Frequency profiles as a tool for tracing the interaction between borrowing and word formation

It is often difficult to determine whether complex words are the result of word formation or borrowing. In many European languages, there are complex words with a Latin or Greek origin, for which a borrowing from classical languages is not possible, because the words are not attested in the classical stages of these languages. Examples include formations in *-ation* and its variants in several languages. These nouns are often analysed as the result of a word formation rule taking a corresponding verb as its input. In German, we find triples such as (1).

(1)   a. Lokalisation   ('localization')
      b. lokalisieren   ('localize')
      c. Lokalisierung   ('localization')

For (1a), the question is whether it is a word formation result based on (1b) or a borrowing. For (1b), the verb can either have been borrowed or be a back formation based on (1a). In German, French is the main candidate as a source of borrowings of this type. In the case of verbs such as (1b), the ending *-ieren*, which serves as a condition for the nominalization in *-ation*, corresponds to the French infinitive. The noun (1c) is an unambiguous word formation result, derived from (1b).

Frequency profiles are a tool for determining the likelihood of the different scenarios in constellations such as (1). The question is not which words in a particular case were borrowed or resulted from word formation, but how this generalizes for a process. The underlying assumption is that the base word used in a word formation rule is usually more frequent than the derived word. By looking at large numbers of similar words, the outliers where this assumption does not hold will be outnumbered by the regular cases. The degree to which this effect takes place tells us something about the relative strength of the borrowing and the word formation hypotheses for a particular word formation constellation. In the case illustrated in (1), the questions are whether nouns in *-ation* are rather formed from verbs or borrowed as nouns and whether verbs in *-ieren* should rather be considered borrowings or back formations.

In order to interpret frequency data, we need to calibrate the measures with respect to unambiguous cases. For this calibration, we can use the Slovak counterparts, as in (2).

(2)   a. lokalizácia   ('localization')
      b. lokalizovať   ('localize')
      c. lokalizovanie   ('localization')

In Slovak, the only source for the base of (2a) is a borrowing. In this case, Latin is the most likely source. This means that (2b) must be a backformation based on (2a). The alternative noun in (2c) is derived from the verb in (2b). Comparing the frequency profiles of German triples such as (1) and Slovak triples such as (2) will then provide information about the likelihood that nouns in *-ation* are borrowed or result from word formation.

Frequency profiles are the characteristic distribution of frequency between base words and derivatives in a large corpus. They can be used to determine the likelihood that borrowing or word formation is at the origin of a lexical relationship. By comparing frequency profiles for controversial cases with those for cases where there is no doubt about the relationship, we can determine the most likely analysis for the former. We applied this to German nouns in *-ation* and Slovak nouns in *-ácia*. Further research will be necessary to calibrate the frequency profiles and compare them for a larger set of cross-linguistic data.