

UDK: 811'322

811'373.47

Izvorni znanstveni rad

Rukopis primljen 5. XII. 2022.

Prihvaćen za tisak 24. V. 2023.

<https://doi.org/10.31724/rihjj.49.1.4>

Anna Bączkowska

Institute of English and American Studies

Wita Stwosza Street 51

PL-80-308 Gdańsk

Eviden, Big Data and Security

<https://orcid.org/0000-0002-0147-2718>

anna.baczkowska@ug.edu.pl

Dagmar Gromann

Centre for Translation Studies, University of Vienna

Gymnasiumstrasse 50, AT-1190 Vienna

<https://orcid.org/0000-0003-0929-6103>

dagmar.gromann@univie.ac.at

FROM KNOBHEAD TO SEX GODDESS: SWEAR WORDS IN ENGLISH SUBTITLES, THEIR FUNCTIONS AND REPRESENTATION AS LINGUISTIC LINKED DATA

Swear words represent an important social vehicle for human communication that beyond mere insults are conventionally used for social bonding and bantering among other functions. However, to the best of our knowledge, no systematic typology of swear word functions has been proposed. In this article, such a typology is proposed in a top-down manner drawing on literature as well as bottom-up by analysing a concrete corpus of real-world sufficiently filthy dialogues. For this first case study, the analysis is limited to English and the analysis of subtitles of a single movie. We found that specific types of swear words, e.g. bodily functions, appear across functions with some preferences, e.g. body parts are particularly utilised in jocularity, criticism, and anger. Furthermore, the resulting data are represented as linguistic linked data, extolling the virtues of this format for fine-grained linguistic analyses, e.g. filtering and visualising all swear words pertaining to a specific function.

1. Introduction

Originally, swear words represented a negative function of insult, slant or even expose faults of another person. As such, swearing has been classified as offensive language or even abusive (Andersson and Trudgill 2007: 197) and as an important vehicle to express strong emotions, particularly anger and aggression. However, swearing can also be used in a positive and potentially non-offensive way as a form of expressing humour (Andersson and Trudgill 2007: 197). While several publications propose a detailed account on one, two or three specific function(s) (e.g. Daly et al. 2004), we propose a systematic typology of twelve unique swear word functions that is empirically validated by annotating swear words in English subtitles of *Bridget Jones' Diary*.

This research focuses on the question of which types of functions swear words can fulfil in context, i.e., dialogues, how their degree of offensiveness varies, and whether proposed functions in literature are sufficient to classify all of their occurrences. In literature, either individual or comparative accounts of functions are provided, which at times conflict. As a first step, we collected swear word functions and their definitions from literature to create an initial typology in a top-down manner. In the chosen subtitle corpus, swear words were detected automatically, and two domain experts annotated them according to their function. In this second, bottom-up step, it became clear that the initial typology needed to be adapted and refined, the result of which we present in this article. Furthermore, this step clearly showed that one and the same swear word might fulfil a number of different functions within the context of a single movie. To the best of our knowledge, this is the first systematic typology of swear word functions in context that has additionally been empirically validated, albeit only on a single, yet filthily rich movie as a first case study. This typology might not only spark future theoretical discussions on systemizing swear word functions but has direct practical implications for classifying swear words. For instance, approaches to automatically detect swear words conventionally focus on profane language or hate speech, including all of the swear words discussed in this article. However, as we show in this article, swear words might not necessarily be profane or hate speech.

To ensure reusability of both the proposed typology of functions as well as the annotated swear words, both resources are represented as Linguistic Linked Open Data (LLOD). Thereby, a use case is provided to existing LLOD modelling approaches as well as an incentive for linguists investigating offensive language to benefit from the interoperability and ease of reuse of LLOD resources. The approach is interoperable in the sense of being able to map from one swear word annotated dataset to another, allowing for analysis across corpora, text genres, languages, and corpus formats. For instance, swear words grouped by function can be analysed with a simple query or frequencies of functions and thus swear words can be compared. Furthermore, the typology of swear words can easily be reused for annotating other datasets. For instance, as future work we plan to extend this annotation effort to translations of subtitles to, e.g. Polish and German. The basis for this representation is the OntoLex-Lemon model (Cimiano et al. 2016) and its extension Module for Frequency, Attestation and Corpus Information (OntoLex-FrAC)¹ to represent corpus data.

2. Swear words defined

Swearing is a linguistic practice that involves the use of social, religious or cultural taboos that is some ‘forbidden’ words (Ljung 2011: 5). It typically capitalizes on words revolving around sex or bodily parts and functions, in particular excretory (e.g. *piss*) and scatological aspects (e.g. *shit*), parts of the body and genitals (*ass*, *cunt*), sexual practices (*fuck*, *wank*, *having sex*); it also makes reference to objects of religious cult (*God*, *Jesus*) (Stapleton 2010: 289–290). The non-religious swearing involves the use of vernacular words, in lieu of scientific ones, that are vulgar or at least embarrassing; hence *excrement* is replaced by *shit* and *copulate* by *fuck* (Ljung 2011: 7). Most scholars (e.g. Andersson and Trudgill 2007: 195; Ljung 2011: 4) identify three features of bad language: (1) they invoke stigmatised or taboo concepts; (2) they are interpreted non-literally; (3) they convey speaker’s strong emotions or attitude. For other scholars (e.g. McEnery 2006: 27; Singleton 2009: 138–140), the literal use is inclusive of swear words, as in *We fucked*, *Let’s fuck*, or *Man wanna fuck my wife*. Swearing typi-

¹ <https://acoli-repo.github.io/ontolex-frac/>

cally entails negative feelings, expressing criticism, anger or frustration. These reasons for swearing may be classified under the rubric of “abusive swearing” (Andersson and Trudgill 2007: 197). However, along with this prototypical approach, swearing may also be perceived in a more positive way, as a form of expressing humour, and thus being classified as “humorous swearing” (Andersson and Trudgill 2007: 197). In both cases, they are typically employed for emphasis (Dewaele 2016: 113). Furthermore, the same swear words may be perceived by interlocutors as either offensive or non-offensive, depending on the situation, degree of formality of the language used (Jay 2009) as well as the relationship previously established by the interactants. Thereby, what might seem to be a vituperative remark for one addressee in a specific situation, can be approached as a term of affection by another one or in a different situational context, and “words that are not commonly viewed as emotion-laden may acquire emotional connotations in discourse” (Pavlenko 2008: 148). On the other hand, swear words frequently repeated become semantically bleached and lose their original strong, offensive load.

3. Functions of swear words

Some of the most oft-cited functions conveyed by swear words comprise expressing solidarity and bonding with the interlocutor (e.g. Daly et al. 2004), alignment with the community of practice (e.g. Beers Fägersten 2001), or in-group identity strategy in a community of practice (e.g. Stapleton 2010: 297).

To these contexts, the function of catharsis, that is emotional release of tension, is added by some scholars (e.g. Jay 1999). This psychological function has no referential meaning. In linguistics, such swear words are usually classified as expletives; that is, typically stand-alone lexemes, albeit they can also be integrated with a sentence, that illustrate self-talk and are uttered in order to release strong emotions, such as anger, irritation, fear, pain, surprise, etc. In the analysis which follows, the ontology of swear words encompasses both linguistic terms, i.e. swear words and expletives, and for convenience they are subsumed by the superordinate category of swear words.

4. Method for swear word detection and classification²

Extending existing lists of swear words (e.g. proposed by Dewaele 2016), this approach automatically detects swear words in English subtitles by means of a profanity filter. In this chapter, the method of annotating detected swear words with their function and representing the results as LLOD resources is presented.²

4.1. Material

The corpus consists of roughly 9,080 words in 1,500 English sentences from the official subtitles of the movie *Bridget Jones' Diary*.³ As a romantic comedy produced in Great Britain, the movie is rich in number and variety of swear words. Furthermore, translations of the movie and subtitles are available in a wide variety of languages, which is important for our future endeavour of analysing swear word translations and potential shifts in function. In this case study, we analyse swear words in their contexts to determine their function. In the original file format, the SubRip file format SRT, the subtitles are split by timestamp. We developed a Python script to automatically compose these sequences to sentences to provide a context for the swear words.

4.2. Swear word detection

To automatically detect swear words, we employ the Python library Profanity-Filter,⁴ which is equipped with an initial English dictionary of swear words and detects derivative and distorted profane words. We compared this dictionary with existing lists of swear words in literature proposed by Beers Fägersten (2007), Dewaele (2016), McEnery (2006) and Love (2021) as well as the inventory on Wiktionary.⁵ Euphemistic forms of swear words (*Gosh* for *God* or *shoot* for *shit*, etc.) were excluded from our analysis. Contrary to Dewaele's list, the words we analysed involve not only nouns and adjectives but a gamut of other

² https://github.com/dgromann/SwearWords_SubtitleAnalysis

³ <https://yifysubtitles.ws/movie/bridget-jones's-diary/tt0243155>

⁴ <https://pypi.org/project/profanity-filter/>

⁵ https://en.wiktionary.org/wiki/Category:English_swear_words

options as well. This leads to an extension of the Profanity-Filter list by 32 swear words (see Appendix A), which amounts to 407 English swear words automatically checked by the Profanity-Filter.

4.3. Typology of functions in our study

Twelve functions have been initially proposed as a template for our classificatory analysis of the English original subtitles and these comprise: advice, anger, banter, bonding, criticism, evaluation, fear, jocularity, pain, request, ridicule, and surprise.

The category of *evaluation* is understood here in line with Hunston (2004: 157), for whom *evaluation* is an “indication that something is good or bad,” and it is assessed as such on the basis of goal achievement or attempted actions (Partington 2007). Other terms used to indicate *evaluation* comprise, inter alia, attitudinal or affective language (Ochs 1989), *attitude* (Halliday 1994), *stance* (Conrad and Biber 2000), *appraisal*⁶ (Martin 2000; White 2015); in earlier publications, the following terms were also employed: *overtones* (Ullmann, 1962), *connotation* (Lyons, 1977), and *valuation* (Hartman, 1967). There is a preference to use the term valuation in the context of business (in the sense of ‘calculation’), as well as in philosophical publications devoted to axiology (from which axiological linguistics evolved), while evaluation is more common in semantics. *Evaluation* expresses the speaker’s opinion about the target’s behaviour or views; thus, it may refer to situations, events or objects (associated with the target). The aim of *evaluation* is not to hurt the interlocutor, rather to present one’s stand, and, in the case of (potentially) offensive contexts, presenting it in a blunt way. Evaluation should not be seen as synonymous with affective language, as the former focuses on sheer expression of one’s opinion while the latter on emotions and feelings that accompany expressing one’s opinion (Bednarek 2006: 19).

By *criticism* we mean non-jocular remarks that are meant to genuinely hurt and/or denigrate an interlocutor or to describe some negative features of an absent

⁶ In systemic-functional linguistics, evaluation has been studied under the rubric of appraisal, which refers to evaluative statements (monoglossic) or presuppositions (heteroglossic) about the human behaviour (judgments), processes, states or entities (appreciation), as well as emotional response (affect). The term *evaluation* used in this paper is not tantamount to *appraisal*.

target. Criticism differs from anger, as one may be critical of somebody without demonstrating strong emotions through annoyance or fury expressed forcefully and/or in a loud voice, which is typical of *anger*. Moreover, anger may be targeted at an interactant but it may also be self-directed, typically used to release the speaker's emotions. Criticism expressed emotionally may be combined with anger. On the other hand, negative jocular overtones are present in criticism aiming at jibing the target, which will be spanned by the function of *ridicule*. *Jocular* draws on funny remarks for the purpose of entertainment without entailing pretended offensiveness. If funny remarks are jointly constructed humorous exchanges of retorts or repartees, and they rely on pretended disparaging comments, and are geared towards mutual entertainment, then we deal with *banter* (which overlaps with the term *teasing* – see for details below). Banter is appreciated by both interactants, which is often signalled by post-utterance mutual laughter (Haugh 2010: 5). Operating within the mutually agreed upon apparently adversarial yet jocular frame, banter is thus not offensive for the interactants who are well-known to each other (Buglass et al. 2011: 288) as they regard apparently impolite remarks and as obviously untrue (Haugh and Bousfield 2012: 1019). In our analysis, banter is reserved for pretended insults, i.e. “humorous insults” or “jocular mockery” (Haugh and Bousfield 2012: 1100), which may be used for flirting purposes, and is applied only to an exchange, and not to self-talk. Rapport-building and peer-group solidarity oriented, pretended offensive language on the other hand is reserved for *bonding*.

A few words on teasing are in order to elucidate the term. *Teasing* is a concept delineating a playful comment, even benevolent, or provocative, mocking, or even critical and malignant (Haugh 2010; Boxer and Cortés-Conde 1997: 279); it is “a personally addressed remark with a bite often performed in front of a public” (Kotthoff 2007: 271) that may have a prosocial function with “permitted disrespect” aiming at enhancing positive feelings about the addressee and strengthening their relationship (Haugh 2017: 205). Teasing may vary in intensity, it may be placed on a cline from sheer bonding to nipping, with the intermediate stage of biting (Boxer and Cortés-Conde 1997). Thus, it may have a wide remit of functions, from being a token of positive emotions to aggression. In Fig.1, we show our understanding of the term teasing in relation to other terms we distinguish, which is in line with Boxer and Cortés-Conde (1997).

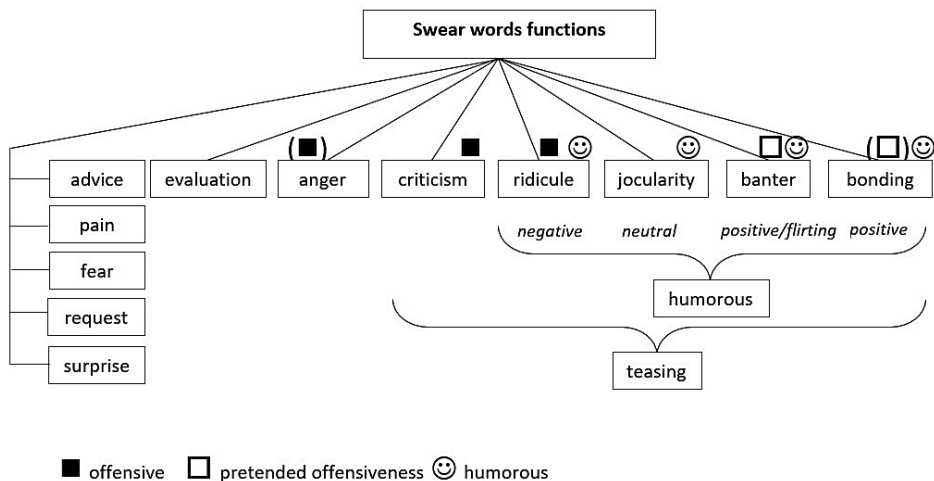


Figure 1: Swear word functions in the proposed schema

To sum up the humorous terms, neutral humorous comments are dubbed *jocularity*, negative humorous remarks stand for *ridicule* (while negative non-humorous for *criticism*), and mutually agreed upon negative humorous exchanges are spanned by the notion of *banter*.

While the terms elucidated above generally pose some definitional problems, and are thus not determined in an undisputable way (hence we propose our own understanding of them), the remaining terms seem self-explanatory, and they are understood in line with lexicographic explanations (the definitions are based on the online Cambridge Dictionary). Thus, *advice* is offering somebody a suggestion of what one should do or how one should act in a given situation. *Request* is asking for something in a polite or official manner. *Surprise* is about the occurrence of an unexpected event. An unpleasant feeling (emotion or thought) one has when being frightened or worried by something that is, or potentially can be, dangerous, painful or bad is dubbed *fear* in our taxonomy. Finally, a feeling of physical or mental suffering resulting from some illness, injury or a situation/event will be used to signify *pain*.

The novelty of the typology presented in Fig. 1 resides in marshalling a number of categories stemming from various accounts in one schema and marking the relations among them, i.e., in trying to systematise them, as well as making an attempt to capture their distinctive features. Whilst the terms have already

received due scholarly attention (see the discussion above), they have been discussed either as separate (albeit complex) notions (e.g. *teasing* described by Boxer and Cortés-Conde 1997), a juxtaposition of two or three functions (e.g. *teasing* vs. *mockery* in Haugh 2010), or as very broad theoretical concepts that are distinguished from various “interaction practices” in which they are used, as in the case of *banter* being used in such practices as *teasing*, *mocking*, *jocularity*, etc. (Haugh and Bousfield 2012), which we find confusing. Moreover, in the proposed schema, we try to pinpoint their crucial distinctive features and/or show the gradability of their meaning. As a result, the schema encompasses a wide range of potential contexts, and the functions are positioned on a cline spanning neutral cases (*advice* and other functions presented vertically in Fig. 1), offensive contexts, pretended offensiveness and humorous contexts. The humorous cases in turn are ascribed some axiological charge. In sum, some existing typologies and definitions have been adapted and inscribed into a more comprehensive schema we offer, which accommodates a wider range of contexts and more semantic features and their gradability.

4.4. Swear word annotation

To annotate the detected swear words with the functions presented in Section 4.3., each swear word and its sentence of occurrence as well as the preceding and subsequent sentence for more context in their order of appearance in the movie were listed in a spreadsheet. The annotation material also included a list of functions and their definitions. A pilot annotation of a small proportion of the detected swear words was performed in order to test and finally refine the typology of functions. Two domain experts with a high command of English on level C2 of the Common European Framework of Reference for Languages (CEFR) and a background in linguistics separately annotated each swear word with a function.

4.5. Annotation validation

Agreement of annotators was calculated utilising Cohen’s kappa value for Inter-Rater Reliability. An initial calculation of the kappa value led to a moderate

agreement of 0.429, which upon refining the definitions of the functions could be pushed to 0.634. Remaining cases of disagreement were jointly discussed by the annotators to create a gold standard dataset for swear word functions. For some of these cases, it was essential to consult the movie sequence to determine the intended use of the swear word and its function.

5. Results

In 9,080 English words in the corpus, 111 swear words could be detected, some of which occur in close proximity or direct sequence of one another, e.g. *Oh, shit. Double shit. Bollocks! Oh, God*. (swear words underlined). The overall count and list of swear words per alphabetically ordered function is depicted in Table 1, where one and the same swear word might fulfil several functions in context. There are 113 function counts for 111 swear words since two occurrences fulfilled two independent functions, that is, *Careful, you ham-fisted cunt!* was found to express *criticism* and *anger* simultaneously and *Oh, bloody hell* expresses *anger* and *surprise* at the same time. Only in these two instances, two functions had to be assigned to the swear words, which only became clear to the annotators when consulting the intonation and situation in the movie. Furthermore, five occurrences of *God* and *breasts* were excluded due to their non-offensive use, e.g. *thank God!*, and two detected swear words were erroneously included, that is, *tit* is a typing error in the subtitles in *a tit like rabbits* which is *at it like rabbits* in the movie and *fools* is part of song lyrics playing in the background. For the remainder of swear words one function could be assigned.

Table 1: Total count of functions

Functions	Count	Swear words
advice	5	<i>fuck, fucking asses, sod 'em all, fool, sod</i>
anger	27	<i>ass, bastards, bloody hell, bollocks, bollocks, bollocks, breasts, bugger off, damn, double shit, fuck, fuck 'em, fucking, God, ham-fisted cunt, idiot, piss, shit, sod, stick</i>
banter	7	<i>dirty bitch, full sex, god damn, sexy, stupid ass, tits, sex goddess</i>
bonding	8	<i>bollocks, daft cow, fuck, fucking, silly</i>

criticism	24	<i>big knobhead, dull bastard, fannies, fat-ass, fuckwit boyfriend, fucked, fucking, fucking idiot, ham-fisted cunt, idiot, knob, nasty bastard, pissed, prematurely middle-aged prick, pretty nasty beast, shit, stupid, screwed up, wanker</i>
evaluation	12	<i>ass, cracked, crap, fuck, fuck me, God, hell, idiot, sex</i>
fear	2	<i>shit-faced, God</i>
jocularity	5	<i>bugger, crack, pissed, shit</i>
pain	2	<i>fuck</i>
request	0	-
ridicule	5	<i>fanny, sexy, tits pervert</i>
surprise	9	<i>bloody hell, bollocks, fuck, fuck me, God, shit</i>
not offensive	5	<i>God, breasts</i>
errors in subtitles	2	<i>tit (at it), fools (song lyrics)</i>
Total	113	

As can be seen from Table 1, *anger* with 27 and criticism with 24 counts were by far the most frequent functions, followed by *evaluation* and *surprise*. The most frequently occurring swear word across the corpus and functions is *fuck* with overall 29 occurrences, 10 of which express *anger*, e.g. *Fuck 'em*, and 5 criticism, e.g. *fuckwit boyfriend*. *Fuck me* is utilised to express pain, e.g. *Fuck me, that hurt!*, evaluation, e.g. *Oh, fuck me, I love Keats.*, and surprise, e.g. *And these are fuck me, absolutely enormous panties*. *Fucking* is generally utilised as intensifier, e.g. *fucking tuna*, expressing anger. Regarding the overall nature of swear words, we observed excretory, e.g. *piss*, and scatological examples, e.g. *shit*, body parts, e.g. *bollocks*, sexual practices, e.g. *wanker*, religious references, e.g. *hell*, animals, e.g. *cow*, mental illnesses or anti-social behaviour, e.g. *pervert*, and explicit insults, e.g. *fool*. The body with its functions and parts seems to be an excellent canvas on which to paint criticising and angry valuations as well as to strengthen social relations with *jocularity* and *bonding*. It is important to note that directionality of swearing is not represented in Table 1, that is, the offence might be self-directed, e.g. *I was so stupid*, or directed at others, e.g. *You stupid ass*. Furthermore, some swear words are strongly culture-dependent, e.g. the criticising *knobhead* as in *He's just a big knobhead with no knob*. inspired part of the title of this article and represents British slang. Religious references, e.g.

God, hell, goddess, are mainly used to express surprise, fear, evaluation and anger, where the *bantering sex goddess* lends itself to part of this article’s title. The only animal occurs in *bonding* and mental illnesses occur in *bantering, criticism, and ridicule*. Explicit insults of a person are utilised in almost all functions apart from surprise, pain, jocularity, and fear. Thus, what might appear as an insult in a literal reading becomes a vehicle for bonding, bantering or reinforcing advice given to an intimate friend when considering functions.

6. LLOD representation

To associate swear words and their functions, two separate Linguistic Linked Open Data (LLOD) resources are created: a concept set describing and specifying the swear word functions and a lexicon providing all detected swear words, their contexts of usage, and the identified function. To this end, we rely on the OntoLex-Lemon model (Cimiano et al. 2016), the current standard for representing linguistic data as Resource Description Framework (RDF), to represent all functions as `ontolex:ConceptSet` (see Fig. 2).

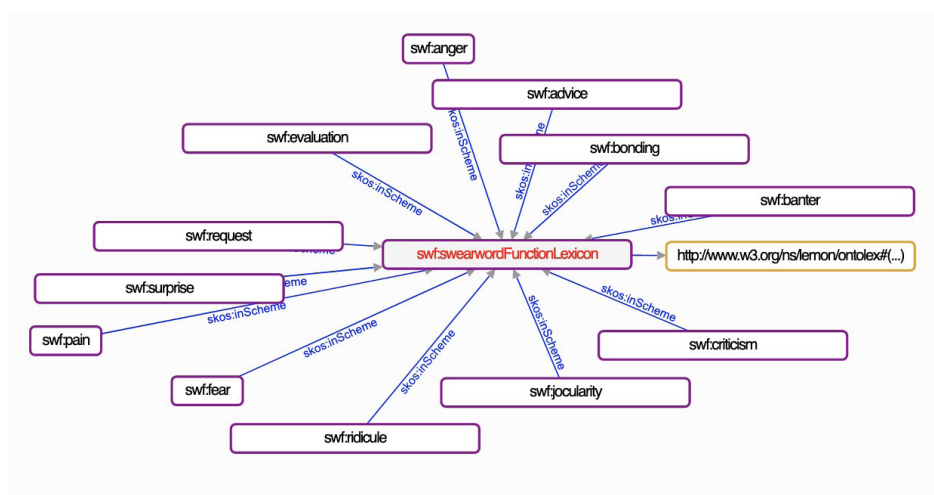


Figure 2: Visualisation of `ontolex:ConceptSet` of swear word functions⁷

⁷ This visualisation has been created with isSemantic.net: <https://issemantic.net/>.

Each function in the Swear Word Function concept set with the prefix `swf` is represented as a lexical concept in `ontolex` and associated with either a DBpedia or DBnary concept, e.g. for banter there is no adequate DBpedia but a DBnary concept as depicted in Fig. 3. Each function is associated with a natural language definition proposed in this article and modelled as `skos:Definition` as well as a lexical entry for its written representation. Advantages of this RDF representation, among others, are easy visualisation as shown in Fig. 2 and reusability to analyse swear words in other corpora or text types.

```
swf:anger a ontolex:LexicalConcept ;
  ontolex:isConceptOf <https://dbpedia.org/page/Anger> ;
  ontolex:isEvokedBy swf:anger_lex ;
  skos:definition "Demonstrating strong emotions through annoyance or fury expressed forcefully and/or in a loud voice; it may be targeted at an interactant but it may also be self-directed, typically used to release the speaker's emotions."@en ;
  skos:inScheme swf:swearwordFunctionLexicon .

swf:anger_lex a ontolex:LexicalEntry ;
  ontolex:canonicalForm swf:anger_lemma .

swf:anger_lemma ontolex:writtenRep "anger"@en .

swf:banter a ontolex:LexicalConcept ;
  ontolex:isConceptOf dbnary-eng:__cf_banter_Noun__1 ;
  ontolex:isEvokedBy swf:banter_lex ;
  skos:definition "Humorous pretending insulting between two or more interlocutors with the aim of supporting social bonding, thus creating an aura of intimacy; also called humorous insults."@en ;
  skos:inScheme swf:swearwordFunctionLexicon .

swf:banter_lex a ontolex:LexicalEntry ;
  ontolex:canonicalForm swf:banter_lemma .

swf:banter_lemma ontolex:writtenRep "banter"@en .
```

Figure 3: RDF representation of swear word functions⁸

In the Bridget Jones Swear Word Dictionary (BJSD) we represent swear words, their function and, where available, their definitions utilising `OntoLex` and its extension `OntoLex-FrAC`. This extension module allows us to assign a context example from the corpus to each lexical entry representing a specific swear word by means of the object property `frac:attestation` as depicted in Fig. 4. Swear words with the same meaning and function in the corpus are assigned a corpus frequency count with `frac:frequency`. To associate each swear word

⁸ The highlighting of the RDF representation in this article was generated with the Semantic Web – Turtle Editor <https://ci.mines-stetienne.fr/teaching/semweb/turtle.html>.

entry with its function, we introduce `swf:hasSwearwordFunction` that links lexical entries with lexical concepts. For each separate meaning of a swear word, a separate entry is created equipped with a definition taken from Wiktionary and DBNary (S  rasset 2015). The advantage of the latter is that each sense is represented as a separate entity, whereby we can link to a very specific sense of the word, e.g. *ass* may refer to Buttocks (Sense 1) or is used to express something unpleasant (Sense 4) of Etymology 2. This link to DBNary also assigns a word class to each entry.

```

bjsd:knobhead_en a ontolex:LexicalEntry, ontolex:Word ;
  dct:language <http://id.loc.gov/vocabulary/iso639-2/en>, <http://lexvo.org/id/iso639-1/en> ;
  ontolex:denotes dbnary-eng:knobhead__Noun__1, <https://en.wiktionary.org/wiki/knobhead#English>;
  ontolex:canonicalForm bjsd:knobhead_en_form ;
  swf:hasSwearwordFunction swf:criticism ;
  frac:attestation [
    a frac:Attestation ;
    frac:quotation "He's just a big knobhead with no knob."@en ;
    dct:source <https://yifysubtitles.ws/movie/bridget-jones's-diary/tt0243155> ].

bjsd:knobhead_en_form a ontolex:Form;
  ontolex:writtenRep "knobhead"@en .

bjsd:ass_1_n a ontolex:LexicalEntry, ontolex:Word ;
  dct:language <http://id.loc.gov/vocabulary/iso639-2/en>, <http://lexvo.org/id/iso639-1/en> ;
  ontolex:denotes dbnary-eng:_ws_1_ass__Noun__2, <https://en.wiktionary.org/wiki/ass#English>;
  ontolex:canonicalForm bjsd:ass_en_form ;
  swf:hasSwearwordFunction swf:anger ;
  frac:attestation [
    a frac:Attestation ;
    frac:quotation "Someone who insists I call him uncle while he gropes my ass and asks me the question dreaded by all Singletons."@en ;
    dct:source <https://yifysubtitles.ws/movie/bridget-jones's-diary/tt0243155> ].

```

Figure 4: OntoLex-FrAC example of swear word representation

Representing swear words and their functions as LLOD resources enables highly detailed analyses. For instance, all functions and different definitions of *fuck* can be investigated with a simple query or a visual representation of the data. All swear words pertaining to a specific function can be investigated easily. This representation also allows analysing the different types a single word might be assigned, e.g. *bollocks* refers to nonsense and testicles, and their respective functions, i.e., anger, bonding, and surprise.

6. Discussion and Conclusion

With this classification of swear words by their nature and function, we contribute to a systematic analysis method of swear words in context. While specific functions, such as criticism, aim at offending the interlocutor, others serve the more subtle aim of establishing or maintaining (bonding), reinforcing (advice), or potentially attracting (banter) an intimate social relationship. Thus, offensive language might not necessarily offend. Furthermore, swear words tend to occur in situations involving primordial emotions, such as pain, fear, anger or surprise, which, interestingly, collocate with religious expressions, especially *God*. The proposed typology and method of classifying swear words, to the best of our knowledge, is the first systematic approach in literature. Nevertheless, a clear limitation of this case study is its restriction to a single movie, language and culture. It would be highly interesting to investigate its generalizability across text type, languages and cultures.

To create the proposed typology, definitions of functions from literature represented a starting point followed by a first pilot annotation of real-world dialogues from a movie to investigate the typology's completeness and distinctiveness, leading to several changes. For instance, we initially included teasing as a separate category, but finally omitted it (see Section 4.3. for a detailed explanation). In distinguishing criticism and evaluation, the main question was whether criticism also applies if the person being criticised is not present. After careful consideration, we opted for classifying all negative evaluations as criticism, irrespective of the presences of the referent, and all positive or neutral cases as evaluation. For request, not a single case could be detected, however, this function might still be useful in other corpora.

Further confusion arose with the directionality of the (mock) insult. For instance, if *You stupid ass.* is self-directed, it more likely expresses *anger*, whereas directed at another person, it might be an instance of banter. Banter was generally found to require the presence of two interlocutors. Thus, it became evident that the knowledge of the pragmatic context of an utterance is crucial for determining its function. In this regard, the corpus selection turned out to be ideal since it provided the option to consult the movie scenes during annotation.

In terms of LLOD representation, the recently proposed OntoLex-FrAC module turned out to be vital for representing our resulting datasets. One remaining challenge was modelling the relationship between functions and swear words in LLOD, for which we introduced a new object property

To conclude, the pragmatic context is as important for annotating swear words as culture-specific awareness. A correct interpretation of functions requires to know interlocutors, their intonation, the situation within the plot, among many other pragmatic considerations. In terms of cultural awareness, both annotators were not originally from the UK and culture-specific nuances might have been missed. To mitigate this aspect, both annotators were advised to consult online definitions of swear words as well as to revisit the movie for further context. Nevertheless, a comparative study with annotations by first language speakers of UK English with the proposed annotations might provide interesting further insights.

Acknowledgement

This article is based upon work from the COST Action NexusLinguarum – “European network for Web-centered linguistic data science” (CA18209), supported by COST (European Cooperation in Science and Technology) www.cost.eu.

References

- ANDERSSON, LARS, G.; TRUDGILL, PETER. 1990. *Bad Language*. Penguin Books Ltd. London.
- BEDNAREK, MONIKA. 2006. *Evaluation in Media Discourse*. Bloomsbury, London.
- BEERS FÄGERSTEN, KRISTY ANINA. 2001. A descriptive analysis of the social functions of swearing in American English. *Dissertation Abstracts International, A: The Humanities and Social Sciences* vol. 61. 3142–A.
- BEERS FÄGERSTEN, KRISTY, ANINA. 2007. A sociolinguistic analysis of swearword offensiveness. *Saarland Working Papers in Linguistics* 1. 14–37. doi.10.22028/D291-23494.
- BOXER, DIANA; CORTÉS-CONDE, FLORENCIA. 1997. From bonding to biting: Conversational joking and identity display. *Journal of Pragmatics* 27. 275–294. doi.org/10.1016/S0378-2166(96)00031-8.

- BURGERS, CHRISTIAN; VAN MULKEN, MARGOT; SCHELLENS, PETER J. 2011. Finding irony: An introduction of verbal irony procedure. *Metaphor and Symbol* 26/1. 186–205. doi.org/10.1080/10926488.2011.583194.
- CIMIANO, PHILIPP; MCCRAE, JOHN P.; BUITELAAR, PAUL. 2016. Lexicon model for ontologies: community report, 10 May 2016. *Ontology-Lexicon Community Group under the W3C Community Final Specification Agreement (FSA)*.
- CONRAD, SUSAN; BIBER, DOUGLAS. 2000. Adverbial Marking of Stance in Speech and Writing. *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Eds. Hunston, Susan; Thompson, Geoffrey. Oxford University Press. Oxford. 56–73.
- DAWAELE, JEAN-MARC. 2016. Thirty shades of offensiveness: L1 and LX English users' understanding, perception and self-reported use of negative emotion-laden words. *Journal of Pragmatics* 94. 112–127. doi.org/10.1016/j.pragma.2016.01.009.
- DALY, NICOLA; HOLMES, JANET; NEWTON, JONATHAN; STUBBE, MARIA. 2004. Expletives as solidarity signals in ftas on the factory floor. *Journal of Pragmatics* 36. 945–964. https://doi.org/10.1016/j.pragma.2003.12.004.
- HALLIDAY, M.A.K. 1994. *An introduction to functional grammar*. Arnold. London.
- HAUGH, MICHAEL. 2010. Jocular mockery, (dis)affiliation and face. *Journal of Pragmatics* 42. 2106–2119.
- HAUGH, MICHAEL. 2017. Teasing. *The Routledge Handbook of Language and Humor*. Ed. Attardo, Salvatore. Routledge. New York. 204–218.
- HAUGH, MICHAEL; BOUSFIELD, DEREK. 2012. Mock impoliteness, jocular mockery and jocular abuse in Australian and British English. *Journal of Pragmatics* 44/9. 1099–1114. doi.org/10.1016/j.pragma.2012.02.003.
- HARTMAN, ROBERT. 1967. *The Structure of Value: Foundations of Scientific Axiology*. Southern Illinois University Press. Carbondale.
- HUNSTON, SUSAN. 2004. Counting the uncountable: Problems of identifying evaluation in a text and in a corpus. *Corpora and Discourse* 9. 157–188.
- JAY, TIMOTHY. 1999. *Why We Curse: A Neuro-Psycho-Social Theory of Speech*. John Benjamins. Amsterdam. doi.org/10.1075/z.91.
- KOTTHOFF, HELGA. 2007. Oral genres of humor: On the dialectic of genre knowledge and creative authoring. *Pragmatics* 17/2. 263–296. doi.org/10.1075/prag.17.2.04kot.
- MCENERY, TONY. 2006. *Swearing in English. Bad Language, Purity and Power from 1586 to the Present*. Routledge. Abingdon.
- LJUNG, MAGNUS. 2011. *Swearing: A Cross-Cultural Linguistic Study*. Palgrave Macmillan. London.
- LYONS, JOHN. 1997. *Semantics*. Cambridge University Press. Cambridge.

- LOVE, ROBBIE. 2021. Swearing in informal spoken English: 1990s–2010s. *Text & Talk* 41(5-6). 739–762. doi.org/10.1515/text-2020-0051.
- MARTIN, JAMES. 2000. Beyond exchange: appraisal systems in English. *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Eds. Hunston, Susan; Thompson, Geoff. Oxford University Press. Oxford. 142–175.
- OCHS, ELINOR. 1989. The Pragmatic of Affect. *Text*. 9/1.
- PARTINGTON, ALAN. 2007. Irony and reversal of evaluation. *Journal of Pragmatics* 39. 1547–1569. doi.org/10.1016/j.pragma.2007.04.009.
- PAVLENKO, ANETA. 2008. Emotion and emotion-laden words in the bilingual lexicon. *Bilingualism* 11/2. doi.org/10.1017/S1366728908003283.
- SÉRASSET, GILLES. 2015. DBnary: Wiktionary as a Lemon-based multilingual lexical resource in RDF. *Semantic Web*. 6/4. 355–361.
- SINGLETON, DAVID. 2009. Unspeakable words: the taboo fringe of the lexicon. *Advances in discourse approaches*. Ed. Dynel, Marta. Cambridge Scholars Publishing. Newcastle. 130–146.
- STAPLETON, KARYN. 2010. Swearing. *Interpersonal Pragmatics*. Ed. Locher, Miriam A.; Sage L. Graham. De Gruyter Mouton. Berlin – New York. 289–306. doi.org/10.1515/9783110214338.2.289.
- ULLMANN, STEPHEN. 1962. *Semantics: An Introduction to the Science of Meaning*. Blackwell. London.
- WHITE, PETER. 2015. Appraisal Theory. *The International Encyclopedia of Language and Social Interaction*. Ed. Tracy, Karen; Ilie, Cornelia; Sandel, Todd. Wiley-Blackwell. Boston. 1–8.

Appendix A. Extension of swear word dictionary

God, god, knobhead, fucking hell, prick, arsehole, wanker, Wanker, loser, Loser, stupid, moron, thick, bollocks, lunatic, bugger, jerk, idiot, wacko, fruitcake, bonkers, nutter, weird, damn, lost mind, maniac, fool, silly, daft, comedian, sod, crack

Od šupljoglavca do božice seksa: psovke u engleskim titlovima, njihove funkcije i prikaz modelom jezičnih povezanih podataka

Sažetak

Psovke su važno društveno sredstvo za ljudsku komunikaciju koje se, osim za puke uvrede, konvencionalno upotrebljavaju i za društveno povezivanje i zafrkanciju. Međutim, koliko nam je poznato, dosad nije predložena nijedna sustavna tipologija funkcija psovki. U ovom se radu takva tipologija predlaže metodom odozgo prema dolje oslanjajući se na književnost, kao i metodom odozdo prema gore analizom konkretnoga korpusa provokativnih dijaloga iz stvarnoga svijeta. Predlaže se dvanaest funkcija psovki, a to su: šaljivo zbližavanje, zafrkancija i šala, uvredljiva kritika, ljutnja i ismijavanje, neutralno procjenjivanje, zatim savjet, zahtjev, iznenađenje, bol i strah. Za ovo prvo istraživanje odabrali smo titlove britanskoga filma *Dnevnik Bridget Jones*, u kojem smo automatski detektirali 111 psovki primjenjujući filter vulgarnosti. Priroda otkrivenih psovki varirala je od navođenja tjelesnih dijelova i njihovih funkcija do upućivanja na vjeru ili mentalne bolesti. Dijelovi tijela, na primjer, posebno su korišteni za izražavanje šaljivosti, kritičnosti i ljutnje u spomenutom kontekstu, npr. engl. *bollocks*. Kako bismo prikazali rezultate, stvorili smo dva LLOD izvora: konceptualni skup funkcija i leksikon psovki. Svakom značenju psovke dodijeljen je zaseban leksički unos u OntoLexu i povezan s definicijom i njezinom funkcijom. Kako bi se svaki unos opremio primjerom konteksta i učestalošću pojavljivanja, korišten je modul OntoLex-FrAC. Kako bismo povezali funkcije iz skupa pojmova s leksičkim unosom psovke, uvodimo novo svojstvo objekta. U budućnosti se planira ponovna primjena tipologije psovki u analizi prijevoda titlova te u istraživanju uočavanja promjene u funkciji ili prirodi psovki.

Ključne riječi: psovke, funkcije, LLOD, titlovi

Keywords: swear words, functions, LLOD, subtitles

