Matt Gee*

"There was much new to grok": an analysis of word coinage in science fiction literature

Abstract: As can be witnessed in projects such as *The Oxford Dictionary of Science Fiction* (Prucher, Jeff. 2007. *Brave new words: The Oxford dictionary of science fiction.* Oxford: Oxford University Press), science fiction has been fertile ground for the creation of new words and concepts. Whereas the aforementioned dictionary was constructed by eliciting examples and citations from volunteers, this paper presents an initial foray into data-driven methods for uncovering lexis unique to science fiction. Words unique to science fiction texts are extracted by comparing a science fiction corpus against the British National Corpus (BNC Consortium. 2007. *The British National Corpus, XML edition.* Oxford Text Archive. http://hdl.handle.net/20.500.12024/2554 (accessed 29 June 2022)) to produce a list of 306 neologisms from 74 texts. In addition, this study seeks to examine the ways in which authors impart the meaning of such words to the reader, drawing on frameworks of semantic word relations and work in cognitive linguistics. This reveals the use of definitions and glosses by the authors, both in narration and direct speech, co-occurrence with synonyms, and the drip-feeding of attributes pertaining to the concept being referenced. In addition, characters can be shown to struggle with the concepts to which neologisms refer, allowing authors to explore themes of alienation and other-worldliness.

Keywords: corpus linguistics; science fiction; lexis; neology; cognitive linguistics

1 Introduction

Discovering new words within the extensive body of science fiction works is no small task. Perhaps the best known study of the science fiction lexicon is *Brave New Words: The Oxford Dictionary of Science Fiction* (Prucher 2007). The dictionary was compiled following the work of the Science Fiction Citations Project, in which the public were asked to contribute citations for words from science fiction texts. Since the development of the *COBUILD Dictionary* (Sinclair 1987), corpora have been important resources for evidencing language use when developing dictionaries (for an overview, see Hanks 2009; Moon 2007). Although recent developments in the digitization of classic pulp magazines (Rogers 2021) have provided data to help verify the citations in what is now named the *Historical Dictionary of Science Fiction* (Sheidlower 2020), public submissions remain the primary method by which words are suggested to the dictionary's compiler.

The goal of this study is to uncover science fiction word coinage using a data-driven approach. It draws on a corpus of 74 science fiction books first compiled in the late 1990s, but to date unused in academic study. Corpus studies have shown the value of data-driven approaches to identifying neology. For example, Renouf (1993) and Kristiansen and Andersen (2012) used regularly updated corpora of newspaper articles to identify neologisms, keeping a record of previously known forms and identifying new forms at each corpus update. The approach taken in this study is also to find words which occur in the science fiction texts that are otherwise unknown. However, with the corpus not being organized into regular time periods, a general reference corpus will be used as a record of known words against which the science fiction texts will be compared (see Section 2).

In addition to discovering new words, this study seeks to determine how writers present newly created forms and the concepts they denote. After all, in the words of science fiction author John Scalzi, "the thing about making up new terminology … it has to make sense not only within the universe that you're building but also in the

*Corresponding author: Matt Gee**, Birmingham City University, Birmingham, UK, E-mail: matt.gee@bcu.ac.uk. https://orcid.org/0000-0003-4499-5196

universe of the reader" (Rogers 2021: Para. 4). To uncover how this is achieved linguistically, the analysis notes processes of word formation (Section 2) before investigating the textual context of the first occurrences of the coinages in each novel (Section 3). In doing so it draws upon semantic relationships between words and cognitive linguistic theories relating to prototypes, subordinates, and attributes (see Ungerer and Schmid 2006 for an overview). Conclusions are presented in Section 4.

2 Data and method

The science fiction corpus used in this study (henceforth referred to as SFC) consists of 74 science fiction books, totalling 3,321,066 words. The corpus was compiled in the late 1990s by Professor Antoinette Renouf and colleagues in the Research and Development Unit for English Studies, now based at Birmingham City University.[1] The books come from multiple authors and cover the years 1890–1994 (see Appendix for details), with many of the texts appearing regularly in lists of popular or bestselling science fiction works. The SFC contains books of varying types, including novels, novellas (shorter than a novel but longer than a short story), and collections of short stories. The texts were digitized by scanning the books page by page using optical character recognition (OCR) tools available at the time of compilation (OCR errors were corrected by hand, although some remain). For the purpose of this study, the corpus has been part-of-speech (POS) tagged using the Stanford CoreNLP tagger (Manning et al. 2014).

The first stage of analysis was to compare the SFC against a reference corpus. In this case, the written portion of the BNC XML Edition (BNC Consortium 2007; Burnard 2007) was used, due to its size, balance of text types, genres, and subject domains, and the fact that it covers a time period no more recent than the SFC (up to 1993).[2] Word lists were extracted from both corpora and compared to identify words that occur in the SFC word list but not the BNC word list. In addition, POS information was retained in the word lists enabling proper names to be removed.

In total, 19,264 words (types) were found as a result of the comparison. However, many of these were errors remaining from the OCR conversion, spelling variations across varieties of English, or POS tagging errors in which proper names were tagged as other grammatical types. To facilitate the manual inspection of each word, only words occurring five or more times in the SFC were included for further analysis, resulting in a list of 471 types. Through inspecting concordances for these types, the errors could be identified and excluded, leaving 306 types. The grammatical split was 42 adjectives, 253 nouns, and 11 verbs, with no words of other grammatical types with frequency of five or more. Table 1 shows the top 50 words, ordered by number of occurrences in the corpus, along with POS and morphological information. The words found to be unique to the SFC (when compared against the BNC) will be referred to as "neologisms" or "coinages" throughout, but note that no assumption is made about their use outside of the texts in which they occur.

Looking briefly at morphology and drawing on established ideas of word formation (see Bauer and Huddleston 2002; Beliaeva 2019), 18 (34 %) of the 50 types in Table 1 are compounds (e.g., *dropship*, *wristpad*, *dreamgum*), 12 (26 %) are derivations (e.g., *hypersleep*, *unslave*), 9 (18 %) are inflections (of words occurring in the same texts, e.g., *grokked*, *unslaves*), 7 (14 %) are formations in which the source words are modified (either lexical blends or clipped compounds, e.g., *telescreen*, *endtendant*, *simstim*), and 4 (8 %) are clippings (e.g., *trodes*, *precog*). In addition, seven words without clear morphological components can be observed in the list of neologisms (*grok*, *triffid*, *kemmer*, *poietic*, *shifgrethor*, *yuny*, and *lortel*). The semantic aspects of such word formations, especially compounds and lexical blends, can be placed in the context of frames and the conceptual blending subtheory of cognitive linguistics (Fauconnier and Turner 2002). For example, the meaning of *dreamgum* could be inferred by combining knowledge of the act of dreaming with knowledge of gum as a type of foodstuff, suggesting

---

1 I was not part of the team involved at the time. It is remarkable that the physical copies and OCR versions of the corpus are still archived at Birmingham City University; but it is also thought that the corpus originally contained more texts.
2 The British National Corpus is comprised of British English texts, while the SFC contains multiple varieties. Differences in varieties of English will be accounted for during the analysis.

Table 1: The 50 most common words unique to the science fiction corpus used here in comparison to the British National Corpus.

Number of occurrences	Word	Part of speech	Morphology	Components
308	*grok*	V, N	unclear	
99	*grokked*	V	inflection	*grok + ed*
90	*telescreen*	N	blend	*tele(vision) + screen*
80	*triffid*	N	unclear	
73	*femmish*	Adj	derivation[a]	*fem + ish*
67	*dropship*	N	compound	*drop + ship*
62	*hypersleep*	N	derivation	*hyper + sleep*
53	*kemmer*	N	unclear	
51	*endtendant*	N	blend	*end + (at)tendant*
41	*grokking*	V, N	inflection	*grok + ing*
41	*Karhidish*	Adj	derivation	*Karhid(e) + ish*
39	*poietic*	Adj	unclear	
36	*EEV*	N	acronym	*emergency escape vehicle*
36	*simstim*	N	clipped compound	*sim(ulated) + stim(uli)*
36	*underpeople*	N	derivation	*under + people*
32	*grailstone*	N	compound	*grail + stone*
32	*shifgrethor*	N	unclear	
31	*bloodmother*	N	compound	*blood + mother*
31	*childpack*	N	compound	*child + pack*
31	*Gethenian*	Adj	derivation	*Gethen + ian*
30	*field-minder*	N	compound	*field + minder*
29	*wristpad*	N	compound	*wrist + pad*
27	*yuny*	N	unclear	
26	*conapt*	N	blend	*con(dominium) + ap(artmen)t*
26	*pulse-rifle*	N	compound	*pulse + rifle*
25	*groks*	V	inflection	*grok + s*
25	*penner*	N	embellished clipping	*pen(-propeller) + ner*
25	*precog*	N	clipping	*precog(nitive)*
25	*trodes*	N	clipping	*(elec)trodes*
23	*carry-fems*	N	compound & inflection[a]	*carry + fem + s*
23	*kinotrope*	N	neoclassical compound	*kino + trope*
22	*unslave*	N	derivation	*un + slave*
22	*unslaves*	N	derivation & inflection	*un + slave + s*
21	*chickenhead*	N	compound	*chicken + head*
21	*cold-pac*	N	clipping/clipped compound	*cold-pac(k)/cold + pac(kage/kaging)*
21	*mohole*	N	blend	*moho + hole*
20	*copseyes*	N	compound & inflection	*cops + eye + s*
20	*cutboy*	N	compound	*cut + boy*
20	*self-song*	N	derivation	*self + song*
20	*temperfoam*	N	compound	*temper + foam*
20	*dreamgum*	N	compound	*dream + gum*
19	*lortel*	N	unclear	
19	*man-apes*	N	compound & inflection	*man + ape + s*
19	*pedecab*	N	blend	*pede(strian) + cab*
19	*sharemothers*	N	compound & inflection	*share + mother + s*
19	*unmen*	N	derivation	*un + men*
19	*copseye*	N	compound	*cops + eye*
18	*kemmering*	N	inflection/derivation[b]	*kemmer + ing*
18	*whileawayans*	N	compound & derivation	*while + away + ans*
18	*texian*	Adj	derivation	*tex(as) + ian*

Notes: [a]*fem* and *fems* are also clippings of *female(s)*, but both terms occur in the SFC (in the same texts as *femmish* and *carry-fems*) and in the BNC. [b]On first inspection, *kemmering* appears to be an inflection of *kemmer*, but as both only occur as nouns, *kemmering* could be considered a derivation formed with a non-standard suffix.

dreamgum to be food that when eaten creates or enhances sleep to invoke dreams; but a full discussion of conceptual blending is beyond the scope of this article. Of course, the above presentation ignores the context available in written prose, which may hold essential information to understanding a neologism, especially when the morphological components are not immediately obvious. The remainder of the analysis discusses the kinds of clues to meaning that are provided in the textual context of the words.

3 Results

Perhaps the simplest way to introduce a new term to the reader is to provide a definition. Examples (1)–(4)[3] below show the first use of selected coinages and highlight how definitions may be provided.

(1)   Ever since mankind had gone through the Rediscovery of Man, bringing back governments, money, newspapers, national languages, sickness, and occasional death, there had been the problem of the **underpeople** – people who were not human, but merely humanly shaped from the stock of Earth animals. They could speak, sing, read, write, work, love, and die; but they were not covered by human law, which simply defined them as "homunculi" and gave them a legal status close to animals or robots.
(*The Ballad of Lost C'Mell*)

(2)   Twice the usual number of **copseyes** floated overhead, waiting. Gold dots against blue, basketball-sized, twelve feet up. Each a television eye and a sonic stunner, each a hookup to police headquarters, they were there to enforce the law of the Park.
(*Cloak of Anarchy*)

(3)   "What's a 'sharemother'?"
"One who shares the mothering of your child with you. I'm one of your **sharemothers**. Sheel, unfortunately perhaps, is another."
(*Motherlines*)

(4)   PFC Spunkmeyer was the **dropship** crew chief, the man responsible along with Pilot-Corporal Ferro for safely conveying his colleagues to the surface of whichever world they happened to be visiting, and then taking them off again in one piece.
(*Aliens*)

Example (1) shows an instance where the word *underpeople* is introduced and immediately followed by a definition, similar to that which may be found in a dictionary (e.g., see Saeed 2009: 58–59). The gloss is short and succinct, providing sufficient detail to understand the word. Even so, the subsequent sentence provides more information, elaborating on the physical description and adding information about the legal context. In example (2), the term *copseyes* is introduced by the narrator and then described in the subsequent sentences, starting with a physical description and moving on to information about the functions and purpose of *copseyes*. However, the gloss is less explicitly marked than in (1) and, whereas (1) reads like a stand-alone definition, (2) references the events that are occurring in the story at that moment. Example (3) concerns the word *sharemothers*, a term which is introduced earlier in the novel but not explained. Subsequently, as shown in (3), one of the characters asks what it means and another character gives a definition. The definition is a short gloss, much like (1), but, unlike those above, the introduction of the term and its definition all happen within direct speech rather than narration. In example (4), *dropship* is not itself defined but it appears as part of an entity name *dropship crew chief*, which is then immediately described. In defining the role of the agent (*crew chief*) that operates the entity, the purpose of the entity itself (the *dropship*) is simultaneously defined.

---

3 Book titles are included for each example. For further information see the Appendix.

In examples (5) and (6) below, the words are once again defined, but the definitions are less certain than those above.

(5)   The Numbers were marching along in step in neat ranks of four-hundreds and thousands of them in their sky-blue **yunies*** with the golden badge on each chest bearing each one's state number.
* Probably from the old word uniform.
(*We*)

(6)   a.   No doubt this was all a matter of **shifgrethor** – prestige, face, place, the pride-relationship, the untranslatable and all-important principle of social authority in Karhide and all civilizations of Gethen.
       b.   "I've made some mistake in **shifgrethor**. I'm sorry; I can't learn. I've never even really understood the meaning of the word."
             "**Shifgrethor**? It comes from an old word for shadow."
(*The Left Hand of Darkness*)

Example (5) concerns the words *yuny* and *yunies*. In this case, the etymology of the word is used to indicate its meaning, which is included in a footnote suggesting that the word originates from *uniform*. The author only hints at the origins of the word, even though they clearly know its etymology as they have coined it. In addition, the author does not provide much more detail about *yunies* in the rest of the book, which may be a stylistic choice designed to maintain a shroud of mystery around the type of clothing it represents. Example (6) concerns *shifgrethor*, which is defined by providing a list of near-synonyms ("prestige, face, place, the pride-relationship") and then followed by a short description revealing the cultural significance of the word. The WordNet project (Miller 1995) has shown that defining words as lists of synonyms can be an effective way to convey meaning, assuming the set of synonyms is sufficient to distinguish the meaning in relation to other known concepts (Miller et al. 1990: 240). However, the addition of the gloss revealing the cultural significance of the term, as well as describing it as "untranslatable", suggests the author deemed the synonyms alone insufficient, whether because the meaning would seemingly remain unclear or to stress the importance of the concept to the novel. In addition, several chapters later, more information is provided which reveals the etymology of the word ("an old word for shadow"). It should be noted that, even though all of this information about the meaning of *shifgrethor* is given to the protagonist, the character struggles to understand the concept throughout the majority of the novel. The *Left Hand of Darkness* is not alone in introducing concepts, including giving them novel lexical forms, that are deliberately difficult for the characters to grasp.

The textual context of first occurrences of the neologisms is explored further in the following examples, but in these cases no definition or description is provided for the terms. Instead, potential clues to meaning can be observed through semantic relationships with co-occurring words (indicated through the underlining here), which may also operate in combination with information from the morphological construction of a word.

(7)   "Empty apartments," Rick said. Sometimes he heard them at night when he was supposed to be asleep. And yet, for this day and age a one-half occupied **conapt** building rated high in the scheme of population density; […] The owner of the adjoining pasture, his **conapt** neighbour Bill Barbour, hailed him;
(*Do Androids Dream of Electric Sheep*)

(8)   "She'll have all the kin she needs," Nenisi said. "Alldera Holdfaster the fem will be **bloodmother** to her. Barvaran will be one of her sharemothers […]"
(*Motherlines*)

(9)   At three-thirty A.M. on the night of June 5, 1992, the top telepath in the Sol System fell off the map in the offices of Runciter Associates in New York City. […] The Runciter organization had lost track of too many of Hollis' psis during the last two months; this added disappearance wouldn't do. […] Sleepily, Runciter grated, "Who? I can't keep in mind at all times which inertials are following what teep or **precog**."
(*Ubik*)

In example (7), the word *conapt* is being used as a direct replacement for *apartment*. The word *apartments* is used first and then *conapt* starts to be used in multiword compounds in which *apartment* would seemingly also suffice (*conapt building* and *conapt neighbour*). The relation between the words is also reinforced by assigning the attribute of low usage to both concepts with *apartments* being premodified with "empty" and the *conapt building* with "half occupied". The word is seemingly a blend of *condominium* and *apartment* (Wiktionary 2022; a condominium being a type of apartment or apartment building). However, this information does not appear obvious from the word form alone and the concept of ᴄᴏɴᴀᴘᴛ[4] is not expanded upon in the remainder of the novel, making it unclear to what extent this information would be accessible to the reader.

In example (8), *bloodmother* is shown in proximity to the co-hyponym *sharemother*, which is defined earlier in the novel (example (3) above). Both words are signalled as being part of the category of ᴋɪɴ, but when the term *bloodmother* is used, it is signalled as being incompatible with *sharemother*, with each term being applied to separate characters. Thus, both words can be understood as belonging to a set of related terms, but also as having crucial differences in their meaning. This is reinforced by their morphological similarities and differences. Both words share a root component (*mother*) with the prepended components differentiating the type of mother. Here *blood* designates a biological parent (because they share "blood"), whereas *share* does not.

Example (9) concerns the word *precog*, which may be recognized as an abbreviation of *precognition*, but referring to people with the ability rather than to the ability itself. Furthermore, the word is introduced in a section which also contains other words that could be members of the same category, namely *telepath, psi*, and *teep*, suggesting a similarity in meaning to these terms. Of these, *telepath* may be seen to be a prototypical member (Heider 1971; Rosch 1973, 1975) and offers the greatest clue to *precog*'s meaning (with known attestations from over 80 years prior to the novel's publication according to the Oxford English Dictionary; OED Online 2022a). In example (9), *psi*, *teep*, and *precog* all share the property with *telepath* that they should be under surveillance, as shown by the phrases "fell off the map", "lost track of", and "are following what". A number of clues are available here to suggest the meaning of *precog* (which could be glossed as a person with exceptional powers of the mind who can predict the future), and, indeed, this is made clear as the novel progresses.

Further examples taken from Table 1 show more of the kinds of contextual clues that may appear near first occurrences of a term (here underlined), but also how authors can hold back information for thematic purposes.

(10)   It was the appearance of the **triffids** which really decided the matter for us. […] it was a **triffid** <u>sting</u> that had <u>landed me in hospital</u> […] Nor did their <u>seeds</u> float to us through space […] they were the outcome of a series of ingenious <u>biological</u> meddlings […] it was the <u>seeds</u> of the **triffid** he had set out to bring. […] My introduction to a **triffid** came early. It so happened that we had one of the first in the locality <u>growing</u> in our own <u>garden</u>.
        (*The Day of the Triffids*)

(11)   a.   They were in the first phase of **kemmer**. The large, soft snow danced about them as they stood barefoot in the icy mud, <u>hands clasped</u>, <u>eyes all for each other</u>. […] Harmes, <u>beloved</u> **kemmering** of Emran III, whose beauty is still celebrated […] there were two brothers who <u>vowed</u> **kemmering** to each other […] full brothers were permitted to keep **kemmer** until one of them should <u>bear a child</u> […]
        b.   […] on the 22nd or 23rd day the individual enters **kemmer**, estrus. […] When the individual finds a partner in **kemmer**, hormonal secretion is further stimulated […] until in one partner either a male or female hormonal dominance is established.
        (*The Left Hand of Darkness*)

---

4 This study follows the typographical convention of using small caps for concepts and cognitive categories (see Ungerer and Schmid 2006: ix).

(12)   All four symbols in the query were in Smith's vocabulary but he had trouble believing that he had heard
         them rightly. […] He was filled with mild regret, since there was still so much to **grok** of these new events,
         but no reluctance. […] Then he turned to Smith. "Bowels move?" he asked.
         Smith understood this; Nelson always asked about it. "No, not yet."
         (*Stranger in a Strange Land*)

Example (10) concerns the word *triffid.* The term is introduced in such a manner that pieces of information are drip fed to the reader. The TRIFFID concept is central to the novel so this may be a deliberate tactic designed to invoke suspense and the unusualness of the entity. With each occurrence of *triffid*, additional information is provided which can help the reader to understand the concept, which could be seen as assigning attributes to the concept. For example, the words underlined in example (10) could enable attributes such as "can sting", "is dangerous", "comes from seed", "is biological", and "grows in a garden" to be associated with TRIFFID. To some extent these attributes would relate to a prototypical PLANT, but of course, not all plants can sting and even fewer have a sting capable of hospitalizing a person. However, this should not prevent the reader from including TRIFFID in the PLANT category as it shares a number of attributes with a prototypical PLANT ("comes from seed", "is biological", and "grows in a garden"). From the perspective of basic-level categories (Berlin et al. 1974; Rosch and Mervis 1975; Rosch et al. 1976), PLANT could be the basic-level category, in which case the TRIFFID concept, exhibiting a type-of relationship, would be a subordinate category. In this case the "can sting" and "is dangerous" attributes enable TRIFFID to be distinguished from a prototypical PLANT.

Example (11) concerns the words *kemmer* and *kemmering,* which, similar to example (10), sees the terms introduced in an incremental manner. In this case, the concept associated with these words is not fully explained until Chapter 7 (of 20) in which it is defined in detail in the style of an encyclopedia entry (example (11b)). As can be seen from (11a), it first co-occurs with terms that relate to ROMANCE, MARRIAGE, and CHILDBIRTH, and a vague understanding of KEMMER being a mixture of these may be sufficient for the story up to Chapter 7. The novel in question, *The Left Hand of Darkness,* involves a character travelling to an alien world. In this context, not specifying the meaning of *kemmer* enables the author to invoke a feeling of alienness. Furthermore, "brothers" are the agents noted to be in *kemmer(ing)* with one another. The prototypical notion of BROTHERS is that of male blood relations, which is in conflict with the prototypical notion of MARRIAGE (that of a bond between non-blood relations of opposite sex), and even more so with CHILDBIRTH (which typically requires a female mother). This further adds to the alienness of the KEMMER concept, which is only resolved when the process of partners changing biological sex is explained in Chapter 7 of the novel.

It was noted previously that new terms are used to show a protagonist's struggle with an alien world (e.g., *kemmer* and *shifgrethor*). This tactic is used to an even greater extent in *Stranger in a Strange Land,* where *grok, grokking, grokked,* and *groks* (shown in the article title and example (12)) are terms used to illustrate a concept that does not exist on Earth but is central to a visiting alien's society. Few verbs were found among the neologisms (3.6 %), but *grok* and its inflections are predominantly used as verbs.[5] It is an action that is solely associated with the alien character, highlighting their differences from the other characters. In fact, the other characters are shown to try to understand the concept, but struggle to do so as the novel progresses, simultaneously highlighting a central theme of the novel.

It is also worth noting that *grok* has subsequently been used in other texts outside the original novel and been the subject of previous linguistic study (Wright 2008). It is found in the SFC in *The Female Man*, published 14 years after *Stranger in a Strange Land,* has been used in the domain of computing programming, and is included in the Oxford English Dictionary (which notes *Stranger in a Strange Land* as the first use of the term; OED Online 2022b). This phenomenon is discussed further in the conclusion below.

---

5 The noun derivatives *grok* and *grokking* can also be found, but are less frequent.

4 Conclusions

This study sought to uncover the nature of words created in science fiction stories and how authors present new terms and concepts in their writing. The findings show that many words follow well-understood word formation rules including compounding, derivation, lexical blending, and clipping. As revealed through investigation of the textual context where the neologisms first occur in each text, one strategy used is to provide definitions of new words, taking the form of dictionary-like glosses, notes regarding etymology, description by the narrator, or description during direct speech. The definitions and descriptions were found to give information about the physical aspects and the functions of the denoted concepts. It was also noted that terms can be defined by describing the actions of agents that use an entity, rather than by describing an entity itself.

Where a definition is not provided, other words with semantic relations to the neologism, such as synonyms or co-hyponyms, were observed in the context of the neologism's first occurrence. Alternatively, clues can be discerned from the textual context that provide attribute-like information for a new concept. In some cases, these practices are combined, such that properties and actions may be associated with and shared between the semantically related words, revealing a connection between them. It was also noted that a new concept can be made unusual or alien to the reader by placing the new category at the periphery of existing categories, or by juxtaposing it with prototypical notions of similar concepts. Furthermore, it was observed that authors can associate new words with concepts that central characters struggle to grasp, enabling authors to explore themes of alienation and cultural difference.

This work represents first steps into using data-driven corpus methods to evidence coinage within the lexis of science fiction. To progress further, a great deal more work would be required to increase the scale of the corpus and grow the body of quantitative evidence. Of the 306 neologisms examined, 18 % were found to occur in multiple texts, with only 5 % used by multiple authors. This lack of shared words is somewhat surprising. Indeed, entries in *The Oxford Dictionary of Science Fiction* were required to appear in multiple fictional universes for inclusion, and Prucher (2007: xv–xvi) notes that there "are a great many terms … that are used by many authors in similar ways". However, the dictionary citations were drawn from around 1,160 texts, whereas, at 74 texts, the corpus investigated here is much more limited in size. In addition, science fiction has expanded considerably as a genre (through mainstream TV, movie franchises, and video games, for example) since the publication of the texts studied. It is hoped that resources will emerge from future studies to enable science fiction corpus studies of much greater coverage.

Appendix

See Table 2.

Table 2: The books that make up the science fiction corpus (SFC).

Title	Author	Year first published	Number of words
News from Nowhere	Morris, William	1890	49,755
The Time Machine	Wells, H. G.	1895	32,606
The War of the Worlds	Wells, H. G.	1898	60,522
The Land Ironclads	Wells, H. G.	1903	8,405
A Modern Utopia	Wells, H. G.	1905	87,402
Finis	Pollack, Frank L.	1906	4,022
As Easy as ABC	Kipling, Rudyard	1912	10,329
We (English translation by Gregory Zilboorg)	Zamyatin, Yevgeny	1924	64,506
The Metal Man	Williamson, Jack	1928	4,838
Brave New World	Huxley, Aldous	1932	65,330

Table 2: (continued)

Title	Author	Year first published	Number of words
A Martian Odyssey	Weinbaum, Stanley G.	1934	10,365
Night	Campbell, John W.	1935	8,187
By His Bootstraps	Heinlein, Robert	1941	20,995
Desertion	Simak, Clifford D.	1944	4,496
The Piper's Son	Padgett, Lewis	1945	10,076
The Monster	van Vogt, Alfred E.	1948	6,652
The Sentinel	Clarke, Arthur C.	1948	3,807
1984	Orwell, George	1949	103,967
I, Robot	Asimov, Issac	1950	71,760
The Second Night of Summer	Schmitz, James H.	1950	10,255
The Voyage of the Space Beagle	van Vogt, Alfred E.	1950	74,098
Second Dawn	Clarke, Arthur C.	1951	12,299
The Day of the Triffids	Wyndham, John	1951	92,125
Crucifixus Etiam	Miller, Walter M.	1953	7,425
Fahrenheit 451	Bradbury, Ray	1953	47,022
The Tunnel under the World	Pohl, Frederik	1954	11,499
The Naked Lunch	Burroughs, William	1959	62,071
Guardians of Time	Anderson, Poul	1960	11,275
Stranger in a Strange Land	Heinlein, Robert	1961	213,767
The Ballad of Lost C'Mell	Smith, Cordwainer	1962	7,591
Billenium	Ballard, J. G.	1962	5,977
The Man in the High Castle	Dick, Philip K.	1962	80,657
The Rest of the Robots	Asimov, Issac	1964	51,651
Semley's Necklace	Le Guin, Ursula K.	1964	7,577
Who Can Replace a Man	Aldiss, Brian	1965	3,237
How Beautiful with Banners	Blish, James	1966	3,907
A Criminal Act	Harrison, Harry	1967	5,259
Problems of Creativeness	Disch, Thomas M.	1967	9,008
2001: A Space Odyssey	Clarke, Arthur C.	1968	61,265
Do Androids Dream of Electric Sheep	Dick, Philip K.	1968	63,454
The Left Hand of Darkness	Le Guin, Ursula K.	1969	87,234
Ubik	Dick, Philip K.	1969	68,220
How the Whip Came Back	Wolfe, Gene	1970	5,959
To Your Scattered Bodies Go	Farmer, Philipe J.	1971	77,294
Cloak of Anarchy	Niven, Larry	1972	7,634
A Thing of Beauty	Spinrad, Norman	1973	6,085
Walk to the End of the World	Charnas, Susie McKee	1974	74,268
The Female Man	Russ, Joanna	1975	69,499
Woman on the Edge of Time	Piercy, Marge	1976	140,676
The Screwfly Solution	Sheldon, Raccoona	1977	8,389
Motherlines	Charnas, Susie McKee	1978	78,428
Alien	Foster, Alan D.	1979	68,976
Shikasta	Lessing, Doris	1979	161,590
The Way of Cross and Dragon	Martin, George R. R.	1979	7,409
The Marriages between Zones Three, Four and Five	Lessing, Doris	1980	99,630
Burning Chrome	Gibson, William	1982	7,960
Neuromancer	Gibson, William	1984	80,933
Silicon Muse	Schenck, Hilbert	1984	8,043
The Handmaid's Tale	Atwood, Margaret	1985	94,360
Aliens	Foster, Alan D.	1986	78,577
A Gift from the Culture	Banks, Iain M.	1987	6,496
Cleaning Up	Banks, Iain M.	1987	5,749
Descendant	Banks, Iain M.	1987	8,309
Karl and the Ogre	McAuley, Paul J.	1988	6,039
Road of Skulls	Banks, Iain M.	1988	1,599
Odd Attachment	Banks, Iain M.	1989	2,017

Table 2: (continued)

Title	Author	Year first published	Number of words
Piece	Banks, Iain M.	1989	3,020
The State of the Art	Banks, Iain M.	1989	34,305
Swarm	Sterling, Bruce	1989	9,678
The Difference Engine	Gibson, William	1990	140,059
Piecework	Brin, David	1991	10,485
Alien 3	Foster, Alan D.	1992	57,707
Red Mars	Robinson, Kim Stanley	1992	213,794
The Furies	Charnas, Susie McKee	1994	121,206

References

Bauer, Laurie & Rodney Huddleston. 2002. Lexical word formation. In Rodney Huddleston & Geoffrey K. Pullum (eds.), *The Cambridge grammar of the English language*, 1621–1721. Cambridge: Cambridge University Press.

Beliaeva, Natalia. 2019. Blending in morphology. *Oxford Research Encyclopedia of Linguistics*. https://doi.org/10.1093/acrefore/9780199384655.013.511.

Berlin, Brent, Dennis Breedlove & Peter Raven. 1974. *Principles of Tzeltal plant classification*. New York: Academic Press.

BNC Consortium. 2007. *The British National Corpus, XML edition*. Oxford: Oxford Text Archive. http://hdl.handle.net/20.500.12024/2554 (accessed 29 June 2022).

Burnard, Lou. 2007. *Reference guide for the British National Corpus (XML edition)*. http://www.natcorp.ox.ac.uk/docs/URG/(accessed 29 June 2022).

Fauconnier, Gilles & Mark Turner. 2002. *The way we think: Conceptual blending and the mind's hidden complexities*. New York: Basic Books.

Hanks, Patrick. 2009. The impact of corpora on dictionaries. In Paul Baker (ed.), *Contemporary corpus linguistics*, 214–237. London: Continuum.

Heider, Eleanor. 1971. "Focal" color areas and the development of color names. *Developmental Psychology* 4. 447–455.

Kristiansen, Marita & Gisle Andersen. 2012. Corpus approaches to neology and their relevance for dynamic domains. *Neologica* 6. 43–62.

Manning, Christopher, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard & David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In Kalina Bontcheva & Jingbo Zhu (eds.), *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System demonstrations*, 55–60. Stroudsburg: Association for Computational Linguistics.

Miller, George, A. 1995. WordNet: A lexical database for English. *Communications of the ACM* 38(11). 39–41.

Miller, George A., Richard Beckwith, Christiane Fellbaum, Derek Gross & Katherine J. Miller. 1990. Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography* 3(4). 235–244.

Moon, Rosamund. 2007. Sinclair, lexicography, and the Cobuild Project: The application of theory. *International Journal of Corpus Linguistics* 12(1). 159–181.

OED Online. 2022a. *Telepath, n*. Oxford: Oxford University Press. https://www.oed.com/view/Entry/37623394?rskey=h4uls1 (accessed 29 June 2022).

OED Online. 2022b. *Grok, v*. Oxford: Oxford University Press. https://www.oed.com/view/Entry/81710?redirectedFrom=grok (accessed 29 June 2022).

Prucher, Jeff. 2007. *Brave new words: The Oxford dictionary of science fiction*. Oxford: Oxford University Press.

Renouf, Antoinette. 1993. A word in time: First findings from dynamic corpus investigation. In Jan Aarts, Pierter de Haan & Oostdijk Nelleke (eds.), *English language corpora: Design, analysis and exploitation*, 279–288. Amsterdam: Rodopi.

Rogers, Adam. 2021. A new way to trace the history of sci-fi's made-up words. *WIRED*, 27 January. https://www.wired.com/story/historical-dictionary-of-science-fiction/ (accessed 29 June 2022).

Rosch, Eleanor. 1973. On the internal structure of perception and semantic categories. In Timothy Moore (ed.), *Cognitive development and the acquisition of language*, 111–144. New York: Academic Press.

Rosch, Eleanor. 1975. Cognitive representations of semantic categories. *Journal of Experimental Psychology: General* 104. 193–233.

Rosch, Eleanor & Caroline Mervis. 1975. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology* 7. 573–605.

Rosch, Eleanor, Caroline Mervis, Wayne Gray, David Johnson & Penny Boyes-Braem. 1976. Basic objects in natural categories. *Cognitive Psychology* 8. 382–439.

Saeed, John I. 2009. *Semantics*, 3rd edn. Malden, MA: Wiley-Blackwell.

Sheidlower, Jesse. 2020. *Historical dictionary of science fiction*. https://sfdictionary.com/ (accessed 29 June 2022).

Sinclair, John. 1987. The dictionary of the future. *Library Review* 36(4). 268–278.

Ungerer, Friedrich & Hans-Jörg Schmid. 2006. *An introduction to cognitive linguistics*, 2nd edn. London: Routledge.

Wiktionary. 2022. *Conapt, n*. San Francisco: Wikimedia Foundation. https://en.wiktionary.org/wiki/conapt (accessed 30 October 2022).

Wright, David E. 2008. Do words have inherent meaning? *ETC: A Review of General Semantics* 65(2). 177–190.