A longitudinal corpus-based study of *it*-extraposition constructions in Italian EFL academic writing
Erik Castello (University of Padua, Italy)

*It*-extraposition constructions have been researched extensively in various fields, including corpus linguistics (e.g. Kaltenböck 2000, 2003; Groom 2005; Thompson 2009), learner-corpus research (e.g. Römer 2009; Hasselgård 2009; Herriman & Boström Aronsson 2009), and corpus-based research aiming at providing fine-grained descriptions of lexico-grammatical patterns for English Language Teaching (e.g. Francis, Hunston & Manning 1996, 1998). These constructions are of the utmost importance to written academic English and to the field of English as a Foreign Language (EFL), as they enable writers to objectively and impersonally make statements or express opinions and ultimately develop a successful academic persona (e.g. Hewings & Hewings 2001, 2002; Larsson 2016). They can, however, pose a series of challenges to EFL learners (e.g. Hinkel 2013), which can persist even at an advanced level of linguistic proficiency, at least in the case of Italian learners (e.g. Castello 2015).

Learner-corpus research on *it*-extraposition has mainly investigated cross-sectional data, while longitudinal studies are less common. This paper attempts to fill this gap by exploring the use of *it*-extraposition constructions in academic texts written by two cohorts of Italian undergraduate language students at the University of Padua (Italy), each followed over a period of two academic years, respectively from year 1 to year 2 (138 students) and from year 2 to year 3 (73 students). This longitudinal data (189,051 words in all) is part of the Italian component of the Longitudinal Database of Learner English (LONGDALE)\(^1\). It is compared to a sub-corpus of the Louvain Corpus of Native English Essays (LOCNESS)\(^2\) (35,399 words) and to another one from the Italian component of the International Corpus of Learner English (ICLE_IT)\(^3\) (Granger, Dagneaux & Meunier 2002) (107,204 words).

The study involved the manual inspection of the concordance lines for all the instances of *it*-extraposition constructions in the corpora, which were retrieved by means of the corpus query system The Sketch Engine (Kilgariff et al. 2014). Following Francis, Hunston & Manning (1996, 1998), the constructions were grouped according to the "pattern group" they belong to, including incorrect patterns. The following are the most recurrent patterns accompanied by examples:

1. *it* (mod) *V* (adv) adj: e.g. For a language learner *it is almost impossible* to completely master a second language.
2. *is* (adv) adj: e.g. I think *is absolutely possible* being able to speak a language like a native speaker.
3. *it* (mod) *be* n/prep: e.g. Perhaps *it would be a good thing* if people would learn it at the beginning of the process.
4. *it* (mod) *be* V-ed: *It can be claimed* that going abroad is a successful way to know and get in touch with foreign cultures.
5. *it* (mod) *V*: e.g. Moreover, *it seems that* the endangered areas are more frequently run by foreign tourism companies.

---

Subsequently, a complementary analysis was conducted into the typologies of extraposed constituents in the corpora: to-clauses, that-clauses, for/to clauses, wh-clauses, ing-clauses and erroneously *extraposed noun phrases. The following is an example of an extraposed noun phrase:

(6) *It should also be obligatory a standard set of guidelines and a formal system of accreditation.

The analysis revealed that the instances of the “it (mod) be V-ed” pattern and those of extraposed that-clauses increased from year 1 to year 3 in the LONGDALE-IT data, while those of the “it (mod) be (adv) adj” pattern and of extraposed to-clauses decreased. The paper discusses these and other related results.

The last part of the study focused specifically on the erroneous uses of it-extraposition constructions with a view to exploring the causes that bring them about. The case of the mistaken extraposition of noun phrases, in particular, was explored not only considering the role of L1 interference but also the lack of successful interplay of lexical, syntactic and discourse resources (Lozano & Mendikoetxea 2010). In order to conduct this part of the analysis, the 52 instances of extraposition of noun phrases in LONGDALE-IT and in ICLE_IT were lumped together and divided into two groups according to whether they belong to the “it (mod) be (adv) adj” pattern or to the “it (mod) be V-ed” pattern. The syntactic complexity of the sentences in which noun phrases are erroneously extraposed was measured in terms of the number of words they contain (“length of sentences”) and of the number of main and secondary clauses in them (“ranking clauses/sentence”). The complexity of the extraposed noun phrases was quantified by counting the words they consist of (“length of noun phrases”). The Mann-Whitney U test was performed to check for the statistical significance of the differences between the two groups of patterns with regard to the three variables. It turned out that the instances of the “it (mod) be V-ed” pattern with extraposed noun phrases tend to be used in longer sentences and that the differences between the values for “sentence length” in the two groups are statistically significant. These findings support the hypothesis that the combined use of it-extraposition and the impersonal passive voice - both of which are inherent features of the “it (mod) be V-ed” pattern - in long sentences is the most likely cause for the erroneous extraposition of noun phrases in the Italian data.

The paper discusses these findings as well as their implications for EFL teaching and learning.

References


