

## Chapter 7. Appendix 1. Definitions to Key Terminology

The terms referred to in the analyses are corpus, concordance, word and lemma, function words<sup>1</sup> and lexical words, token and type, frequency, and collocation

A corpus is a systematic collection of textual or discursal material that has been accumulated for a specific purpose (Knowles 1996: 5). For this study, the corpus referred and analysed is that of a compilation of mythological tales from the Literary Readers and some History books,<sup>2</sup> sourced from Irish National School textbooks, from 1924 to 2001. As explained in Chapter 5, the English Literary Reader was the core classroom text in Irish National Schools up to and beyond the introduction of the New Primary School Curriculum in 1971 and thus received most usage in the school learning environment. Also a small percentage of the target tales are drawn from History books.

A concordance is an index to the words in a text that gives access to many important language patterns in texts (Sinclair 1991: 170). It is a complete and exhaustive listing of all the words in a corpus; in this study it is a complete compilation of all the salient tales sourced from the schoolbooks. A concordance is accompanied by the words' immediate contexts and is also normally alphabetically ordered (Knowles 1996: 20). The computer-generated concordance can be very flexible as the context of a word can be selected on various criteria, for example providing information in relation to the number and type of words on either side of the word being examined (Sinclair 1991: 171). Most specifically, the fundamental idea in utilising a concordance is "to direct .. [one's] .. attention to the immediate linguistic environment of the specified word".<sup>3</sup> One of the most widely used forms of concordancing is that of Key Word in Context (KWIC).

A word is a lexical item that can stand alone in meaning, that can act as a sentence constituent and that can also directly correlate with an object of thought as a general reflection of reality or imagination (Knowles 1996: 21). In this study and in the descriptions and discussions arising from my analyses, the term 'word' is used in the sense of its dictionary form, its headword, also referred to as its lemma. I make no morphological, phonological or orthographic distinctions in my analyses, each word is recognised individually for its semantic properties in the texts. I therefore am not encompassing the family of variations that exist around each word, known in lexicography as the lexemes or inflected forms (e.g. 1. lemma = go; inflected forms = goes, going, went, gone. 2. lemma = great; lexeme = greater, greatest). So regardless of its inflection, in this study I take each lexical representation at its own face value. For example in the verb frequency list, 'said' is found to be the most frequent item – I do not have this word grouped under its headword, its lemma 'say'; so 'say' and 'said' are thus recognised and listed as separate frequency items in the concordances I conducted. Likewise in the common type frequency list, 'great' is highlighted and discussed for its own properties and contributions, as opposed to grouping it together with 'greatest' or 'greater.' Knowles points out that the degree of distributional identity bonding between the members of a lemma set may well be quite surprisingly low and these members often have quite independent existences in terms of their semantics (ibid: 8).

Function words are a closed set of words used to express the structure of an utterance or statement rather than to express the meaning or message delivered in the text.<sup>4</sup> These words (e.g. pronouns, determiners, modal verbs, conjunctions) play a major part in English grammar, they provide grammatical information such as quantity, tense and modality and are thus also referred to as

<sup>1</sup> Chapter 7. Appendix 2. Sample List of Function Words.

<sup>2</sup> Approximately 10% of the corpus derives from History school books. See Chapter 7 Appendix 3 and Appendix 6.

<sup>3</sup> Kings College London Digital Humanities : 5. <http://www.cch.kcl.ac.uk/legacy/teaching/av1000/textanalysis/method.html>

<sup>4</sup> Chapter 7. Appendix 2. Sample List of Function Words.

‘grammatical words’ or ‘structure words’ (Greenbaum & Quirke 1990: 16). They are words that are empty of any lexical meaning; they carry no message in the expository text. On the other hand, ‘lexical words’ are words that infuse a text with meaning(s) and that carry the message(s) in the text. They are considered open classes of words and comprise of nouns, adjectives, full or lexical verbs and adverbs (ibid). The analyses conducted for this thesis all centre on lexical word content in the target texts.

In the generic descriptions when discussing corpora contents and findings in lexical studies, a token is any instance of a word in running text. For example, a text length can be expressed as having 7,234 tokens. In my analyses, tokens are counted as each headword – so ‘old’ and ‘oldest’ are counted individually, and ‘man’ and ‘men’ are counted individually. In corpus analysis, when the same word (token) appears more than once it is then referred to as ‘a type’. The example “the cat sat on the mat” contains six (6) tokens, but five (5) types (Knowles 1996: 7 & 21). For this study the term ‘word’ will be used interchangeably with the term ‘type’ to carry the same meaning, while ‘token,’ when referred to, will hold the generic meaning presented above, any word in running text.

The term frequency refers to the actual occurrence number in the text in question of any particular word, which is then referred to as a token. This can also be referred to as its ‘*absolute frequency*’.

Generically in the English language, collocation is the regular co-occurrence of certain lexical patterns<sup>5</sup> that is habitual and oftentimes appears to have neither rhyme nor reason. The most common examples are idioms and fixed expressions (she *takes the biscuit*, he’s *a pain in the neck*, I’m *over the moon*, he *got the wrong end of the stick*). In corpora, collocation is the occurrence of two or more words within a short space of each other in a text where the usual measure of proximity is a maximum of four words intervening (Sinclair 1991: 170). Collocation of particular words can help to highlight the semantic message a specific word may carry in a text. Such collocational environments evidenced in this study have helped to highlight the specific self-image and uniqueness the high frequency words carry in the mythology tales that make up this corpus.

---

<sup>5</sup> Examples of collocation: adjective + noun, noun + noun, verb + adverb, adverb + adjective, adjective + adjective.