

## New developments in morphological typology

Bernhard Wälchli (University of Bern) waelchli AT isw unibe ch

**Morphological typology** investigates the cross-linguistic diversity in morphology. Traditional topics addressed include degree of synthesis, agglutination, morphological complexity, and the presence and extent of any kind of morphemes or morphological processes (suffixes, prefixes, infixes, ablaut, etc.). This talk will not be a survey of all modern approaches to morphological typology, but rather explores – from a subjective point of view – what kind of approaches might be promising for future research.

The word "new" in the title is not understood in the sense of starting completely anew, but rather means the combination of existing approaches, applying them to new datasets by using more efficient processing devices while not being afraid of keeping things simpler than they are to start with.

The starting point for this talk consists of three families of approaches:

- Greenberg's (1954) indices of measuring degree of synthesis, prefixing, suffixing etc. in texts, and recent work inspired by this approach (Popescu et al. 2009; Kortmann & Szendrői forthcoming).
- Modern functional typology, such as recently canonized in WALS (e.g., Dryer 2005).
- Corpus linguistics and computational approaches: statistical natural language processing, unsupervised learning of morphology (e.g., Goldsmith 2001).

By "keeping things simple" I mean defining more modest intermediate goals if the real goals are out of reach. Instead of languages we can only investigate **doculects** (documented lects of languages). For the sake of comparison it is easier to take **parallel texts** rather than more reliable original texts. Rather than having the material analyzed neatly into grammatical words and written phonologically we have to start with **conventional orthography** and a **written-language bias** is hard to avoid. A major challenge is that some **crucial issues in morphological theory** remain unresolved for many examples (inflection vs. derivation, words vs. phrases, morphemes vs. morphological processes).

I will argue that the aim of morphological typology is not to classify languages into types. Rather, Greenberg's (1954) approach is still innovative in applying measures directly to texts, thus allowing for a lower degree of data reduction than practiced in modern functional typology. Greenberg's approach, however, is very labor-intensive, presupposing a full morphological analysis of the texts considered.

The lazy linguist, however, looks for shortcuts. Are there any indirect ways of measuring morphological typology more efficiently? **Indirect measurement** is widespread in science. Simple examples include measuring age by counting annual rings in wood or by the Carbon-14 method or measuring the height of a mountain by trigonometry. In morphology a prime example is measuring productivity by counting hapax legomena (Baayen 1992 and elsewhere).

I will show that it can be useful in typology to start with unanalyzed primary data and do the analysis step-by-step by applying a universal algorithm. Every doculect is treated in exactly the same manner, but the results of the procedural universals applied are highly language-specific. Typologies are a by-product of this analytic procedure where we proceed step-by-step from wordforms (zero-order) to segments (first-order) and to more complex morphological processes (higher-order).

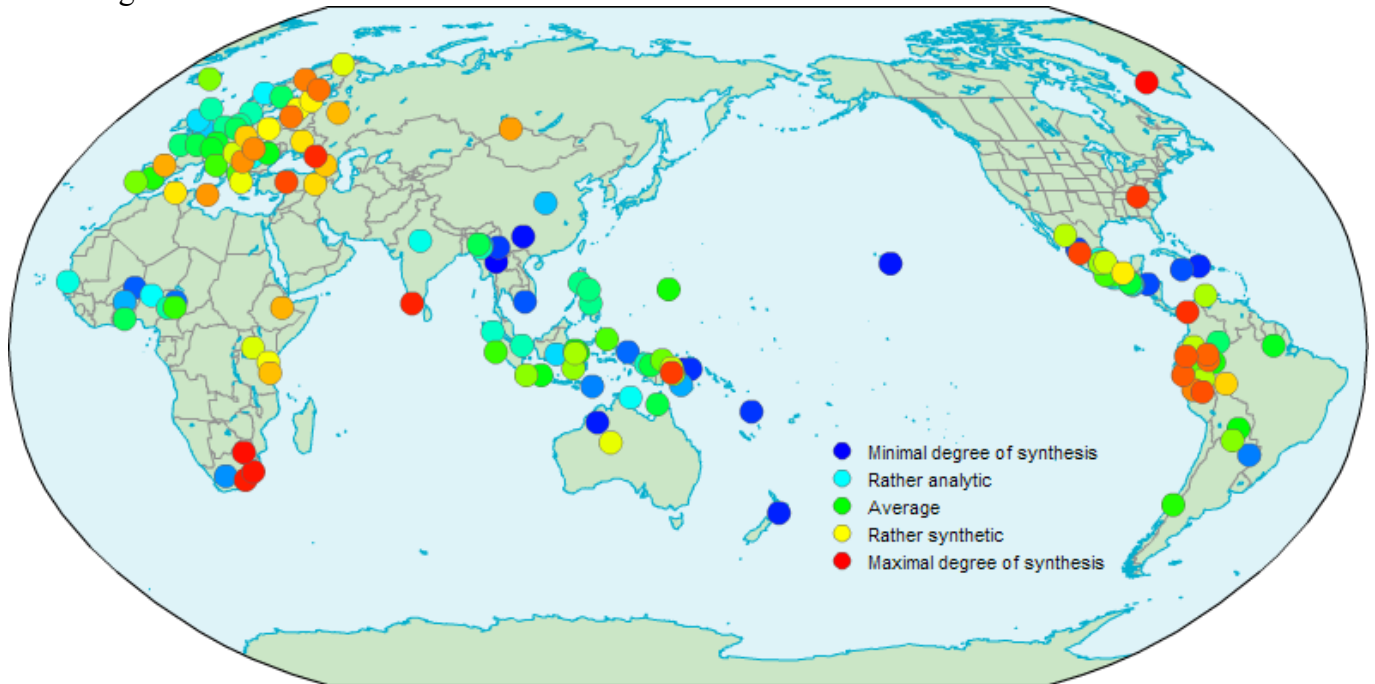
**Degree of Synthesis** can be measured indirectly already with a **Zero-Order Morphology**, a morphology that knows only unanalyzed wordforms in texts. A simple approach is to measure the type-token ratio, which is, however, useful only for texts of the same length with the same contexts. With other material, other measures derived from types and tokens can be used (Popescu et al. 2009).

For **Degree of Suffixing and Prefixing** it is not necessary to have a complete morphological analysis. **First-Order Morphology** ( $M_1$ ) is sufficient, distinguishing only "stems", prefixes and suffixes. "Stems" are simply recurrent sequences in forms of a lexical domain and prefixes and affixes are simply whatever is left if stems are subtracted. Prefixing and suffixing in inflection can be measured automatically in parallel texts with considerable accuracy with a set of language-specific lexical domains as search spaces. The results are evaluated by comparison with results in functional typology (Dryer 2005).

Finally, **Higher-Order Morphology** can be built from First-Order Morphology. For instance,  $M_1$  wrongly detects a Tagalog stem *asok* in the set of forms for 'enter' [*pum-asok*, *pap-asok*, *pin-asok*, *pagkap-asok...*] instead of *p(-)asok*. A method to identify discontinuous stems is to look for consonants occurring in all  $M_1$ -prefixes or -suffixes in the set. Once discontinuous stems have been identified it is possible to search for infixes (whatever is in-between).

Indirect methods of measuring morphological typology will be exemplified in a sample of 150 doculects from forty linguistic stocks. Map 1 illustrates an example from  $M_0$ : degree of synthesis.

Map 1: Degree of synthesis: doculects displayed in ranking order with equal distance across a rainbow-color space. Indirectly measured by means of a type/token-derived measure in translations of the Gospel according to Mark



#### References:

- Baayen, R. Harald 1992. Quantitative aspects of morphological productivity. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 1991*. Dordrecht: Kluwer. 109-149.
- Dryer, Matthew 2005. Prefixing vs. suffixing in inflectional morphology. In Haspelmath, Martin & Dryer, Matthew & Gil, David & Comrie, Bernard (eds.) 2005. *The World Atlas of Language Structures*. (Book with interactive CD-ROM), Chapter 26. Oxford: Oxford University Press.
- Goldsmith, John A. 2001. Unsupervised learning of the morphology of a natural language. *Computational Linguistics*, 27 (2): 153-198.
- Greenberg, Joseph H. 1960 [1954]. A quantitative approach to the morphological typology of languages. *International Journal of American Linguistics* 26: 178-194. First published in Robert F. Spencer, Fs. for Wilson D. Wallis. *Method and Perspective in Anthropology*. University of Minnesota Press.
- Kortmann, Bernd & Szmrecsanyi, Benedikt. To appear. Parameters of morphosyntactic variation in World Englishes: prospects and limitations of searching for universals". In: Siemund, Peter (ed.), *Linguistic Universals and Language Variation*. Berlin: Mouton de Gruyter.
- Popescu, Ioan-Iovitz & Ján Mačutek & Gabriel Altmann. 2009. *Aspects of Word Frequencies*. Lüdenscheid: RAM.