dryer, Gil and Comrie (2005), and then by consultation with informants.