

Error-tagging the German component of LINDSEI: Methodology and application

Susanne Kämmerer & Sandra Götz

For over a decade now, learner corpus research has been combining theoretical approaches from both Second Language Acquisition research (contrastive analysis, error analysis) and corpus linguistics (quantificational approach to data and mechanisms of analysis) (cf. Díaz-Negrillo & Fernández-Domínguez 2006), and this "[has] given rise to a powerful apparatus for quantitative and qualitative study of foreign language learning" (Díaz-Negrillo & Fernández-Domínguez 2006:85). While raw corpora can be used to investigate learner overuse and underuse very easily, only an error-tagged version of the data makes quantitative and exhaustive investigations into the area of errors possible (cf. Kämmerer 2009).

This paper is divided into the following sections. First of all, we will attempt to review the Louvain error-tagging system, which is a flat-level annotation system designed for the annotation of written learner language. We will therefore focus on problems, decisions and adaptations concerning the tag set to cater for the needs of error-tagging spoken learner language (e.g. tags for pronunciation errors, inconsistency in use of variety, self-correction, repetition of errors, disfluencies). We will also treat general tagging problems (e.g. tags for Double Tags, Trebble Tags), as well as additional tags focusing on a learner-centred identification, correction and explanation of errors (tags for interference-induced errors, performance vs. competence). The second section will report on a pilot study on the most common errors in the spontaneous speech of advanced learners. Overall, our pilot study not only reveals constructions that are especially error-prone in the case of German learners of English (e.g. prepositional phrase complements, adverb-adjective selection), but it also shows that there is still a clear systematicity regarding the error types even at an advanced level. Also, specific hesitation phenomena, such as filled or unfilled pauses, tend to co-occur only with certain errors in spontaneous spoken language and not with others (cf. Brand & Götz forthcoming). Our final step will be to present language-pedagogical implications derived from our findings.

References:

- Brand, C. & Götz, S. forthcoming. "Fluency versus accuracy in advanced spoken learner language: A multi-method approach", *International Journal of Corpus Linguistics* 16, (2).
- Díaz-Negrillo A. & J. Fernández-Domínguez. 2006. "Error tagging systems for learner corpora". *RESLA* 19, 83-102.
- Kämmerer, S. 2009. "Error-tagging spoken features of (learner) language: The UCL Error Editor 'revised'". Paper presented at the 30th annual conference of the *International Computer Archive of Modern and Medieval English (ICAME30)*. Lancaster, UK.

Susanne Kämmerer, Justus Liebig University Giessen (Germany)

Susanne.Kaemmerer@anglistik.uni-giessen.de

&

Sandra Götz, Justus Liebig University Giessen & Macquarie University Sydney (Australia)

Sandra.Goetz@anglistik.uni-giessen.de