Selecting quantitative data for qualitative analysis: A case study connecting a lexicogrammatical pattern to rhetorical moves

Annelie Ädel[*]

Dalarna University, Sweden

ARTICLE INFO

ABSTRACT

Learner corpus research involves studying large collections of data to achieve a certain degree of representativeness, which means that it is often not doable to examine a full set of data qualitatively. An important issue, then, is how to select a subset for further qualitative analysis. This study illustrates a selection method, taking quantitative results as a starting-point, for a qualitative study of a lexicogrammatical pattern. Three configurations are examined, involving not only statistically significant differences (overuse and underuse), but also similarities (equal use). What is studied is the anticipatory *it* pattern ("*It is however important to interpret* these findings with caution") in apprentice writing in linguistics by learners and native speakers of English. The method yielded 463 tokens in 62 learner and 82 native-speaker essays. The research questions were (i) What are the connections between the selected subpatterns of anticipatory *it* and specific rhetorical moves? and (ii) Are there indications of learner behaviour in the connections between subpatterns and rhetorical moves? Most subpatterns were found to be specialised for a few moves. The two groups mostly used the subpatterns for the same rhetorical work, but the learners used *important* and *clear* subpatterns for a greater range of moves.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Research into second language writing is increasingly informed by learner corpus data. In the past decade, we have seen corpus-based studies on a range of phenomena, such as phraseology (e.g. Nesselhauf, 2005) and information structure (e.g. Callies, 2009). Learner corpora have become popular research tools thanks to their being based on naturally occurring language and their potential representativeness: they typically "contain data from hundreds (sometimes thousands) of learners and can therefore lay claim to greater representativeness than previous SLA studies" (Granger, 2009: 16). The development of learner corpus research is especially owed to a pioneering project which resulted in the *International Corpus of Learner English* (ICLE; Granger, Dagneaux, Meunier, & Paquot, 2009), a corpus consisting of argumentative essays by advanced learners of English from a range of different L1 backgrounds. Within the project, a systematic approach—Contrastive Interlanguage Analysis (CIA)—was adopted for the comparison of learner and native production of the same language. CIA typically also involves comparison between different learner populations, with different L2s, in order to study interlanguage features.

* Department of English, School of Languages and Media Studies, Dalarna University, 791 88 Falun, Sweden. Tel.: +46 701918616.
  E-mail address: annelie.adel@du.se.

The aim of CIA is to "highlight a range of features of non-nativeness in learner writing […], i.e. not only errors but also instances of under- and overrepresentation of words, phrases and structures" (Granger, 2002: 12) by comparison to native-speaker data. With this approach, the notions of 'underuse' and 'overuse' have become key in learner language research.

Although CIA was originally conceived of as both a quantitative and qualitative approach to learner language (cf. Granger, 2009: 18), research taking this approach—as learner corpus research in general—has tended to be overly oriented to quantification. In an attempt to counter this tendency, the current study focuses on qualitative analysis, and it brings in the discourse level in doing so. In taking quantitative data as a point of departure for a contextualised analysis of the rhetorical work carried out by a linguistic pattern, the present study illustrates a method for how to select a subset of a pattern for closer qualitative analysis. In principle, the method can be used in any kind of combined quantitative and qualitative study, regardless of population(s) studied. But learner status happens to be the variable of primary interest here; the study investigates university student writing in linguistics, comparing a group of advanced learners to a group of native speakers of English. It is assumed that the frequency of the linguistic phenomenon studied is too high to make it feasible to analyse the complete set manually; this then requires criteria for selecting a subset.

In the current study, then, the notions of overuse and underuse are considered not merely from the perspective of frequency, but the data are also contextually analysed to see whether the advanced learner and native-speaker groups examined differ in their actual usage of the linguistic phenomenon studied. In fact, in the interest of looking beyond (exceptional) frequencies, tokens of 'equal use' are also considered. While instances of overuse and underuse clearly signal difference—albeit based on frequency alone—the justification for studying equal use is perhaps not quite as apparent. However, there is a case for considering also equal use, in that a linguistic phenomenon that is drawn on equally often by different groups may still not necessarily be used the same way. Now that learner corpora have become established research tools, we are in a better position to avoid the assumption that 'equal use' amounts to 'what learners get right'. It is only through qualitative analysis that the true degree of equivalence can be determined. In other words, the useful notions of 'overuse' and 'underuse' tell us whether a linguistic phenomenon occurs more or less often in corpus X than in corpus Y, but not whether the phenomenon is used similarly or not in corpus X and corpus Y. Learner corpus studies sometimes also refer to the notion of 'misuse', thereby evaluating the use of a given phenomenon from the perspective of grammatical correctness. But learner corpus research also needs to consider levels above grammatical correctness, relating to important concepts such as communicative competence and rhetorical effectiveness. This is of particular relevance in the case of advanced learners.

The pattern used here to illustrate the method is the anticipatory *it* pattern (as in "*It is however important to interpret* these findings with caution"). Previous research into this pattern has provided quantitative results for a range of different material and populations, some of which is summarised in Section 2.3 below. Previous corpus work comparing learner and native-speaker data, specifically, has shown that, taken as a whole, the pattern is underused by EFL advanced learner groups. In the qualitative analysis, my main interest is in exploring what the lexicogrammatical pattern contributes at a rhetorical level. For this purpose, the notion of 'rhetorical moves' is drawn on to analyse the use of the pattern. The research questions were (i) "What are the connections, if any, between the selected subpatterns of the anticipatory *it* tokens and specific rhetorical moves?" and (ii) "Are there indications of learner behaviour in the connections between subpatterns and rhetorical moves?". In the results section, I first look at the overall connections between subpatterns of anticipatory *it* and rhetorical moves (Section 5.1) and then go on to consider differences between the native-speaker and learner texts (Section 5.2).

Before the results are presented, the forms and functions of anticipatory *it* patterns are introduced (Sections 2.1 and 2.2), followed by a brief overview of previous findings of its use in advanced learner writing specifically (Section 2.3). Then follow a summary of rhetorical moves (Section 3) and a presentation of the material and method used for the study (Section 4).

2. Anticipatory *it* patterns

This section first briefly describes the form of the anticipatory *it* pattern, next its functions as described in the literature, followed by the major findings regarding anticipatory *it* patterns in learner writing specifically. Finally, there is a brief summary of the quantitative findings used as a starting-point for the present study.

2.1. The forms of anticipatory it patterns

The anticipatory *it* pattern involves a type of extraposition—i.e. "postponement which involves the replacement of the postponed element by a substitute form"(Quirk, Greenbaum, Leech, & Svartvik, 1985: 1391)—of a subject. The subject position is filled by the anticipatory pronoun *it*. The extraposed clause is finite—usually a *that*-clause—or non-finite—usually a *to*-infinitive—as in (1) and (2).

(1) It is obvious that these dictionaries prioritize different things. (NNS-038)[1]

(2) It is essential to take the concept cognitive domain into account. (NNS-027)

---

[1] The prefix 'NNS' refers to examples retrieved from the learner material, and 'NS' the native material.