| Category | Feature | Code/ Tag | Examples | Operationalisation | Normalisation unit | As coded by |
|---|---|---|---|---|---|---|
| **Features for which there are no tags in the tagged texts** | | | | | | |
| General text properties | Total number of words | **Words** | *It's a shame that you'd have to pay to get that quality.* (= 14) | The number of tokens as tokenised by the Stanford Tagger, but excluding punctuation marks, brackets, symbols, genitive 's (POS), and filled pauses and interjections (FPUH). Contractions are treated as separate words, i.e.,it's is tokenised as it and 's. Note that this variable is only used to normalise the frequencies of other linguistic features. | NA | Le Foll |
| General text properties | Average word length | **AWL** | *It's a shame that you'd have to pay to get that quality.* (42/12 = 3.50) | Total number of characters in a text divided by the number of words in that same text (as operationalised in the Words variable above, hence excluding filled pauses and interjections, cf. FPUH). | Words | Le Foll |
| General text properties | Lexical diversity | **TTR** | ***It's a shame that you'd have to pay*** *to* ***get*** *that* ***quality*** *.* (12/14 = 0.85) | Following Biber (1988), this feature is a type-token ratio measured on the basis of, by default, the first 400 words of each text only. It is thus the number of unique word forms within the first 400 words of each text divided by 400. This number of words can be adjusted in the command used to run the script (see instructions at the top of the MFTE script). | Words (by default first 400) | Le Foll |
| General text properties | Lexical density | **LDE** | *It's a* ***shame*** *that you'd have to* ***pay*** *to get that* ***quality*** *.* (3/14 = 0.21) | For this feature, tokens which are not on the list of the 352 function words from the {qdapDictionaries} R package, nor individual letters, or any of the fillers listed in FPUH are identified as content words. Lexical density is calculated as the ratio of these content words to the total number of words in a text. | Words | Le Foll |
| General text properties | Finite verbs | **FV** | *He discovered that the method involved imbiding copious amounts of tea. Ants can survive by joining together to morph into living rafts. Always wanted to experience the winter wonderland that Queen Elsa created?* | This feature is not directly listed in the MFTE output tables; however, it is used as a normalisation basis for many other linguistics features (see Normalisation column). It is calculated by tallying the number of occurrences of the following features: VPRT, VBD, VIMP, MDCA, MDCO, MDMM, MDNE, MDWO and MDWS. | NA | Le Foll |
| **Features for which there are tags in tagged version of the texts processed by the MFTE** | | | | | | |
| Adjectives | Attributive adjectives | **JJAT** | *I've got a* ***fantastic*** *idea! I didn't sleep at all* ***last*** *night.* ***Cheap*** *,* ***quick*** and ***easy*** *fix!* | Whereas the Biber Tagger and the MAT first identify predicative adjectives and then consider all remaining J.* tags from the Stanford Tagger to be attributive adjectives, the MFTE proceeds the other way around because it is considerably easier to reliably identify attributive adjectives than it is predicative adjectives. Thus, all adjectives (J.*, as tagged by the Stanford Tagger) followed by another adjective, a noun or a cardinal number, or preceded by a determiner are tagged as attributive adjectives. Once these first attributive adjectives have been identified, an additional loop is run to capture any additional attributive adjectives found in lists of attributive adjectives. | Nouns | Le Foll |

| | | | | | | |
|---|---|---|---|---|---|---|
| Adjectives | Predicative adjectives | **JJPR** | *That's **right**. One of the main advantages of being **famous** ... It must be absolutely **wonderful**.* | Once attributive adjectives have been identified (see JJAT) and tagged as JJAT, all remaining JJ, JJS and JJR tags are overwritten as JJPR. In addition, *ok* and *okay* in the construction *BE ok(ay)* are also tagged as JJPR. These words are otherwise identified as foreign words (FW) by the Stanford Tagger. | Finite verbs | Le Foll |
| Adverbials | Frequency references | **FREQ** | *We should **always** wear a mask. But he had found his voice **again.*** | Assigned to all occurrences of the frequency adverbs listed in the COBUILD (Sinclair et al. 1900: 270): *usually, always, mainly, often, generally, normally, traditionally, again, constantly, continually, frequently, ever, never, infrequently, intermittently, occasionally, often, periodically, rarely, regularly, repeatedly, seldom, sometimes* and *sporadically.* | Finite verbs | Le Foll |
| Adverbials | Place references | **PLACE** | *It's not **far** to go. I'll get it from **upstairs**. It's **downhill** all the way. It's **there** not here.* | Biber's (1988: 224) list of place adverbials was taken from Quirk et al. (1985:514ff) but inexplicably excludes many from this list. Those that do not fulfil other major functions were therefore added: downwind, eastward(s), westward(s), northward(s), southward(s), upwards, downwards, elsewhere, everywhere, here, offshore, nowhere, somewhere, thereabout(s) and there (but occurrences of there tagged as existential there (EX) by the Stanford Taggers were ignored). Only occurrences of far which have not previously identified as TIME references (e.g., so far, thus far) or emphatics (e.g., far better, far more) are tagged as PLACE references. | Finite verbs | Le Foll, adapted from Biber (1988) |
| Adverbials | Time references | **TIME** | *It will **soon** be possible. **Now** is the time. I haven't come across any issues **yet**.* | All occurrences of *afterwards, again, earlier, early, eventually, forever, formerly, immediately, initially, instantly, late, lately, later, momentarily, now, nowadays, once, originally, presently, previously, recently, shortly, simultaneously, subsequently, today, to-day, tomorrow, to-morrow, tonight, to-night, yesterday.* Following Nini (2014: 18), the word *soon* was not tagged as a time adverbial when followed by the word *as*. *Ago, already, beforehand, prior to,* and *far* (the latter only when proceeded by *so* or *thus* and not followed by an adjective or adverb), and *am* and *pm* as adverbs were added to the list, as well as *yet* tokens that have not previously been identified as concessives (CONC). | Finite verbs | Le Foll, adapted from Nini (2014) |
| Adverbials | Other adverbs | **RB** | ***Unfortunately** that's the case. **Exactly** two weeks. He could **so easily** but he knows **better**. He's **still** gonna come **back**.* | Corresponds to all the tokens tagged as RB, RBS, RBR or WRB by the Stanford Tagger apart from those identified as adverbs of frequency (FREQ), place (PLACE) or time (TIME), amplifiers (AMP), emphatics (EMPH), hedges (HDG) and downtoners (DWNT). | Words | Le Foll |
| Determinatives | *s* -genitives | **POS** | *the world **'s** two most populous country, my parents **'** house* | As identified by the Stanford Tagger: the possessive endings on nouns ending in 's or '. Note that these tokens are not counted as Word in the computation of the lexical diversity (TTR) and average word length variables (AWL) features. | Nouns | Le Foll |
| Determinatives | Determiners | **DT** | *Is that **a** new top? **The** first line has to be interesting. Are they **both** Spice Girls? On **either** side of the page. To **another** room. They're five pounds **each**.* | As tagged by the Stanford Tagger (DT) (Santorini 1990: 2), with the exception of *that, this, these* and *those* which are counted as demonstratives (DEMO). Note that this Stanford Tagger category also includes pronouns such as *another* in *Shall I choose <u>another</u> ?* | Nouns | Le Foll |

| | | | | | | |
|---|---|---|---|---|---|---|
| Determinatives | Quantifiers | **QUAN** | ***Such*** *a good time in like* ***half*** *an hour. She's got* ***all*** *these great ideas. It happens* ***each*** *and* ***every*** *time.* | All occurrences of pre-determiners as tagged by the Stanford Tagger, which includes the following "determiner-like elements when they precede an article or possessive pronoun" (Santorini 1990: 4): *nary, quite, rather* and *such* (e.g., quite a mess, rather a nuisance, many a moon), as well as all instances of all (unless immediately followed by *right* , cf. DMA), *any, a bit, both, each, every, few, half, many, much, several, some, lots, a lot (of), load(s) of, heaps of, wee, less* and *more* (as adjectives only). | Nouns | Le Foll |
| Determinatives | Numbers | **CD** | *That's her number* ***one*** *secret.* ***Two*** *eyes glowed just above the surface. It happened on* ***7*** *February,* ***2019*** *.* | All cardinal numbers as identified by the Stanford Tagger. This includes dates written in numbers, e.g., *1994.* In addition, numbers listed as list markers (LS) by the Stanford are overwritten as CD and strings of the type *\b[0-9]+th_ \| \b[0-9]+nd_ \| \b[0-9]+rd_* are also tagged as numbers (CD). | Words | Le Foll |
| Determinatives | Demonstratives | **DEMO** | *What are you doing* ***this*** *weekend? I love* ***that*** *film. Whoever did* ***that*** *should admit it.* | Assigned to all occurrences of *that, this, these* and *those* identified by the Stanford Tagger as determiners (DT). | Words | Le Foll |
| Discourse organisation | Elaborating conjunctions | **ELAB** | ***Similarly*** *, you may,* ***for example*** *, write bullet points* ***insomuch as*** *it helps you to focus your ideas.* | *Assigned to such that (not followed by a determiner), such as, inasmuch as, insofar as, insomuch as, in that, to the extent that, in particular, in conclusion, in sum, in summary, to summarise, to summarize, for example, for instance, in fact, in brief, in any event, in any case, in other words, e(.)g(.), in summary, viz(.), cf(.), i.e., namely, etc(.), likewise, namely, as well as similarly and accordingly when followed by a comma.* | Finite verbs | Le Foll |
| Discourse organisation | Coordinating conjunctions | **CC** | ***Instead*** *of listening to us, he* ***also*** *told John* ***and*** *Jill* ***but*** *at least his parents don't know yet.* | *This category takes the coordinating conjunctions (CC) tagged by the Stanford Tagger as its basis which include and, but, nor, or, yet, "as well as the mathematical operators plus, minus, less, times (in the sense of 'multiplied by') and over (in the sense of 'divided by'), when they are spelled out" (Santorini 1990: 2). However, conjunctions already captured by other variables are excluded from this count: yet is assigned to concessive (CONC). In addition, the following (multi-word) conjunctions are also included in this category: also, besides, moreover, further (when tagged as an adverb), furthermore, in addition, additionally, as well (as) (except when preceded by least), however (provided it is preceded or followed by a punctuation mark), ibid, on the one hand, on the other hand, instead, besides, conversely, by/in contrast, on the contrary, in/by comparison, whereas, whereby, whilst.* | Finite verbs | Le Foll |
| Discourse organisation | Causal conjunctions | **CUZ** | *He was scared* ***because*** *of the costume. Yeah* ***coz*** *he hated it.* | Assigned to all occurrences of *because* , *'cause, cos, cuz* and *coz* . The latter four were not included in Biber's (1988) original variable. According to Biber (1988: 236) *because* "is the only subordinator to function unambiguously as a causative adverbial". Whilst it is true that many subordinators, e.g., *as, for,* and *since* , can fulfil a range of functions, including causative, and were therefore not included in this category, the following adverbs and multi-word conjunctions were added since they mostly fulfil a causative function: *as a result, on account of, for that/this purpose, thanks to, to that/this end, consequently, in consequence, hence, so that, therefore, thus* . | Finite verbs | Le Foll, adapted from Biber (1988) |

| | | | | | | |
|---|---|---|---|---|---|---|
| Discourse organisation | Concessive conjunctions | **CONC** | *Even **though** the antigens are normally hidden…* | Assigned to all occurrences of *although, though, tho, despite, except that, in spite of, albeit, granted that, nevertheless, nonetheless, notwithstanding, whereas, no matter +* WH-word*, (ir)regardless of* , and *granted.* Also assigned to *still* and *yet* when preceded by any punctuation mark or followed by a comma. Multi-word units are only counted as one occurrence of CONC. | Finite verbs | Le Foll |
| Discourse organisation | Conditional conjunctions | **COND** | ***If** I were you… Even **if** the treatment works…* | Assigned to all occurrences of *if, as long as, unless, lest, in that case, otherwise, whether.* | Finite verbs | Le Foll |
| Discourse organisation | Discourse/pragmatic markers | **DMA** | ***Well no** they didn't say **actually** . **Okay I guess** we'll see how things go **right** ?* | Assigned to "interactional signals and discourse markers" (as listed in Stenström 1994: 59 and cited in Aijmer 2002: 2): *actually, all right, anyway, God, goodness, gosh, OK, okay, right* (if tagged as an interjection by the Stanford Tagger)*, well* (only if identified by the Stanford Tagger as an adverb or adjective and not if preceded by *as, how, very, really, quite,* a verb, an adjective or an adverb)*, yes, yeah, yep, sure* (unless it is preceded by the verb *MAKE, for, not* or *you* ). Verbal phrases such as *you know* and *I mean* were excluded from this variable since literal occurrences could not be automatically disambiguated occurrences as discourse markers. A number of markers from Stenström's list are also not assigned this tag because they are captured by other variables: *now* (TIME), *please* (POLITE), *really* (EMPH), *quite* and *sort of* (HDG). The following items were added: *lol, IMO, omg, wtf, nope, mind you, of course, whatever* and *damn* (unless tagged as a verb, or followed by an adjective; in the latter case it is an emphatic, cf. EMPH). | Words | Le Foll |
| Discourse organisation | Filled pauses and interjections | **FPUH** | ***Oh** nooooooo, Tiger's furious! **Wow** ! **Hey** Tom! **Er** I don't know. **Hmm** .* | Assigned to all occurrences of ah+, aw+, oh+, eh+, er+, erm+, mm+, ow+, um+, huh+, uhu+, uhuh, mhm+, hm+ (but not HM), oo+ps woo+ps, hi, hey, and interjections identified by the Stanford Tagger and not assigned to another category. The plus sign (+) signifies that that the preceding letter can appear multiple times, i.e.,ahh and errrr are also assigned this tag. | Words | Le Foll |
| Discourse organisation | *Like* | **LIKE** | *Sounds **like** me. And just **like** his father. And he was **like** this isn't true. I wasn't gonna **like** do it.* | Occurrences of like tagged as a preposition (IN) or adjective (JJ) by the Stanford Tagger are assigned this tag because, in spoken English, like typically fulfils a range of different functions, e.g.,fillers and softeners, and attempts to disambiguate like as a preposition or conjunct proved too error-prone. This category excludes occurrences of like identified as the quotative BE + like (QLIKE) if the QLIKE feature is included (which, by default, it is not, cf. tagger evaluation). | Words | Le Foll |
| Discourse organisation | *So* | **SO** | *She had spent **so** many summers there. **So** there you go.* | Occurrences of *so* tagged as IN by the Stanford Tagger and not previously identified as either an emphatic (*so +* J.**/much/many/little* ; EMPH) or an adverbial subordinator (*so that* + NN.*/J.*; OSUB) are assigned this tag. | Words | Le Foll |
| Discourse organisation | Direct WH-questions | **WHQU** | ***What**'s happening? **Why** don't we call the game off? **How** ? And **who** is Dinah, if I might venture to ask the question?* | Assigned to *what, where, when, how, why, who, whom, whose* and *which* followed by a question mark within 15 tokens. | Finite verbs | Le Foll |

| | | | | | | |
|---|---|---|---|---|---|---|
| Discourse organisation | Question tags | **QUTAG** | *Do they **?** Were you **?** It's just it's repetitive, isn't it **?*** | Assigned to question marks preceded by (1) *innit, init;* (2) a modal verb (MD) or *did* or *had,* and a personal pronoun (P.+); (3) a modal verb or *did* or *had,* a negation (XX0), and a personal pronoun; (4) *is, does, was* or *has* , followed by *it, she* or *he;* (5) *is, does, was* or *has* , followed by a negation, and *it, she* or *he* ; (6) *do, were, are* or *have* , followed by *you, we* or *they* ; (7) *do, were, are* or *have* , followed by a negation, and *you, we* or *they* . In addition, the above patterns are not considered question tags if a question word occurs within six words to the left of the question mark; consequently, *Why did you do it?* is not assigned this tag but rather WHQU. | Finite verbs | Le Foll |
| Discourse organisation | Yes/no questions | **YNQU** | ***Have** you thought about giving up? **May** I take a seat? **Do** you mind?* | Assigned to any form of the verbs BE, HAVE, DO or a modal verb (MD) followed by a personal pronoun (P.+), a noun (NN.*), a negation (XX0) or determiner (DT) and then a question mark within three to 15 tokens, as long as no WH-question (WHQU) or yes/no question tag (YNQU) is present one or two tokens before the auxiliary verb. Note that this variable should not overlap with question tags (QUTAG). | Finite verbs | Le Foll |
| Discourse organisation | *that* relative clauses | **THRC** | *You must be very clever to find a use for something **that** costs nothing. I'll just run a cable **that** goes from here to there.* | Assigned to that identified as introducing a relative clause by the Stanford Tagger (WDT), unless it is immediately followed by a punctuation mark. Any remaining that_WDT tokens are typically mistagged demonstratives and are thus assigned to the DEMO category, e.g., *I don't think that's a problem that is.* | Finite verbs | Le Foll |
| Discourse organisation | *that* subordinate clauses (other than relatives) | **THSC** | *Did you know **that** the calendar we use today was started by Julius Caesar? She resented being told constantly **that** she was ignorant and stupid.* | Assigned to that tokens which have been tagged as IN by the Stanford Tagger and are not immediately followed by a punctuation mark. Remaining that_IN tokens are assigned to the demonstrative category (DEMO): these are end-of-sentences/utterances tokens which are typically misidentified by the Stanford Tagger, e.g., *Who was that?* | Finite verbs | Le Foll |
| Discourse organisation | Subordinator *that* omission | **THATD** | *I mean **[THATD]** you'll do everything. I thought **[THATD]** he just meant our side. You don't think **[THATD]** he's a drug dealer? I know **[THATD]** that's not his thing.* | The THATD tag is assigned to the following patterns: (1) a public, private or suasive verb followed by a demonstrative pronoun (DEMO) or I, we, he, she, it, they and then a verb (V.* or MD); (2) a public, private or suasive verb followed by I, we, he, she, it, they or a noun (N.*), and then by a verb (V.* or MD); (3) a public, private or suasive verb followed by an adjective (J.*), an adverb (RB), a determiner (DT, QUAN, CD) or a possessive pronoun (PRPS), and then a noun (N.*), and then a verb (V.* or MD), with the possibility of an intervening adjective (J.*) between the noun and its preceding word. This tag corresponds to Biber's (1988: 244) category but its operationalisation has been improved to avoid the algorithm erroneously tagging constructions such as *Why would I know that?* and *He didn't hear me thank God.* | Finite verbs | Le Foll, adapted from Biber (1988) |
| Discourse organisation | WH subordinate clauses | **WHSC** | *I'm thinking of someone **who** is not here today. Do you know **whether** the banks are open?* | Assigned when the words *what, where, when, how, whether, why, whoever, whomever, whichever, wherever* and *whenever* have not been previously identified as part of a WH question (WHQU). Though many attempts were made, it proved impossible to reliably disambiguate between relative and other subordinate WH-clauses, which is why they are pooled together in this category. | Finite verbs | Le Foll |

| | | | | | | |
|---|---|---|---|---|---|---|
| Lexis | Total nouns (including proper nouns) | **NN** | *a **cut,** my **coat,** the **findings,** cruelty **, comprehension,** on **Monday** 6 **Aug,** the **U.S.** , on the **High Street*** | Assigned to all singular (NN) and plural nouns (NNS) identified by the Stanford Tagger including proper nouns (NNP and NNPS). This variable differs from the Biber Tagger in that it includes nominalisations. | Words | Le Foll |
| Lexis | Noun compounds | **NCOMP** | *Surely this stone must be the last one to cover the dungeon **entrance** ! Experts say that the rare winter **phenomenon** is a natural occurrence.* | Assigned when two or more nouns follow each other without any intervening punctuation. The algorithm allows for the first noun to be a proper noun but not the second thus allowing for *Monday afternoon* and *Hollywood stars* but not *Barack Obama* and *Los Angeles* . It is also restricted to nouns with a minimum of two letters to avoid OCR errors (dots and images identified as individual letters and which are usually tagged as nouns by the Stanford Tagger) producing too many erroneous NCOMP's. Note that this feature works best with fully punctuated texts (see per-register recall and precision rates in the tagger documentation). | Nouns | Le Foll |
| Lexis | Emoji and emoticons | **EMO** | 😍 🥰 🌈 :-) :DD XD <3 :/ | Assigned to all emojis as of December 2018 (cf. https://unicode.org/emoji/charts/full-emoji-list.html) and to a range of emoticons, in particular three-character emoticons such as *:-)* . The source code also includes three lines which are by default commented out but can be uncommented for texts where short emoticons are expected. It is not recommended to use these lines for general English because they lead to a sharp decrease in precision: many of the shorter emoticons, e.g., *:( :D :3* , are too easy to confuse with poorly scanned texts that are missing spaces, or with the punctuation styles of specific academic journals. | Words | Le Foll |
| Lexis | Hashtags | **HST** | ***#phdlife #Buy1Get1Free*** | Assigned to any string starting with a hashtag followed by at least three letters, digits or underscores. | Words | Le Foll |
| Lexis | URL and e-mail addresses | **URL** | ***www.faz.net https://twitter.com elefoll@uos.de*** | Assigned to all strings resembling a URL or an e-mail address (without claiming to only include valid URLs or e-mail addresses since this is not the aim). Regex for this feature was inspired by: https://mathiasbynens.be/demo/url-regex | Words | Le Foll |
| Negation | Negation | **XX0** | *Why do **n't** you believe me? There is **no** way that's happening any time soon. **Nor** am I.* | Biber's (1988) analytic and synthetic negation features were merged into one negation variable since the latter is too infrequent to be of use in the context of this study. This unique negation tag is assigned to the tokens *not_RB, n't* _RB, all occurrences of the words *nor* and *neither* , and *no* when followed by an adjective (J.*) or noun (NN.*). | Finite verbs | Le Foll |
| Prepositions | Prepositions | **IN** | *The Great Wall **of** China is the longest wall **in** the world. There are towers **along** the wall. I prefer **to** go **to** an art gallery. The objects **on** display are **from** all **over** the world.* | All items tagged as IN by the Stanford Tagger other than those assigned to CAUS, CONC, COND, OSUB, SO and LIKE. | Words | Le Foll |

| | | | | | | |
|---|---|---|---|---|---|---|
| Pronouns | Reference to the speaker/writer | **FPP1S** | ***I*** *don't know. It isn't* ***my*** *problem.* | All occurrences of *me, myself* and *mine* and *I* if tagged by the Stanford Tagger as a pronoun, a list symbol (LS) or a foreign word (FW). | Finite verbs | Le Foll |
| Pronouns | Reference to the speaker/writer and other(s) | **FPP1P** | ***We*** *were told to deal with it* ***ourselves*** *.* | All occurrences of us, we, our, ourselves and ours, as well as the contracted form of us (e.g., in let's). All these terms are case insensitive but an exception for US was added as this usually refers to the United States of America. | Finite verbs | Le Foll |
| Pronouns | Reference to the addressee | **SPP2** | *If* ***your*** *model was good enough,* ***you*** *'d be able to work it out.* | Following Biber (1988), all occurrences of *you, your, yourself, yourselves* . Following Nini (2014: 18), also includes *thy, thee* and *thyself* . In addition, the forms *ur, ye, y'all, ya* , *thine* and the nominal possessive pronoun *yours* were also added. | Finite verbs | Le Foll, adapted from Nini (2014) |
| Pronouns | *it* pronoun reference | **PIT** | ***It*** *fell and broke. I implemented* ***it*** *.* ***Its*** *impact has not yet been researched.* | All occurrences of the pronoun *it* . An exception was added for the all capital form *IT* which most frequently refers to *Information Technology* . Following Nini (2014: 18), also includes all occurrences of *itself* and *its* . | Finite verbs | Le Foll, adapted from Nini (2014) |
| Pronouns | *One* as a personal pronoun | **PRP** | ***One*** *would hardly suppose that your eye was as steady as ever.* | This tag consists of the remaining personal pronouns not yet tagged as either first (FPP1S and FPP1P), second (SPP2) or third (TPP3) person pronouns. In practice, this should only leave *one* . | Finite verbs | Le Foll |
| Pronouns | Reference to one non-interactant | **TPP3S** | ***He*** *is beginning to form* ***his*** *own opinions.* ***She*** *does tend to keep to* ***herself*** *.* | Following Biber (1988), all occurrences of *she, he, her, him, his, himself, herself* and *themself* . Note that the singular *they* form can only be accounted for with the possessive pronoun: *themself.* | Finite verbs | Le Foll |
| Pronouns | Reference to more than one non-interactant | **TPP3P** | *The text allows readers to grapple with* ***their*** *own conclusions. I wouldn't trust* ***them*** *.* | All occurrences of *they, them, themselves, theirs* and *em* when tagged by the Stanford Tagger as a pronoun. | Finite verbs | Le Foll |
| Pronouns | Quantifying pronouns | **QUPR** | *said Alice aloud, addressing* ***nobody*** *in particular.* | All occurrences of *anybody, anyone, anything, each other, everybody, everyone, everything, nobody, none, no one, nothing, somebody, someone* and *something.* | Finite verbs | Nini (2014) |
| Stance-taking devices | Politeness markers | **POLITE** | *Can you open the window,* ***please*** *? Would you* ***mind*** *giving me a hand? I was* ***wondering*** *whether you could help.* | Assigned to all occurrences of *thanks, thank you, cheers, ta* (unless it is preceded by *got* to avoid the confusion with *gotta* ), *please, sorry, apology, apologies,* all forms of the verbs *excuse, I/we wonder, I/we + BE + wondering,* and the multi-word units *you mind* and *don't mind* . No exception was made for *please* as a verb because the Stanford Tagger frequently misidentifies *please* as a verb, e.g., *I was like please_VPRT just please_VB just get there.* | Words | Le Foll |

| | | | | | | |
|---|---|---|---|---|---|---|
| Stance-taking devices | Amplifiers | **AMP** | *I am **very** tired. They were both **thoroughly** frightened.* | Assigned to the amplifiers from Biber's (1988) list: *absolutely, altogether, completely, enormously, entirely, extremely, fully, greatly, highly, intensely, perfectly, strongly, thoroughly, totally, utterly, very. Especially* was added. | Words | Le Foll, adapted from Biber (1988) |
| Stance-taking devices | Downtoners | **DWNT** | *These tickets were **only** 45 pounds. It's **almost** time to go.* | Assigned to all occurrences of *almost, barely, hardly, merely, mildly, nearly, only, partially, partly, practically, scarcely, slightly, somewhat* . In Biber (1988) *almost* is listed as both a hedge and a downtoner. Following Nini (2014), it is only considered a downtoner here. | Words | Nini (2014) |
| Stance-taking devices | Emphatics | **EMPH** | *I **do** wish I hadn't drunk quite **so** much. Oh **really** ? I **just** can't get my head around it.* | Following Biber (1988), assigned to all occurrences of *just, really, most, more, real* + ADJ, *so* + ADJ, *for sure, such a* . The algorithm was improved by adding *so + much/little/many, such a/* an (whilst excluding *such a/an* if proceeded by *of* ), and ensuring that only *DO* + verb in base form (VB) are tagged. *Least* and *far* + J.*/RB were added (the latter only when not proceeded by *so* or *thus* ). To account for recent language change (Aijmer 2018), *bloody, dead* + ADJ, *fucking* and *super* were also added. Multi-word units are counted as one EMPH tag but several Words. | Words | Le Foll, adapted from Biber (1988) |
| Stance-taking devices | Hedges | **HDG** | *There seemed to be no **sort** of chance of getting out. I wish that **kind** of thing never happened. She's **maybe** gonna do it.* | Following Biber (1988: 240) assigned to all occurrences of *maybe, at about, something like,* and *more or less,* as well as *sort of* and *kind of* as long as they are not preceded by a determiner (DT), quantifier (QUAN), cardinal number (CD), adjective (J.*), possessive pronoun (PRPS) or WH word. The condition that *kind* must have been tagged as a noun (NN) by the Stanford Tagger was added to exclude phrases such as *it's very kind of you* . *Kinda* and *sorta* was added as colloquial alternatives to *kind of* and *sort of* and the adverbs *apparently* , *conceivably, perhaps, possibly, presumably, probably, roughly* and *somewhat* were also added to the list. | Words | Le Foll, adapted from Biber (1988) |
| Stative forms | Existential *there* | **EX** | ***There** are students. And **there** is now a scholarship scheme.* | As tagged by the Stanford Tagger: "Existential *there* is the unstressed *there* that triggers inversion of the inflected verb and the logical subject of a sentence" (p. 3). | Finite verbs | Le Foll |
| Stative forms | Be as main verb | **BEMA** | *It was nice to just **be** at home. She'**s** irreplaceable. It'**s** best I think. How **was** your mum on Sunday? It'**s** not long.* | Following Biber (1988), this tag is assigned to the all forms of the verb be when followed by a determiner (DT), a possessive pronoun (PRPS) a preposition (IN), or an adjective (JJ). In addition, Nini (2014: 20) improved the Biber Tagger "by taking into account that adverbs or negations can appear between the verb BE and the rest of the pattern. Furthermore, the algorithm was slightly modified and improved: (a) the problem of a double-coding of any Existential there followed by a form of BE as a BEMA was solved by imposing the condition that there should not appear immediately before or two before the pattern; (b) the cardinal numbers (CD) tag and the personal pronoun (PRP) tag were added to the list of items that can follow the form of BE." This latter improvement by Nini, however, resulted in tag questions also being assigned to BEMA. The present algorithm therefore further excludes any occurrences of BE found one or two to the left of a question tag (QUTAG), as well as BE occurrences one or two to the left of a present participle form tagged as PROG or past participle form tagged as PASS. | Finite verbs | Le Foll, adapted from Nini (2014) |

| | | | | | | |
|---|---|---|---|---|---|---|
| Syntax | Split auxiliaries and infinitives | **SPLIT** | *I would actually **drive** . You can just so **tell** . I can't ever **imagine** arguing with Jill.* | This category merges Biber's (1988) split auxiliaries and split infinitive categories and follows Nini's (2014: 30) operationalisations. Hence, this tag is assigned every time the infinitive marker to (TO) is followed by one or two adverbs and a verb base form, and every time an auxiliary (any modal verb MD, or any form of DOAUX, or any form of BE, or any form of HAVE) is followed by one or two adverbs and a verb form. Nini's algorithm was improved to ensure that negated split auxiliaries would also be identified, e.g., They have not yet developed cancer. | Finite verbs | Le Foll, adapted from Nini (2014) |
| Syntax | Stranded prepositions | **STPR** | *We've got more than can be accounted **for** . Open the door and let them **in** . Where is it **from** ? It's not the sort of music we're **into** .* | As in Biber (1988), assigned to the prepositions *against, amid, amidst, among, amongst, at, between, by, despite, during, except, for, from, in, into, minus, of, off, on, onto, opposite, out, per, plus, pro, than, through, throughout, thru, toward, towards, upon, versus, via, with, within* and *without* followed by any punctuation mark. Following Nini (2014: 30), *besides* was removed from Biber's original list since it also frequently serves as a conjunct and, in this function, is usually followed by a punctuation mark. Note that Nini's (2014:30) operationalisation tagged all occurrences of these word forms as prepositions regardless of how they were tagged by the Stanford Tagger. Here, it was decided to improve accuracy by restricting the query to tokens tagged as IN by the Stanford Tagger (thus excluding many RB and RP tokens, e.g., *Don't take it away ! Tie her up ! He roared out : "Come away !"* ). | Finite verbs | Le Foll, adapted from Nini (2014) |
| Verb features | Verbal contractions | **CONT** | *I do **n't** know. It is **n't my** problem. You **'ll** have to deal with it.* | Following (Nini 2014: 29), all occurrences of an apostrophe followed by a word identified as a verb (V.*, MD) by the Stanford Tagger and all occurrences of the token *n't* _XX0. | Finite verbs | Nini (2014) |
| Verb features | Particles | **RP** | *I'll look it **up.** It's coming **down** . When will you come **over** ? Some of the birds hurried **off** at once.* | As tagged by the Stanford Tagger (RP) (Santorini 1990: 9-10). | Finite verbs | Le Foll |
| Verb features | *BE* -passives | **PASS** | *He must have been **burgled** . They need to be **informed** . He was **found** out. When were they **arrested** ?* | Assigned to past participles (here: VBN or VBD) preceded by the following patterns: 1) any form of the verb BE; 2) BE followed by one or two adverb(s) (RB) and/or a negation (XX0); 3) BE followed by a noun (NN.*) or personal pronoun (PRP); 4) BE followed by a noun (NN.*) or personal pronoun, and an adverb (RB) or negation (XX0). Unlike Biber (1988), no subdivision is made for by-passives and agentless passives. This choice is a) theoretically motivated because passives are too infrequent to be robustly measured at this level of granularity in most texts and b) for practical reasons because the algorithm proposed to identify by-passives resulted in too many false positives (e.g., looking for things that have been made by hand). | Finite verbs | Le Foll |
| Verb features | *GET*- passives | **PGET** | *He's gonna get **sacked.** She'll get me **executed** . It gets **done** all the time.* | Assigned to past participles (here: VBN or VBD) preceded by the following patterns: 1) any form of the verb GET; 2) GET followed by a noun (NN.*) or personal pronoun (PRP); 3) GET followed by a determiner (DT) or a noun (NN.*) plus a noun (NN.*). | Finite verbs | Le Foll |

| | | | | | | |
|---|---|---|---|---|---|---|
| Verb features | *Going to* constructions | **GTO** | *I'm not **gonna** go. You're **going** to absolutely love it there! **Gonna** come along?* | Assigned to all occurrences of *going to* and *gonna* followed by a base form verb (VB), allowing for up to one intervening word between *going to* or *gonna* and the infinitive. GTO constructions are excluded from the progressive (PROG) count. | Finite verbs | Le Foll |
| Verb features | Past tense | **VBD** | *It **fell** and **broke**. I **implemented** it. If I **were** rich.* | As tagged by the Stanford Tagger, except where VBD tags are assumed to have been misassigned by the Stanford Tagger and are instead attributed to the perfect aspect (PEAS), passives (PASS, PGET) or USEDTO categories. | Finite verbs | Le Foll |
| Verb features | Non-finite verb *-ing* forms | **VBG** | *He texted me **saying** no. He just started **laughing**. I remember **thinking** about that.* | All verb forms ending in *-ing* as tagged by the Stanford Tagger, except those identified as progressives (PROG) or *going to* constructions (GTO). This category also includes "putative prepositions" ending in *-ing* such as *<u>according</u> to* and *<u>concerning</u> your request* (Santorini 1990: 11). | Finite verbs | Le Foll |
| Verb features | Non-finite *-ed* verb forms | **VBN** | *These include cancers caused by viruses. Our content is grouped into sections called topics. Have you read any of the books mentioned in the blog?* | As tagged by the Stanford Tagger except for the exclusion of tokens identified as instances of the perfect aspect (PEAS), passives (PASS, PGET) and *used to* constructions (USEDTO). Note that according to the Stanford Tagger rules, this category includes "putative prepositions" ending in *-ed* such as *<u>granted</u> that* and *<u>provided</u> that* (Santorini 1990: 11). | Finite verbs | Le Foll |
| Verb features | Imperatives | **VIMP** | ***Let** me know! **Read** the website and **write** the names of the characters. In groups, **share** your opinion. Always **do** as you're told!* | This tag is first assigned to any verb in base form (VB) occurring 1) immediately after a punctuation mark except a comma (e.g., Okay: do it!), an emoji or emoticon (EMO), a symbol (SYM), hashtag (HST), foreign word (FW) or a list marker (LS), or 2) after a punctuation mark and an adverb (e.g., 1A. Then practice the dialogue), unless the VB token is please or thank or has previously been identified as a DO auxiliary (DOAUX). In a second loop, the VIMP tag is assigned to VB verb tokens (except thank or please) when preceded by an imperative as identified above, with up to two optional intervening tokens, and the tokens and or or (e.g., Describe or draw, Listen carefully and repeat, Read the text and answer the questions). In addition, a number of verbs frequently found in instructions are listed as exceptions (e.g., Complete, Choose, Check) and are always assigned to this category when they are found at the beginning of a sentence regardless of their tag because these were found to be frequently erronouesly identified by the Stanford Tagger as nouns (NN). | Finite verbs | Le Foll |
| Verb features | Present tense | **VPRT** | *It'**s** ours. Who **does** n't love it? I **know**.* | Subsumes the VBP (present tense other than third-person singular) and VBZ (third-person singular present tense) tags assigned by the Stanford Tagger. The MFTE also corrects systematic errors in the Stanford Tagger output by adding VPRT tags in strings such as *I dunno* and *there's.* | Finite verbs | Le Foll, adapted from Nini (2014) |

| | | | | | | |
|---|---|---|---|---|---|---|
| Verb features | Perfect aspect | **PEAS** | *Have you **been** on a student exchange? She'd already **seen** it. He has been **told** before. Is this the last novel you've **read**?* | Assigned to past participles (VBN, VBD) preceded by the following patterns: 1) any form of the verb HAVE; 2) HAVE followed by one or two adverb(s) (RB) and/or a negation (XX0); 3) HAVE followed by a noun (NN.*) or personal pronoun (PRP); 4) HAVE followed by a noun (NN.*) or personal pronoun, and an adverb (RB) or negation (XX0); 5) HAVE followed by a participle tagged as a passive (PASS); 6) HAVE followed by one or two adverb(s) (RB) and/or a negation (XX0), and a passive participle (PASS); 7) HAVE followed by a noun (NN.*) or personal pronoun (PRP), and a passive participle (PASS); 8) *'s* as a verb (VBZ) followed by *been, had, done* or a stative verb; 9) 's as a verb (VBZ) followed by an adverb (RB) or negation (XX0), and *been, had, done* or a stative verb (as listed under JJPR). | Finite verbs | Le Foll |
| Verb features | Progressive aspect | **PROG** | *He wasn't **paying** attention. I'm **going** to the market. I'm **guessing** you're not going to be alone. I must be **getting** home.* | Assigned to any form of BE followed by an *-ing* form of any verb (VBG). The algorithm allows for an intervening adverb (RB), emphatic (EMPH) and/or negation (XX0). The interrogative form is captured as BE followed by a noun (N.*) or personal pronoun (PRP) followed by the VBG token. As for the affirmative version, the latter algorithm also accounts for an intervening adverb (RB) and/or negation (XX0). *Going to* constructions are excluded from this category and are tagged separately (GTO). | Finite verbs | Le Foll |
| Verb features | *HAVE got* constructions | **HGOT** | *He's **got** some. I haven't **got** any.* | Assigned to the word *got* preceded by the following patterns: 1) any form of the verb HAVE; 2) HAVE followed by one or two adverb(s) (RB) and/or a negation (XX0); 3) HAVE followed by a noun (NN, NNP) or personal pronoun (PRP); 4) HAVE followed by a noun (NNP, NNP) or personal pronoun, and an adverb (RB) or negation (XX0). Note that this algorithm overwrites the perfect aspect (PEAS) and passive (PASS) tag. | Finite verbs | Le Foll |
| Verb semantics | DO auxiliary | **DOAUX** | *Should take longer than it **does** . Ah you **did** . She needed that house, **did** n't she? You **don** 't really pay much attention, **do** you? Who **did** not already love him.* | Assigned to do, does and did as verbs in the following patterns: (a) when the next but one token is a base form verb (VB) (e.g., *did it work?, didn't hurt?* ); (b) when the next but two token (+3) is a base form verb (VB) (e.g., *didn't it work)* ; (c) when it is immediately followed by an end-of-sentence punctuation mark (e.g., *you did?* ); (d) when it is followed by a personal pronoun (PRP) or *not* or *n't* (XX0) and an end-of-sentence punctuation mark (e.g., *do you? He didn't!* ); (e) when it is followed by *not* or *n't* (XX0) and a personal pronoun (PRP) (e.g., *didn't you?* ); (f) when it is followed by a personal pronoun followed by any token and then a question mark (e.g., *did you really? did you not?* ); (g) when it is preceded by a WH question word. Additionally, all instances of DO immediately preceded by *to* as an infinitive marker (TO) are excluded from this tag. | Finite verbs | Le Foll |
| Verb semantics | Activity verbs | **ACT** | *I got up and **ran** out. **Bring** your CV. Where have you **worked** before? I **go** to school.* | Assigned to all forms of the verbs: *buy, make, give, take, come, use, leave, show, try, work, move, follow, put, pay, bring, meet, play, run, hold, turn, send, sit, wait, walk, carry, lose, eat, watch, reach, add, produce, provide, pick, wear, open, win, catch, pass, shake, smile, stare, sell, spend, apply, form, obtain, arrange, beat, check, cover, divide, earn, extend, fix, hang, join, lie, obtain, pull, repeat, receive, save, share, smile, throw, visit, accompany, acquire, advance, behave, borrow, burn, clean, climb, combine, control, defend, deliver, dig, encounter, engage, exercise, expand, explore* and *reduce* (cf. Biber 2006: 246, based on the LGSWE, pp. 361–362, 367–368, 370). *Do* is only included when it has not previously been tagged as an auxiliary (DOAUX). *Get* and *go* were removed from Biber's (2006) list due to their high polysemy. Like Biber (2006), for practical reasons, no phrasal verbs were included in this variable. | Finite verbs | Le Foll, based on Biber (2006) |

| | | | | | | |
|---|---|---|---|---|---|---|
| Verb semantics | Aspectual verbs | **ASPECT** | *You should just **keep** talking. I **started** early today.* | Following Biber (2006: 247, based on the LGSWE, pp. 364, 369, 371), assigned to all forms of the verbs: *start, keep, stop, begin, complete, end, finish, cease* and *continue*. | Finite verbs | Biber 2006 |
| Verb semantics | Facilitation and causative verbs | **CAUSE** | *He **helped** her escape. I pleaded with her to **let** me go.* | Following Biber (2006: 247, based on the LGSWE, pp. 363, 369, 370), assigned to all forms of the verbs: *help, let, allow, affect, cause, enable, ensure, force, prevent, assist, guarantee, influence, permit* and *require*. | Finite verbs | Biber 2006 |
| Verb semantics | Communication verbs | **COMM** | ***Describe** it to your partner and **say** why. **Write** a list. **Say** what these words mean.* | Following Biber (2006: 247, based on the LGSWE, pp. 362, 368, 370), assigned to all forms of the verbs: *say, tell, call, ask, write, talk, speak, thank, describe, claim, offer, admit, announce, answer, argue, deny, discuss, encourage, explain, express, insist, mention, offer, propose, quote, reply, shout, sign, sing, state, teach, warn, accuse, acknowledge, address, advise, appeal, assure, challenge, complain, consult, convince, declare, demand, emphasize, excuse, inform, invite, persuade, phone, pray, promise, question, recommend, remark, respond, specify, swear, threaten, urge, welcome, whisper* and *suggest.* British spellings and the verbs *agree, assert, beg, confide, command, disagree, object, pledge, pronounce, plead, report, testify, vow* and *mean* were added. The latter was on Biber's (2006) list for mental verbs but, in most contexts encountered in the present study, it was found to be more likely to be a communication verb. | Finite verbs | Le Foll, based on Biber (2006) |
| Verb semantics | Existential or relationship verbs | **EXIST** | *Weren't they **representing** Jamaica? It encouraged young athletes to **stay** .* | Following Biber (2006: 247, based on the LGSWE, pp. 364, 369, 370–371), assigned to all forms of the verbs: *seem, stand, stay, live, appear, include, involve, contain, exist, indicate, concern, constitute, define, derive, illustrate, imply, lack, owe, own, possess, suit, vary, deserve, fit, matter, reflect, relate, remain, reveal, sound, tend* and *represent* . This variable does not include the copular *be* . *Look* was removed from Biber's original list because it frequently acts as an activity verb, too, e.g., *I was looking for my glasses.* | Finite verbs | Le Foll, based on Biber (2006) |
| Verb semantics | Mental verbs | **MENTAL** | *We **want** to **see** you tomorrow. Did you never **hear** back? I don't **recognize** any.* | Following Biber (2006: 246-247, based on the LGSWE, pp. 362–363, 368–369, 370), assigned to all forms of the verbs: *see, know, think, want, need* (unless identified as a necessity modal; cf. MDNE), *feel, like, hear, remember, believe, read, consider, suppose, listen, love, wonder, understand, expect, hope, assume, determine, agree, bear, care, choose, compare, decide, discover, doubt, enjoy, examine, face, forget, hate, identify, imagine, intend, learn, mind, miss, notice, plan, prefer, prove, realize, recall, recognize, regard, suffer, wish, worry, accept, appreciate, approve, assess, blame, bother, calculate, conclude, celebrate, confirm, count, dare, detect, dismiss, distinguish, experience, fear, forgive, guess, ignore, impress, interpret, judge, justify, observe, perceive, predict, pretend, reckon, remind, satisfy, solve, study, suspect* and *trust.* British spellings were added. *Afford* and *find* which can be found on Biber's original list, were removed for being too polysemous. Note that the phrase *dunno* , which is incorrectly parsed by the Stanford Tagger, was also retagged as *du_VPRT n_XX0 no_VB* and that *no_VB* tokens are also assigned to this category. | Finite verbs | Le Foll, based on Biber (2006) |

| | | | | | | |
|---|---|---|---|---|---|---|
| Verb semantics | Occurrence verbs | **OCCUR** | *Couldn't have* ***happened*** *at a busier time! The cricket* ***lasts*** *all day.* | Following Biber (2006: 247, based on the LGSWE pp. 364, 369, 370), assigned to all forms of the verbs: *become, happen, change, die, grow, develop, arise, emerge, fall, increase, last, rise, disappear, flow, shine, sink, slip* and *occur* . | Finite verbs | Biber 2006 |
| Verb semantics | Necessity modals | **MDNE** | *I really* ***must*** *go.* ***Should*** *n't you be going now? You* ***need*** *not have worried. Everybody* ***needed*** *to be needed.* | As in Biber (1988), all occurrences of *ought, should* and *must.* Contrary to Nini's operationalisation (2014: 27), only occurrences tagged as modals (MD) by the Stanford Tagger were included. In addition, *need* when tagged as a modal by the Stanford Tagger (mostly when followed by *not* or *n't* ) or when immediately followed by *to* not tagged as a preposition (IN) was also added to this variable. | Finite verbs | Le Foll, adapted from Biber (1988) |
| Verb semantics | Modal *can* | **MDCA** | ***Can*** *I give him a hint? You* ***can*** *not. I* ***ca*** *n't believe it!* | All occurrences of *can* and c*a* tagged as modals by the Stanford Tagger (MD). *Ca* was included because the Stanford Tagger parses *can't* as *ca + n't.* | Finite verbs | Le Foll |
| Verb semantics | Modal *could* | **MDCO** | *Do you think someone* ***could*** *have killed her? Well, that* ***could*** *be the problem.* ***Could*** *you do it by Friday?* | All occurrences of *could* tagged as a modal by the Stanford Tagger (MD). | Finite verbs | Le Foll |
| Verb semantics | Modals *may* and *might* | **MDMM** | ***May*** *I have a word with you? But it* ***might*** *not be enough.* | All occurrences of *may* and *might* tagged as modals by the Stanford Tagger (MD). | Finite verbs | Le Foll |
| Verb semantics | *will* and *shall* modals | **MDWS** | *It* ***wo*** *n't do. Yes it* ***will*** *.* ***Shall*** *we see?* | The tokens *will* and *shall* and their contractions *'ll, wo* and *sha* when tagged as modals by the Stanford Tagger (MD). | Finite verbs | Le Foll |
| Verb semantics | modal *would* | **MDWO** | ***Would*** *n't you like to know? If I could afford to buy it I* ***would*** *. I* ***'d*** *like to think it works.* | The tokens *will* and *shall* and their contractions *'ll, wo* and *sha* when tagged as modals by the Stanford Tagger (MD). | Finite verbs | Le Foll |
| Verb semantics | *be able to* | **ABLE** | *It should be* ***able*** *to speak back to you. Would you be* ***able*** *to?* | Assigned to occurrences of the bigram (un)able to, whenever (un)able has previously been identified as a predicative adjective (JJPR). These occurrences of (un)able are subsequently excluded from the JJPR count. | Finite verbs | Le Foll |
| | | | | **Tags not counted by MFTE but important to understand the operationalisation of other features** | | |
| Lexis | Foreign words | **FW** | *I chose turkish delight and* ***panna cotta . Merrry christmasss! Yo im*** *gonna love it!* | All remaining words tagged by the Stanford Tagger as foreign words and not identified as other variables by the MFTE. Frequently includes words spelt with non-standard spellings, missing apostrophes, and poorly OCR'ed due to unusual fonts. Note that this feature is not counted by the MFTE. | NA | Stanford Tagger |
| Lexis | Symbols | **SYM** | ***â*** *2 € a go. I hope so* ***†*** *. That's* ******* *all* ******* *they said!* | All remaining non alphanumeric tokens tagged by the Stanford Tagger as symbols (SYM) or list markers (LS) and not identified as other variables by the MFTE. Also frequently includes words poorly OCR'ed due to unusual fonts or poorly encoded text. Note that this feature is not counted by the MFTE. | NA | Stanford Tagger |

| | | | | | | |
|---|---|---|---|---|---|---|
| Verb features | *to* -infinitives | **TO** | *They were trying **to** find a solution. We like **to** think it's doable. I went in there **to** kinda like celebrate.* | Following Nini (2014: 21), all occurrences of *to* except when followed by another _IN token, a number (CD), determiner (DT), adjective (J.*), possessive pronoun (PRPS), WH-word (WPS, WDT, WP, WRB), pre-determiner (PDT), noun (N.*) or pronoun (PRP). Note that, unlike Nini (2014), this feature is only used to identify other linguistic features. All occurrences of *to* are counted as prepositions (IN) in the MFTE output tables. | NA | Nini (2014) |
| Verb features | Verb base form | **VB** | *She would **sit** and **read** most afternoons. What do you **use** it for? Ask your parents to **drive** you to your friend's house.* | As tagged by the Stanford Tagger, except those identified as imperatives (VIMP). This feature is not included in the tables of counts outputted by the MFTE because it overlaps with other features (e.g., all the modal verb features). However, it is used to identify many other linguistic features. | NA | Le Foll |
| Verb semantics | Private verbs | NA | *I don't **think** this should be **assumed**. I **suspect** he can't even **remember** it.* | As in Biber (1988, based on 1985: 1181), all forms of the verbs accept, *anticipate, ascertain, assume, believe, calculate, check, conclude, conjecture, consider, decide, deduce, deem, demonstrate, determine, discern, discover, doubt, dream, ensure, establish, estimate, expect, fancy, fear, feel, find, foresee, forget, gather, guess, hear, hold, hope, imagine, imply, indicate, infer, insure, judge, known, learn, mean, note, notice, observe, perceive, presume, presuppose, pretend, prove, realize, reason, recall, reckon, recognize, reflect, remember, reveal, see, sense, show, signify, suppose, suspect, think* and *understand.* Note that this category is only used to identify *that*- omissions (THATD). | NA | Biber 1988 |
| Verb semantics | Public verbs | NA | *She **promised** she'd **write** back.* | As in Biber (1988, based on 1985: 1181), all forms of the verbs *acknowledge, add, admit, affirm, agree, allege, announce, argue, assert, bet, boast, certify, claim, comment, complain, concede, confess, confide, confirm, contend, convey, declare, deny, disclose, exclaim, explain, forecast, foretell, guarantee, hint, insist, maintain, mention, object, predict, proclaim, promise, pronounce, prophesy, protest, remark, repeat, reply, report, retort, say, state, submit, suggest, swear, testify, vow, warn* and *write.* Note that this category is only used to identify *that* -omissions (THATD). | NA | Le Foll, adapted from Biber (1988) |
| Verb semantics | Suasive verbs | NA | *They were **determined** to make this work. I'd **prefer** to do it that way.* | As in Biber (1988, based on 1985: 1182–3), all forms of the verbs *agree, allow, arrange, ask, beg, command, concede, decide, decree, demand, desire, determine, enjoin, ensure, entreat, grant, insist, instruct, intend, move, ordain, order, pledge, pray, prefer, pronounce, propose, recommend, request, require, resolve, rule, stipulate, suggest, urge* and *vote* . Note that this category is only used to identify *that* -omissions (THATD). | NA | Biber 1988 |
| | | | | **Features removed from the MFTE feature portfolio post-evaluation of v.2.9 (Note that the corresponding lines are commented out in v.3.0+ and may still be run, if wished)** | | |
| Verb semantics | Quotative BE + *like* | **QLIKE** | *I was **like** oh this is really good. And everyone is **like** let's do this.* | Assigned to any form of *BE* followed by *like* tagged as a preposition (IN) by the Stanford Tagger and not followed by a noun (NN.*), adjective (J.*), determiner (DT), preposition (IN) or a full stop, comma, exclamation or question mark. This feature is deactivated by default but can be uncommented in the script. | Finite verbs | Le Foll |

| | | | | | | |
|---|---|---|---|---|---|---|
| Verb features | *Used to* constructions | **USEDTO** | *You'll get **used** to it in time. It works but not like it **used** to .* | Assigned to all occurrences of the bigram *used to* . These occurrences of *used* are excluded from the VBN/VBD counts. This feature is deactivated by default but can be uncommented in the script. | Finite verbs | Le Foll |
| Discourse organisation | Phrasal coordination | **PHC** | *read **and** write, positive **or** negative, nouns **and** adjectives* | All occurrences of *and, &, or* and *nor* in the following patterns: adverb + *and/or/nor* + adverb, adjective + *and/or/nor* + adjective, verb + *and/or/nor* + verb, noun + *and/or/nor* + noun. This feature was removed post-evaluation and these occurrences of *and* and *or* are now all included in coordinating conjunctions (CC). | NA | Biber 1988 |