Tolkien as a stylist: a corpus-based investigation into The Lord of the Rings

Dirk Siepmann (University of Osnabrück, Germany)

Opinions on the style of The Lord of the Rings differ widely, ranging from scathing criticisms to the effect that "prose and verse are on the same level of professorial amateurishness" (Wilson, 1995, p. 59) to praise of Tolkien's ordinary yet dignified diction (Kirk, 1977, Rosebury, 2003, p. 22).

The present study takes a more objective, corpus-stylistic approach to Tolkien's masterpiece, asking whether it displays any stylistic features that are either significantly more common than in other texts or even distinctively its own. Among the characteristics considered are word frequency, key words, key collocations, key semantic areas and key n-grams.

Two separate methodologies were used for computing key words and key collocations, respectively. With regard to key words, a detailed comparison was undertaken between key words in 19th century fiction, late 20th century fiction and the Lord of the Rings, as compared to a 'mixed' reference corpus. The 19th century corpus was about 30 million words in size, comprising mainly canonical writers such as Jane Austen, Thomas Hardy, Charlotte and Emily Brontë, George Eliot, Elizabeth Gaskell, George Meredith, etc., with whose writings Tolkien was almost certainly acquainted. The post-war fiction corpus comprised around 160 million running words and was fairly evenly balanced between eight major genres (belles lettres, general fiction, science fiction, adventure, romance, mystery and detective, fantasy, and children's fiction). Key words were computed using the Key words tool of WordSmith 5.0 (Scott, 2008). The tool settings were as follows: minimum frequency = 3, maximum key words = 7,500; p value = 0.000001; procedure = log likelihood, meaning that a word, or rather a word form, was considered to be key if a) it occurred at least three times in the text and b) turned out to be outstandingly frequent when compared to the reference corpus by means of Dunning's (1993) log likelihood function. Following the findings of Scott (2009), which suggest that an optimum reference corpus should be both large and heterogeneous, a reference corpus of 94 million words was collected, comprising volumes II, III, IV and IX of the Cambridge History of English and American Literature (1 m), a wide range of academic texts from the National Academy Press (10 m), a general encyclopedia (10 m) and a medical encyclopedia (1 m), the Times on CD-ROM 1995 (30 m), the Cheddar Valley Gazette and the Brentwood Gazette 2006 (2 m), excerpts from Hansard (the verbatim report of proceedings in the House of Commons and the House of Lords) (35 m), and a selection of transcripts from American talk shows and news programs (5 m). In addition, Wmatrix (Rayson, 2008) was used to identify key semantic fields.

Key collocations were computed by means of a comparison between two lists, one containing collocations specific to Lord of the Rings by comparison with the imaginative section of the BNC, the other containing collocations specific to Lord of the Rings by comparison with a 30-million word corpus of nineteenth-century novels. First, the individual lists were compiled using the Sketch Engine's word list function ('collocation' setting). Next, collocations that appeared in either of the two reference

corpora and Lord of the Rings were eliminated from the lists, so that only collocations specific to Lord of the Rings remained.

A number of interesting findings emerge from the key word analysis. Most importantly perhaps, it is evident that the impression of archaicity which any reader will experience on reading Lord of the Rings is partly due to three simple lexical causes: the 'overuse' of words borrowed from nineteenth-century fiction, the avoidance of words associated with the modern world, and the comparatively dense use of new coinages, unusual grammatical patterns, rare or obsolescent words.

What is particularly noteworthy with respect to key words is the dense presence of nineteenth-century words from four lexical sets: (1) landscape description; (2) position or direction; (3) warfare and the struggle between good and evil; (4) other descriptive verbs and adjectives. Among the highest scoring 200 key words shared with nineteenth-century novels we find a large number of terms describing features of the landscape as well as position or direction, such as mountain(s), hills, hilltop, mountain-side, topmost, bough, hiding-place; road, river, cross-roads, passage, winding; above, westward, northward, southward, near, under, whither, thither, thence, whence, upwards, yonder, afar; journey (v), hewn, descend. While it might be argued that these are reasonably to-be-expected characteristics of a novel revolving around a prolonged journey, it is still significant that so many of the keywords used to describe its settings were obsolete or obsolescent at the time of writing. Another interesting finding is that Tolkien borrows a host of words from his translation of Beowulf (e.g. valour, valiant, corslet), stretching them creatively to cover a wider range of uses than was possible in Old or Middle English (e.g. Branched lightning smote down upon the eastward hills.)

As for the modern key words found in Lord of the Rings, it is worth noting that these are usually part of the core vocabulary of present-day English, such as they, big, wake, wall, unfriendly, tired or smoke. However, Tolkien tends to use many of these words in nineteenth-century patterns. One example among many is smoke, which is commonly preceded by the indefinite article in Lord of the Rings (e.g. … a black smoke swirled in the air.), a use that is notably infrequent in modern fiction. One area in which Tolkien tends to follow modernist tendencies is his abundant use of descriptive verbs (i.e. verbs which apart from denoting an event contain an additional semantic feature that assumes the function of a manner adverb, (cf. Snell-Hornby, 1983, p. 25) and, to a lesser extent, nouns (e.g. haze, stench). Among the key verbs shared with modern fiction are the following: clutch, cower, crack, crawl, curl, flicker, frown, gape, gasp, gaze, gleam, glimmer, glint, glitter, growl, hiss, huddle, loom, paw, peer, pile, plod, reek, shroud, slant, stumble, stow, stride, stumble, twist.

Turning now to those key words that are exclusive to Lord of the Rings (i.e. non-key in 19th and 20th century fiction) and hence particularly distinctive of Tolkien's style, we find that these, too, fall into a number of relatively distinct semantic categories: warfare (e.g. armies, armouries, besieged, destroyed), features of the natural environment (e.g. branches, coasts, dike, downs, fen, fern, dingle, glades, land, starless, starlight, thickets; hewed, hewn, encircling), artefacts made by the inhabitants of Middle Earth (fastness, cities, boats, horn, sheath, shield), people and animals (dwarves, herdsmen, lore-masters, gaffer, hornblower, horse-men, riders, wolf-riders; dragon, steed), time expressions (season, yesteryear, yuledays) as well as certain prepositions and adverbs (along, among, outwards). As

can be seen from the above listings, most of these words are comparatively rare in present-day English and, although not archaic, work together to create an atmosphere of medievalism or otherworldliness.

As is to be expected given the logic of numbers, many of the significant collocations in Lord of the Rings are based on key words such as road, altogether, ancient or mighty, to give a few randomly chosen examples. A cursory look at the collocational range of certain key words suggests that Tolkien shows a marked preference for the most homely collocations available. However, Tolkien also uses everyday adjectives in what would have been considered unusual collocations even in his day. Great, for example, is most commonly associated with three meaning groups: abstract nouns such as peril, deed, force, power, strength; concrete objects such as trees, stones, halls, horns, gates; natural phenomena such as clouds, storms, shadow, wind or smoke. These collocations are derived from 19th century fiction, where they occur much less frequently, though.

While we have already noted Tolkien's preference for geographical terms, darkness and light, location and direction as well as colour adjectives, an analysis focussing on key semantic domains, apart from providing further detail about the domains just mentioned, highlights other domains which key word analysis misses entirely. One such domain is 'plants', where certain affinities are found with nineteenth-century fiction (e.g. reeds), but the domain is in fact much larger than suggested by the key word analysis, including the following items:

tree: elm-tree, alder-tree, fir-tree, pine, beech, acorn, willow; lobelia, nightshade, hemlock, rowan, ivy, hyacinth, marigold, hawthorn, fern, holly, shrub, sward, etc.

Other major sources of descriptivity include 'sailing/swimming', 'shape', 'weight: heavy', 'speed:fast (slow)', 'entire/maximum', 'sound:loud', 'temperature:hot/on fire', 'long, tall and wide', the last five domains lending additional weight to the hypothesis that Tolkien tends to resort to hyperbole. Semantic domains which are also prevalent in the realist novel (cf. Kullmann, 2009, p. 44) and partake in the creation of circumstantial realism include 'general appearance and physical properties' (e.g. hard, stony, bare, plain, bow, kneel, hewn, bold, hollow, blank, splendour, furry, weather-beaten) 'geographical terms' (e.g. land, forest, mountains, marshes, vale, wilderness, cave), and, to a minor extent, 'farming and horticulture' as well as 'fixing and mending'.

In summation it can be said that, far from being 'amateurish', Tolkien's prose effectively employs several centuries' worth of linguistic developments, putting old words to new uses and vice versa. In a much less obvious sense than in the case of James Joyce or Virginia Woolf, he is a linguistic postmodernist.

References

Dunning, T. (1993). Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics*, 19(1), 61-74.

Kirk, E. (1977). I Would Rather Have Written in Elvish: Language, Fiction, and The Lord of the Rings. Novel 5, 5 – 18. Reprinted in M. Spilka (Ed.) (1977) *Towards a Poetics of Fiction* (pp. 289-303). Bloomington: Indiana University Press.

Kullmann, T. (2009). Intertextual Patterns in J.R.R. Tolkien's The Hobbit and The Lord of the Rings. *Nordic Journal of English Studies*, 8(2), 37-56.

Rosebury, B. (2003). *Tolkien. A cultural phenomenon*. London: Palgrave Macmillan.

Scott, M. (2009). In Search of a Bad Reference Corpus. In D. Archer (Ed.), *What's in a word-list? Investigating word frequency and keyword extraction* (pp. 79-92). Oxford: Ashgate.

Snell-Hornby, M. (1983). *Verb-descriptivity in German and English*. Heidelberg: Winter.

Wilson, E. (1956). "Oo, Those Awful Orcs". A Reader's Companion to The Lord of the Rings and The Hobbit. *The Nation*, April 14.