

So what would a native speaker do?
From native-speaker preference to (different) learner preferences

Stefan Th. Gries
University of California, Santa Barbara

One primary question of SLA/FLA research is how learners' usage of lexical items and syntactic patterns/constructions differs from that of native speakers. Thus, much research in usage-based approaches to SLA/FLA has been done on how (i) learners' over-/underuse of linguistic elements contributes to the non-nativeness of learner language and (ii) learners from different L1s differ in their use of one and the same target L2/FL. For some time now, researchers working on native language have been exploring speakers' subconscious choices of one kind of expression over another on the basis of multifactorial methods applied to comprehensively annotated corpus data (see Gries 2000 for one of the first studies of this kind in cognitive linguistics). However, compared to how widespread this analytical approach is now in especially cognitive- and psycholinguistic studies of alternations in native speech and writing, this approach is yet underutilized/-developed in SLA/FLA studies.

The present paper exemplifies the potential of this approach by presenting two applications: As for the first, I exemplify what the classic kind of multifactorial regression offers to SLA/FLA studies using the example of how Chinese and German learners of English behave both similarly and differently from native speakers of English in a word order choice, namely the order of two prenominal adjectives. For instance, learners and native speakers exhibit the effects of noun-specific frequency (compatible with a usage-based approach) or semantic openness, but differ with regard to the impact of the adjectives' affective load or the phonological effect of segment alternation.

As for the second, I discuss how a new method (cf. Gries & Wulff 2012 and Gries & Deshors, submitted) involving the successive application of separate multifactorial regression models helps study differences between native and learner data in more detail than is customary. Specifically, I introduce a new regression-analytic approach to identify how learners' usage is different from what that of native speakers in the same linguistic/situational context on the basis of a lexical choice, namely can vs. may by native speakers and French learners of English. The method involves four steps:

1. the comprehensive annotation of relevant instances of the lexical/ordering choice in native and learner data;
2. a regression analysis M that is based only on the native data to identify the factors governing the choices of native speakers;
3. the application of M to learner data which predicts what a native speaker would do if he were placed in the same linguistic/contextual situation as the learner;
4. the comparison of what the native speakers would have done to what the learners actually did.

The results of this method are manifold and both support and complement the more traditional type of analysis. In particular, they reveal where learners behave as native speakers would (e.g., can in neutral aspect or with accomplishment/achievement verbs), where learners exhibit the greatest mismatches/difficulties (e.g., may with animate subjects or process verbs and can with perfective/progressive aspect), and they in particular highlight differences in the learners' usage related to factors involving frequency effects and processing.