

THE *HOWS* AND *WHYS* OF CODING CATEGORIES IN A LEARNER CORPUS
(OR “HOW AND WHY AN ERROR TAGGED LEARNER CORPUS IS NOT
IPSO FACTO ONE BIG COMPARATIVE FALLACY”)

KARI TENFJORD*, JON ERIK HAGEN**, HILDE JOHANSEN***
University of Bergen

Abstract

A new electronic learner corpus of annotated language test essays of Norwegian as a Second Language developed at The University of Bergen is presented in outline, and in this connection the question whether the practice of error recording and error coding in itself is theoretically misguided by virtue of the so-called “comparative fallacy” argument (Bley-Vroman 1983) is discussed. Error tagging is defended on the basis of a broader discussion of relations between theory and data in SLA generally, and the overall conclusion is that error recording and error coding is not methodologically misguided since error analysis is not a theory of SLA but rather a method, a method that can, in principle, service any theory. We think that the so-called “comparative fallacy” charge often levelled against classical error analysis or any preoccupation with errors at all stems from failure to distinguish sufficiently between these two notions within SLA.

Keywords: learner corpus, comparative fallacy, error, error tagging, SLA methodology

INTRODUCTION

This article will present an electronic corpus of written test essays produced by learners of Norwegian as a second language known as the “ASK corpus”¹ (“ASK” being acronymic for the three constituent morphemes of Norwegian “andrespråkskorpus”), its design and scientific potential. The main aim of constructing this corpus was to enhance the facilities for empirical studies on the acquisition of Norwegian as a second language and perhaps SLA studies more generally. The corpus will provide a database not only of the essay texts themselves, but also annotations of various kinds such as parts of speech, morphological categories and error tags. These resources in combination with an efficient user interface system, make it possible to test hypotheses generated by previous studies in Norwegian as a second language and it may also be a rich source for explorative

¹ For a presentation of ASK visit : <http://spraktek.aksis.uib.no/projects/ask>

*Associate professor Kari Tenfjord, **associate professor Jon Erik Hagen, ***graduate student Hilde Johansen. We are all affiliated to the Dept. of Scandinavian Languages and Literature, University of Bergen.