



# A lexical database for public textual cyberbullying detection

**Aurelia Power**<sup>1</sup>

Institute of Technology Blanchardstown, Dublin

**Anthony Keane**

Institute of Technology Blanchardstown, Dublin

**Brian Nolan**

Institute of Technology Blanchardstown, Dublin

**Brian O'Neill**

Dublin Institute of Technology

## ABSTRACT

Public textual cyberbullying has become one of the most prevalent issues associated with online safety of young people, particularly on social networks. To address this issue, we argue that the boundaries of what constitutes public textual cyberbullying needs to be first identified and a corresponding linguistically motivated definition needs to be advanced. Thus, we propose a definition of public textual cyberbullying that contains three necessary and sufficient elements: the personal marker, the dysphemistic element and the cyberbullying link between the previous two elements. Subsequently, we argue that one of the cornerstones in the overall process of mitigating the effects of cyberbullying is the design of a cyberbullying lexical database that specifies what linguistic and cyberbullying specific information is relevant to the detection process. In this vein, we propose a novel cyberbullying lexical database based on the definition of public textual cyberbullying. The overall architecture of our cyberbullying lexical database is determined semantically, and, in order to facilitate cyberbullying detection, the lexical entry encapsulates two new semantic dimensions that are derived from our definition: cyberbullying function and cyberbullying referential domain. In addition, the lexical entry encapsulates other semantic and syntactic

---

<sup>1</sup> **Corresponding author** – Department of Informatics and Creative Digital Media, Institute of Technology Blanchardstown, Blanchardstown Road North, Dublin 15, D15 YV78 – Dublin, Ireland.

Email: [aurelia.power@itb.ie](mailto:aurelia.power@itb.ie)



information, such as sense and syntactic category, information that, not only aids the process of detection, but also allows us to expand the cyberbullying database using WordNet (Miller, 1993).

*Keywords: cyberbullying, lexical database, linguistic analysis, natural language processing*

## **1. Introduction**

Cyberbullying is one of the most prevalent risks encountered by young people on the Internet (Livingstone et al., 2011). Similar to traditional forms of bullying, cyberbullying can be understood in terms of the three fundamental criteria: intention of harm, repetition, and power imbalance between the victim and the bully (Hinduja & Patchin, 2009). However, traditional forms of bullying are restricted by temporal and geographical constraints, whereas, the specific environment in which cyberbullying occurs – the cyberspace – is characterised by four key aspects that allow the act of cyberbullying to transcend such constraints: (1) the ability to persist over time, (2) the ability to be searched for repeatedly, (3) the ability to be replicated numerous times, and (4) the ability to multicast to potentially large invisible audiences (Boyd, 2007). As a result, the cyberspace allows for the elements of repetition and power imbalance to be achieved in a very unique way. For instance, a hurtful message/post on social media can remain in cyberspace for an indefinite period of time, and, as such, it can be viewed over and over again not only by the targeted victim, but also by a potentially large number of bystanders; this, in turn, can lead to an intensified sense of powerlessness, given the larger and more persistent audience (Lagos, 2012).

Previous research in the field of cyberbullying detection has been driven by two paradigms which focus on two different types of analysis: network analysis and content analysis, although many studies combine both types of analyses. The social network analysis paradigm focuses on modelling interactions among users of a social network (Li & Tagami, 2014; Honjo et al., 2011; Hosseinmardi et al., 2014), based in particular on the usage statistics of digital media, such as sender and receiver IDs, account IDs, users' names, content, time and location of communication. Content analysis focuses on detecting cyberbullying instances based on text analytics and the use of Natural Language Processing (NLP) techniques, and, from this perspective, the task of cyberbullying detection was previously approached as a classification task (Yin et al., 2009) that involves data acquisition and pre-processing, feature extraction and classification. NLP techniques were used mostly in targeting explicit textual cyberbullying behaviours (Chen et al., 2012; Dadvar et al., 2012; Dinakar et al., 2011; Dinakar et al., 2012; Kansara & Shekokar, 2015; Kontostathis et al., 2013; Munezero et al., 2014; Nahar et al., 2013;

Nitta et al., 2013; Ptaszynski et al., 2010; Reynolds et al., 2011; Xu et al., 2012; Yin et al., 2009), but they were also used to identify subtle forms in several studies (Chen et al., 2012; Dinakar et al., 2012; Nitta et al., 2013; Ptaszynski et al., 2010).

Despite its merits, previous research has several limitations. First, from a lexical perspective, the content analysis paradigm typically used a limited set of linguistic features, such as predefined lists of profanities (Chen et al., 2012; Dadvar et al., 2012; Dinakar et al., 2011; Dinakar et al., 2012; Yin et al., 2009), bad words (Reynolds et al., 2011), foul terms (Nahar et al., 2013), bullying terms (Kontostathis et al., 2013), pejoratives and obscenities (Chen et al., 2012), or emotemes and vulgarities (Ptaszynski et al., 2010), as well as personal pronouns (Dadvar et al., 2012; Nahar et al., 2013; Yin et al., 2009). Another major limitation of previous approaches is that they do not attempt to set clear boundaries for what constitutes cyberbullying, and they target in fact the detection of offensive, hurtful, profane and violent language rather than cyberbullying instances.

To address such limitations, in this paper we focus on the relevant lexical, syntactic, and semantic aspects of the language used in public textual cyberbullying and how they can be integrated into a formal lexical database to effectively support the cyberbullying detection process. First, we propose a definition of public textual cyberbullying that identifies three necessary and sufficient elements. We then describe how the database organises the lexical entries using the cyberbullying function and the sense. A formal definition of the lexical database is also provided and each syntactic category is considered in terms of the grammatical properties that contribute to cyberbullying detection. Finally, we describe two novel semantic features specifically related to cyberbullying detection: the cyberbullying function and the cyberbullying referential domain.

The paper is organised as follows. In section 2, we put forward a linguistically motivated definition of public textual cyberbullying. Section 3 describes the overall morphological, syntactic and semantic aspects of the lexical entries and how the entries are organised in the database. In section 4, we formally define the cyberbullying lexicon and specify the attribute matrix. Section 5 is concerned with the syntactic categories and their respective grammatical properties relevant to cyberbullying detection. In section 6, we describe the contribution of sense, while in section 7 we identify the cyberbullying information included in the lexical database. Section 8 concludes the paper, summarising the main contributions of the present research.

## **2. Defining public textual cyberbullying**

The view that we take here is that the presence of explicit terms does not suffice for a message or post to be classified as public textual cyberbullying; it must be linked to or it must target a particular person, or group of people. In addition, explicit terms are not the only means by which textual cyberbullying can be realised. For instance, the messages/posts *You don't deserve a mum* or *You are not pretty* or *You look like a donkey* do not contain any explicit profane or offensive, or violent terms. They are instances of textual cyberbullying that use negation and simile techniques to hurt or offend the victim(s). Thus, we argue that a clear operational definition of what constitutes public textual cyberbullying is a necessary step in the processes of detection, and we propose the following definition of textual cyberbullying:

- (1) **Definition:** A given public textual instance (whether expressed as a message, a post or a sentence) can be classified as cyberbullying if it contains all of the following three elements: (1) the personal marker/pointer, (2) the explicit or implicit dysphemistic element, and (3) the link between the personal marker/pointer and the dysphemistic element.

The personal marker/pointer refers to that element that is used to identify or point to the victim(s), while the dysphemistic element refers to what has been defined by Allan & Burrige (2006) as the “word or phrase with connotations that are offensive either about the denotatum<sup>2</sup> and/or to people addressed or overhearing the utterance” (p.31); the link between the previous two elements refers to that element by means of which the dysphemistic element targets the victim(s) identified or pointed to by the personal marker/pointer. The characteristics of explicitness and implicitness are inherent characteristics that apply to the dysphemistic element only, and refer to whether cyberbullying instances contain explicit (profane, offensive, or violent) terms or not. However, each of the three elements need not be explicitly present in a given instance, as long as it can be inferred from other contextual elements, such as the sentential structure or the previous instances. For example, in the sentence *You are a cunt*, the dysphemistic element is both explicitly present and realised by means of explicit profane language, but, in the sentence *You are not smart*, although the dysphemistic element is explicitly present, it is not realised by means of explicit language, but by means of negation. On the other hand, the instance *You clearly are*, although it contains no dysphemistic language, was labelled as public textual cyberbullying in our dataset<sup>3</sup>, and it was only when we considered

<sup>2</sup> Here, by denotatum is meant what the speaker/sender refers to, while by connotations is meant the set of semantic meanings related to the denotatum and the contexts associated with the denotatum.

<sup>3</sup> Our dataset originates from Ask.fm which contains conversations extracted from several users.

the previous post/message uttered by a different user - *I am not pathetic* - that we were able to identify the dysphemistic element in the form of the offensive adjective *pathetic*; thus, we were able to infer the missing cyberbullying element from the previous instance and resolve the sentence to the following full form: *You clearly are pathetic*.

Based on our dataset, the types of dysphemistic<sup>4</sup> terms that we observed to be associated with public textual cyberbullying can be divided into (1) explicit, such as profane/obscene, violent, and offensive/insulting, and (2) implicit, which tend to be associated with techniques such as negation, animal metaphors and similes. The latter terms can be divided into positive and neutral, based on the fact that when used as part of certain expression they become hurtful and offensive. For instance, the term *nice*, which semantically is associated with positive connotations, can be negated in order to hurt one's feelings - *You are not a nice person*, while the term *spell*, which is semantically associated with neutral connotations, can be used to target one's mental ability: *You can't even spell!*

Personal markers/pointers are terms that refer or point to a victim or victim's characteristics, family, possessions, or ideas, either on their own, or as part of a phrase. When explicitly present, they are typically expressed by means of personal pronouns (for example, *you, she, they, me, your*), proper names (for example, *Helen, Michael*) and screen proper names (*Flower\_Power16*). However, at other times, instances do not contain explicit pronouns or names, but only dysphemistic terms or expressions, yet they can still be qualified as cyberbullying, because they can only refer to people. For example, the instance *Retard!* can be labelled as cyberbullying, since the term *retard* can only refer to a person. Thus, certain common nouns that can only refer to people, such as *child, woman, idiot, retard*, can also act as personal markers/pointers. Yet, at other times, the explicit presence of personal marker is not required, as long as it can be inferred from other elements, such as the structure of the sentence. This is particularly the case of imperative sentences where the subject can be dropped. For example, in the instance *Rot in hell!*, the inferred subject is a person, since only people can communicate online.

Finally, the linguistic element that acts as the link between the personal pointers and dysphemistic element is the conjugated verb, regardless of whether its presence is explicit (for instance, *She is stupid*) or it can be inferred from the underlying copular

---

<sup>4</sup> We also include euphemisms (e.g. *willy*) and orthophemisms (e.g. *penis*) in addition to direct dysphemism (e.g. *dick*), since they can also be used dysphemistically. As Allan & Buridge (2006) suggest, it depends entirely on the context whether terms such as *penis* or *vagina* are used as orthophemisms or as dysphemisms; likewise, it depends on the context whether terms such as *poo* or *willy* are used dysphemistically or euphemistically.

structure (for example, *You an idiot* can be rephrased using as *You are an idiot*). But the verb can also represent the dysphemistic element itself, thus holding both functions simultaneously. In this case, when the verb serves as the link and the dysphemistic element at the same time, the link is reflexive since it links the personal marker back to itself. For example, the sentence *I hope your mom dies* contains a reflexive link in the form of the violent dysphemistic verb *die* which links the personal marker *your mom* back to itself.

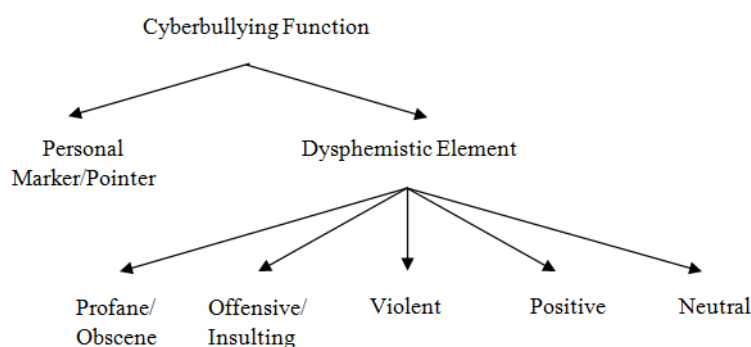
### 3. Architecture of the lexical database

Based on our definition of public textual cyberbullying, we propose a novel lexical cyberbullying database that is enriched with grammatical and semantic information, mainly, cyberbullying specific information. Morphologically, the cyberbullying lexical database organisation follows the lemma-based model and any inflectional cases such as plural, past tense, third person singular, and present and past participle are handled by a lemmatisation process that reduces inflected forms to their respective lemmas. For instance, *cunts* is reduced to *cunt*, *died* is reduced to *die*, *rides* is reduced to *ride*, *shagging* is reduced to *shag*, and *fucked* is reduced to *fuck*, respectively. However, from a derivational perspective, the cyberbullying lexical database recognises all derived forms as lemmas in their own right. For example, the term *fucker* has its own entry in the cyberbullying lexical database, despite the fact that it is a noun derived from the verb *fuck*. Graphical variations and abbreviations are stored separately in a dictionary that maps such informal variations to their respective typical form, in order to represent the lexical entry in a more economical manner; for instance, the informal pronoun *meself* is mapped to its formal counterpart *myself*, while the variation *betch* is mapped to the lemma *bitch*. To account for phrasal verbs, we specify their constituents: the main constituent – represented by the headword or lemma, and the particle constituent – represented by two options: *none* or the concrete particle, such as *off* or *up*.

Syntactically, the cyberbullying database is organised around syntactic categories. Each lexical entry encapsulates the syntactic category and the corresponding grammatical and semantic information that is relevant to cyberbullying detection. For instance, important features specific to nouns and pronouns are type and gender, while transitivity, valency and argument types are features of the verb that we consider relevant to cyberbullying detection. For adjectives and adverbs, an important feature is their type, that is, whether they are descriptive or relational. One of the intuitions behind such organisation is that it allows us to expand the cyberbullying database using the WordNet lexical database, since WordNet is also organised around syntactic categories (Miller, 1993). However, unlike WordNet, our cyberbullying database includes both open categories, such as nouns, verbs,

adjectives, adverbs, and closed categories, such as pronouns.

Overall, the architecture of the cyberbullying lexical database is determined semantically, and, in order to facilitate cyberbullying detection, the lexical entry encapsulates cyberbullying specific information which is functionally derived from the elements that our definition of textual cyberbullying identifies. Thus, we organise the lexical entries based on the cyberbullying function which represents the first two dimensions of the cyberbullying definition: (1) the personal marker/pointer and (2) the dysphemistic element, which can be further subdivided into obscene/profane<sup>5</sup>, insulting/offensive<sup>6</sup>, violent, on the one hand, and positive and neutral, on the other hand. The third element, the link between the personal marker and the dysphemistic language, is not included as a function in the cyberbullying lexical database, since it does not characterise a lexical entry, but rather the relation among sentential components. As such, the third element is detected by other means, such as the grammatical dependency relations. The hierarchical structure of the cyberbullying function is depicted in Figure 1.



**Figure 1.** The cyberbullying function hierarchy.

The lexical entries are further grouped together in sets of synonyms<sup>7</sup>, each group of synonyms representing a unique sense. Such organisation also draws from that of

<sup>5</sup> We group together profane and obscene terms based on the WordNet definitions of *profanity* (“vulgar or irreverent speech or action”) and *obscenity* (“an indecent word or phrase”), since both are defined in terms of vulgar and indecent qualities (the term vulgar is defined as “conspicuously and tastelessly indecent”).

<sup>6</sup> We also group together insulting and offensive terms, since *insult* is defined in WordNet as a rude expression intended to offend or hurt.

<sup>7</sup> The term *synset* was first introduced by Miller (1995) and refers to a given set of synonyms, or a group of entries that hold synonymous senses.

WordNet lexical database (Miller, 1995) and it was designed in this manner so that we could further expand the cyberbullying database using the WordNet. Thus, in addition to the syntactic category, we also use the semantic sense and the related concept of synonymy to expand the cyberbullying lexical database with terms found in WordNet.

The cyberbullying lexical database initially consisted of relevant terms that were extracted from the dataset used for analysis (excluding those terms that are typically considered stop words). However, these terms are specific to our dataset and account only for a small portion of the lexical richness found in textual cyberbullying. As such, we use the WordNet lexical database (Miller, 1995) to extract more cyberbullying relevant terms, based on the type of relations defined in WordNet and syntactic categories. For instance, we were able to retrieve from the synset of the verb *want* synonymous verbs, such as *desire*, *wish*, *require*, *need*, while for the noun *cunt* we were able to retrieve synonyms such as *bitch*, *puss*, *pussy*, *slit*, *snatch*, *twat*. We also used the WordNet hyponymy/hypernymy relations to extract more nouns that inherit same properties; for instance, we used the *animal* entity to extract more instances of animals, such as *bull*, *cat*, etc. Other ways of enriching the lexicon included a list of proper names (Babble, 2016) which contains girls, boys and unisex names, a list of ethnic slurs (Wikipedia, 2016), a list of religious slurs (Wikipedia, 2016), a list of violent words (Vocabulary University, 2016), and a list of profanities and obscenities (Free Web Headers, 2016; Luis von Ahn's Research Group, 2016; NoSwearing, 2016; Urban Dictionary, 2016). For each of these ways of enriching the lexicon, corresponding functionalities were created.

At a higher level, the cyberbullying lexical database is a functional network of concepts where a given cyberbullying functional domain subsumes senses that are related to that particular function. In turn, each sense subsumes the entries whose senses are the same. For instance, the *cock*<sub>2</sub> entry is related to *prick*<sub>1</sub> entry by means of their synonymous sense, which defines them as *obscene terms for penis* and places both entries under the same function: obscene/profane dysphemistic element. Figure 2 shows an example of how the senses are semantically related by means of the cyberbullying function and how the entries are lexically related by means of synonymy or membership to the same sense domain or synset.



Profane/Obscene Cyberbullying Function		
Obscene terms for penis	Obscene terms for female genitals	Sense_n
<i>cock</i> <sub>2</sub>	<i>cunt</i> <sub>1</sub>	<i>entry</i> <sub>x</sub>
<i>dick</i> <sub>1</sub>	<i>puss</i> <sub>2</sub>	<i>entry</i> <sub>y</sub>
<i>pecker</i> <sub>1</sub>	<i>pussy</i> <sub>2</sub>	.
<i>prick</i> <sub>1</sub>	<i>slit</i> <sub>2</sub>	.
<i>pecker</i> <sub>2</sub>	<i>snatch</i> <sub>1</sub>	.
<i>putz</i> <sub>1</sub>	<i>twat</i> <sub>2</sub>	<i>entry</i> <sub>z</sub>
<i>shaft</i> <sub>1</sub>		
<i>tool</i> <sub>1</sub>		

**Figure 2.** An example of the semantic and lexical structure of the cyberbullying database.

Although we have presented the cyberbullying lexical database in a modular fashion, the morphological, syntactic and semantic levels actively interact with one another. For instance, the morphological properties, such as the particle<sup>8</sup> of phrasal verbs, can determine the sense of an entry, as exemplified by the lemma *fuck*, typically associated with sexual intercourse, but when the particle *off* is added it has the sense of “leave immediately”. The syntactic properties of entries can also determine the semantic properties in several ways. First, the syntactic category affects the sense of the lexical entry. For instance, the term *fuck* has several entries in the cyberbullying lexical database, and they have different senses associated with each syntactic category: *fuck*<sub>1</sub> as a noun corresponds to the sense of “slang for sexual intercourse”, while *fuck*<sub>2</sub> as an intransitive verb corresponds to “have sexual intercourse”. Furthermore, the sense is determined by certain properties associated with each syntactic category. For instance, the verb *fuck* has different senses that are determined by transitivity type, as well as by its particle: *fuck*<sub>2</sub> as an intransitive verb has the sense of “have sexual intercourse”, while *fuck*<sub>3</sub> as a transitive verb has the sense of “have sexual intercourse with”. But the syntactic category also imposes restrictions on other semantic properties of the lexical entry, such as the cyberbullying function and the cyberbullying referential domain. For instance, personal pronouns can only function as personal makers/pointers and their referential domain is restricted to person only. Proper names also have their referential domain restricted to person only. On the other hand, verbs, adjectives and adverbs can only be grouped under the dysphemistic function, whether explicit (profane, offensive, or violent) or not (positive or neutral).

<sup>8</sup> Typically, prepositions function as particles of the phrasal verbs.

#### 4. Formal definition of the cyberbullying lexical database

Formally, the cyberbullying lexical database  $L$  is a monolingual lexicon for English and we define it as the set of indexed pairs consisting of a lemma or headword, and a list of its associated attributes  $(l_i, A_i)$  as follows:

- (2)  $L = \langle (l_1, A_1), (l_2, A_2), \dots, (l_n, A_n) \rangle$ ,  
 where  $n$  denotes the number of entries in the lexicon, or its size.

Each lexical entry or item is represented using the attribute-value matrix. The subscript index  $i$ , which is associated with each lemma  $l$ , serves two purposes. First, the subscript index has the role of creating unique entries represented computationally as strings, which can facilitate many computational operations, for example, searching. Secondly, it allows lexical entries to be organised in terms of their corresponding senses, since we associate each entry with a single sense. On the other hand, the list of attributes,  $A$ , encodes syntactic information, such as syntactic category, type and gender, as well as semantic information and cyberbullying specific information that corresponds to sense, cyberbullying function, and cyberbullying referential domain, respectively. The general structure of the lexical entry is shown in (3):

- (3) **LEMMA<sub>index</sub>**: LEXICAL ENTRY  
 HEADWORD: applicable to all entries  
 SYNTACTIC CATEGORY: applicable to all entries  
 TYPE: applicable to all entries  
 GENDER: applicable to nouns and pronouns  
 PERSON: applicable to personal pronouns only  
 PARTICLE: applicable to verbs only  
 ARGUMENTS: applicable to verbs only  
 SENSE: applicable to all entries, except those acting solely as personal markers  
 CYBERBULLYING FUNCTION: applicable to all entries  
 CYBERBULLYING REFERENTIAL DOMAIN: applicable to all entries

The syntactic attributes of entries correspond to properties associated with each syntactic category, and some apply to all categories, while others only to specific ones. For instance, the TYPE attribute applies to all entries, but the PARTICLE attribute applies only to verbs. Sense applies to all open categories, with the exception of proper nouns, which do not typically encapsulate senses<sup>9</sup>. Personal

<sup>9</sup> Although proper nouns are said to have particular meanings associated with their roots, we are not interested in such meanings, since proper names are given outside the context of cyberbullying, at birth, for instance. However, some historical names can be used dysphemistically

pronouns do not encapsulate senses either, since they only act as personal markers/pointers. Each of these attributes is further discussed in the next sections.

## 5. Syntactic categories

From a cyberbullying perspective, the syntactic category<sup>10</sup> of a given entry (sometimes referred to as *lexical category*<sup>11, 12</sup> or *part of speech*) plays a major role in the detection process. There are several syntactic categories that we included in the cyberbullying lexical database, and they belong mainly to the open class categories, such as nouns, verbs, adjectives and adverbs. But we also include some closed class categories, such as personal pronouns, to account for portion of the personal markers/pointers diversity.

### 5.1. Nouns

The noun category is very common in cyberbullying and is used to target victim(s) by making direct reference to people or their family, behaviours, body parts, possessions, ideas, sexuality, religion, or indirect reference through comparisons to animals or other objects. It usually has several grammatical functions: the subject (*Tom is gay*) or the subject complement (*She is a bitch*), or the object of a verb, direct (*Fuck Tom*) or indirect (*I'll give Dana such a box in the face*), but it can also be the object of a preposition (*I'll box the head of that bitch!*) or be an object complement (*I consider you an idiot*). Additionally, nouns can modify other nouns, indicating possessive modifier functions (*Tom's dick*) or attributive functions (*monkey face*).

There are several features of the noun that are directly relevant to cyberbullying

---

and, in this case, they have a corresponding sense attached, and they also specify the dysphemistic functional domain to which they belong.

<sup>10</sup> Because syntactic categories allow us to expand the database using WordNet (Miller, 1995), and to maintain consistency with the terminology used in designing the WordNet, we use throughout the present paper the term *syntactic category*.

<sup>11</sup> According to Van Valin (2001), lexical categories can also be used to refer to noun, verb, adjective, adverb or adposition (preposition and postposition) categories.

<sup>12</sup> Many grammars, such as Transformational Grammar, Government and Binding Theory, Minimalist Program, etc., also distinguish between lexical categories – viewed as categories that belong to open classes and which have full semantic content – and functional categories – viewed as categories that belong to closed classes and which have no particular semantic content on their own; from this perspective, lexical categories include categories such as nouns, verbs, adjectives, and adverbs, while functional categories include conjunctions, determiners, auxiliaries and prepositions.

detection. First, type – common and proper<sup>13</sup> – plays an important role in cyberbullying detection. Common nouns are those that describe classes of entities (*face, man, dog*), while proper nouns are those that refer to particular/given entities (*Lena, Brandon, China*). From a cyberbullying perspective, although common nouns can function as personal markers (*woman, child, dude*), they are most likely used as dysphemistic terms (*bitch, idiot, cut*) or part of dysphemistic expressions (*stupid woman*). On the other hand, proper names are used as personal markers/pointers to refer to people. For instance, the proper name *Maria* clearly refers to a person in online interaction, unless specified otherwise. However, some historical proper names are used as dysphemisms and, thus, they have senses that are relevant to cyberbullying. For instance, names such as *Einstein* can be used dysphemistically in sentences such as *You are no Einstein*, where *Einstein* can be understood as a person that is very intelligent or smart.

Secondly, gender is an important feature that can support the process of cyberbullying detection, particularly of those forms based on gender discrimination and sexuality. For instance, the message/post *Shauna is a man!* is intended to imply that Shauna has masculine characteristics or even to ridicule Shauna’s sexuality, by contrasting her name which is a girl’s name with the noun *man* which is associated the male gender. Gender, in the cyberbullying lexical database, is represented as masculine (for example, *Brendan* or *faggot*), feminine (for instance, *Maria* or *lesbian*), dual (for example, *Alex* which is a unisex name or *bastard*) and neutral (for instance, *dog*). Here, we do not use gender in a grammatical sense, since in Modern English nouns do not have a corresponding gender, but we use gender to describe whether the referent that a noun denotes describes a female (feminine), or a male (masculine), or both (dual), or an asexual entity (since innate entities do not typically have a sex, and, as such, they are labelled as neutral). Two examples of noun entries are shown in (4) and (5), respectively.

- (4) **NIGGA**<sub>1</sub>: LEXICAL ENTRY  
 HEADWORD: nigga  
 SYNTACTIC CATEGORY: noun  
 TYPE: common  
 GENDER: dual  
 SENSE: extremely offensive name for a Black person  
 CYBERBULLYING FUNCTION: offensive  
 CYBERBULLYING REFERENTIAL DOMAIN: person

<sup>13</sup> Common nouns can be further divided into concrete and abstract, and the concrete nouns can be further subdivided into count, mass and collective nouns; however, we are interested here only in the top-level dichotomy, since it has a direct bearing on cyberbullying detection.

- (5) **SNAKE**<sub>1</sub>: LEXICAL ENTRY  
 HEADWORD: snake  
 SYNTACTIC CATEGORY: noun  
 TYPE: common  
 GENDER: dual  
 SENSE: limbless scaly elongate reptile  
 CYBERBULLYING FUNCTION: neutral  
 CYBERBULLYING REFERENTIAL DOMAIN: animal

These noun entries can occur in sentences such as *Nigga!* and *I don't entertain snakes like you*, respectively.

## 5.2. Verbs

The verb is an important category in cyberbullying detection from several perspectives. As the main predicator, it constitutes the direct link between the personal marker/pointer and the dysphemistic element (*She looks awful*), as well as being the dysphemistic element itself (*I hate you*). The infinitive forms can serve as verb complements (*You deserve to rot in hell*), as well as adjective complements (*You should consider yourself lucky to suck my dick*). Although verbs in the form of participials play an important role in cyberbullying, they are treated as adjectives and/or adverbs, since they typically act as modifiers.

An important feature of the verb that is paramount to the linkage system between the personal marker/pointer element and the dysphemistic element is transitivity type which indicates whether the verb can take objects or not. Particularly, for cyberbullying detection, an intransitive verb indicates that the personal marker/pointer is the subject or modifies the subject. For example, in the sentence *Do you masturbate?*, the subject of the verb *masturbate* is the pronoun *you* which represents the personal marker, while *masturbate* is an intransitive verb that at the same time constitutes the dysphemistic element and the link between the personal marker and the dysphemistic element (reflexive link). Similarly, in the sentence *Your sister stinks*, the subject of the intransitive verb *stink* is *sister* which is modified by the possessive personal marker *your*. On the other hand, a transitive verb indicates that the personal marker/pointer can be either a subject or a direct object. For instance, in the sentence *I hate you*, the verb *hate* is a transitive verb that takes as direct object the personal marker indicated by the pronoun *you*, while in the sentence *You deserve death* the subject is the one representing the personal marker and the transitive verb *deserve* takes as object a violent dysphemistic element. The personal marker can also be the indirect object of a ditransitive verb, such as *give* (*I'll give you such a box in the face*).

Verb valency and the semantic type of its arguments are two related concepts that are also relevant to cyberbullying detection. Valency provides the number of arguments a verb takes (subject and objects) and is particularly important in understanding the conceptual content of a clause, since a clause is complete only when the valency requirements of the main verb of the clause are met (Huddleston & Pullum, 2005). For instance, the sentence *She has* is grammatically incomplete, since it only specifies one of the two arguments of the verb *have*<sup>14</sup>. We are mostly interested in three types of valency<sup>15</sup>: monovalency – describing all intransitive verbs (*die, burst*), divalency – describing transitive verbs that typically take a subject and a direct object (*hate, deserve*), and trivalency – describing verbs that typically take a subject, a direct object and an indirect object (*give, put*). The semantic type of the verb arguments is important in identifying the verb’s selectional preferences and is particularly important for detecting metaphoric forms of cyberbullying, since metaphors seem to violate the verb’s semantic selectional preferences described as the type of concepts that verb arguments typically represent (Wilks, 1978; Wilks, 2007). For instance, the verb *die* is a monovalent verb that normally takes as syntactic subject a noun or noun phrase denoting an animate entity (*I hope your mom dies*), but when it takes an inanimate entity as subject, the verb *die* is used metaphorically (*The light was dying when we arrived*).

To identify the semantic type of verb arguments we use the 25 nominal concepts described by Miller (1993) which broadly cover distinct lexical and conceptual domains found in WordNet, and which also serve as the semantic primes (unique beginners) of the hierarchies found in WordNet where the features of the unique beginner/conceptual domain are shared by all its hyponyms/subordinates. These concepts are listed in Figure 3.

---

<sup>14</sup> Within discourse (such as, conversational context), *She has* may be perceived as complete, for example, as an answer to the question *Who has finished eating?*

<sup>15</sup> Typically, verbs are divided into four classes according to their valency type; however, the avalent class is made up mostly of meteorological verbs that are not directly relevant to cyberbullying detection.

{act, action, activity}	{quantity, amount}
{natural object}	{event, happening}
{animal, fauna}	{relation}
{natural phenomenon}	{feeling, emotion}
{artefact}	{shape}
{person, human being}	{food}
{attribute, property}	{state, condition}
{plant, flora}	{group, collection}
{body, corpus}	{substance}
{possession}	{location, place}
{cognition, knowledge}	{time}
{process}	{motive}
{communication}	

**Figure 3.** The 25 nominal concepts found in WordNet (Miller, 1993).

We have also used as argument types the 15 verbal concepts found in WordNet (Fellbaum, 1993). These verbal concepts are divided into those that express actions and events, and those that express states. Actions and events are further subdivided into 14 semantic domains, while stative verbs are a standalone class, expressing states of being, having, as well as spatial relations. The verbal concepts are listed in Figure 4.

{verbs of bodily functions and care}	{motion verbs}
{verbs of change}	{emotion verbs}
{verbs of communication}	{perception verbs}
{competition verbs}	{verbs of possession}
{consumption verbs}	{verbs of social interaction}
{contact verbs}	{weather verbs}
{cognition verbs}	{stative verbs}
{creation verbs}	

**Figure 4.** The 25 nominal concepts found in WordNet (Miller, 1993).

In addition to the nominal and verbal concepts found in WordNet, we added the *underspecified* type of argument to represent arguments that can belong to any conceptual domain; for example, the bivalent verb *hate* typically takes as subject an animate entity (person or animal), however, its direct object could be anything, since one can hate all sorts of things, and, thus, its second argument is semantically underspecified.

Formally, we define valency and verb arguments type together as the ARGUMENTS feature, describing the verb argument structure, such as the number and the semantic type of the verb's syntactic arguments: subject, direct object and indirect object, respectively. Note that we only specify the semantic type of the required arguments: for monovalent verbs, we specify the type of the subject, for bivalent

verbs we specify both the subject type and the direct object type, and for trivalent verbs we specify the semantic types of all three arguments. Examples of what verb entries can encapsulate are shown in (6) and (7), respectively.

- (6)           **FUCK**<sub>5</sub>: LEXICAL ENTRY  
 HEADWORD: fuck  
 SYNTACTIC CATEGORY: verb  
 TYPE: intransitive  
 PARTICLE: off  
 ARGUMENTS: [person]  
 SENSE: leave or go away from a place immediately  
 CYBERBULLYING FUNCTION: profane  
 CYBERBULLYING REFERENTIAL DOMAIN: person
- (7)           **HATE**<sub>2</sub>: LEXICAL ENTRY  
 HEADWORD: hate  
 SYNTACTIC CATEGORY: verb  
 TYPE: transitive  
 PARTICLE: n/a  
 ARGUMENTS: [person/animal, underspecified]  
 SENSE: dislike intensely; feel antipathy or aversion towards  
 CYBERBULLYING FUNCTION: violent  
 CYBERBULLYING REFERENTIAL DOMAIN: person

The first verb entry can occur in sentences such as *Fuck off!*, while the second can appear in sentences such as *I hate you!*

### 5.3. Adjectives

The adjective category is also very common in cyberbullying and they are used to specify hurtful, insulting or offensive information (true or untrue) about the victim(s). We also include here participial modifiers, both present (for example, *shagging woman*, where *shagging* is the present participle form of the verb *shag*) and past (for instance, *infested mouth*, where *infested* is the past participle form of the verb *infest*), since they function mainly as adjectives. In cyberbullying interactions, adjectives typically serve several syntactic functions: attributive (*stupid cunt*), predicative (*Sinead is ugly*) and as object complement (*He considers you stupid*).

Like in WordNet, adjectives are divided into descriptive and relational (Fellbaum et al., 1993). Descriptive adjectives have the ability to appear in both prenominal and postcopula/predicative positions. For instance, the adjective *ugly* can appear in sentences such as *Ugly bitch!*, but also in sentences like *That bitch is ugly*. Relational or domain adjectives are usually derived from nouns (for example, *intellectual* is



derived from *intellect*) and they are mostly restricted to prenominal positions (although few may appear in the predicative position). In addition, relational adjectives cannot typically be graduated (statements such as *the extremely facial surgery* are not acceptable). Relational adjectives do not refer to specific qualities of the noun they modify, but to the conceptual domain they belong to. For instance, the adjective *facial*<sup>16</sup> can appear successfully in sentences such as *You need a facial surgery*, but the sentence *The surgery is facial* sounds contrived. The distinction between descriptive and relational/domain adjectives is particularly important for metaphorical language (Sullivan, 2007), since most of the domain adjectives can also be used descriptively, but in a metaphorical sense, rather than literal. Take for instance the adjective *criminal*<sup>17</sup> derived from the noun *crime*; as a domain adjective, it can be used literally to modify the nouns describing general domains or concepts (*criminal law*), however, as a descriptive adjective, it can metaphorically describe specific noun referents, such as *criminal chat* (to refer to a brilliant conversation). Two examples of adjective entries are shown in (8) and (9), respectively.

- (8)           **UGLY**<sub>r</sub>: LEXICAL ENTRY  
                   HEADWORD: ugly  
                   SYNTACTIC CATEGORY: adjective  
                   TYPE: descriptive  
                   SENSE: displeasing to the senses  
                   CYBERBULLYING FUNCTION: offensive  
                   CYBERBULLYING REFERENTIAL DOMAIN: generalised
- (9)           **FACIAL**<sub>r</sub>: LEXICAL ENTRY  
                   HEADWORD: facial  
                   SYNTACTIC CATEGORY: adjective  
                   TYPE: relational  
                   SENSE: of or concerning the face  
                   CYBERBULLYING FUNCTION: neutral  
                   CYBERBULLYING REFERENTIAL DOMAIN: generalised

The adjective entry shown in (8) can occur in sentences such as *Your face is ugly*, while the adjective entry in (9) can be used in sentences such as *You need a facial surgery*.

#### 5.4. Adverbs

<sup>16</sup> *Facial* is also a noun that denotes “care for the face that usually involves cleansing and massage and the application of cosmetic creams” (WordNet 2.1).

<sup>17</sup> The term *criminal* can also be a noun denoting “a person that acts against the law” (WordNet 2.1).

A small portion of the cyberbullying lexical database is dedicated to adverbs, which can serve several modifying functions in order to target the victim(s) in a more efficient and intense manner: to modify a verb (*Have you actually looked in the mirror?*), to modify an adjective (*She is extremely ugly*), to modify another adverb (*I hope you die very soon*), and to modify a participial modifier that functions as adjective (*You are so fucked*).

Like adjectives, adverbs can be divided into descriptive (*extremely, beautifully, horribly*) and relational (*spiritually, criminally, verbally*), a distinction that can help identify metaphorical usages (Sullivan, 2007). Some adverbs, such as *fucking*, have the same form as their adjectival counterparts. For instance, *fucking* can act as an adjective (*fucking bitch*) or as an adverb (*fucking ugly*). Similarly, the adjective *long* can modify a noun (*long nose*), while the adverb *long* can modify a verb (*how long can you fuck?*). In such cases, the lexical database accommodates entries for both syntactic categories. Two examples of how adverb entries are encoded in the cyberbullying lexical database are shown in (10) and (11), respectively.

(10)       **HARDLY**<sub>1</sub>: LEXICAL ENTRY  
               HEADWORD: hardly  
               SYNTACTIC CATEGORY: adverb  
               TYPE: descriptive  
               SENSE: almost not  
               CYBERBULLYING FUNCTION: neutral  
               CYBERBULLYING REFERENTIAL DOMAIN: generalised

(11)       **BADLY**<sub>1</sub>: LEXICAL ENTRY  
               HEADWORD: badly  
               SYNTACTIC CATEGORY: adverb  
               TYPE: descriptive  
               SENSE: without skill or in a displeasing manner  
               CYBERBULLYING FUNCTION: offensive  
               CYBERBULLYING REFERENTIAL DOMAIN: generalised

The first adverb entry can occur in sentences such as *You are hardly a beauty* while the second can be used in sentences such as *You spell quite badly*.

## 5.5. Pronouns

Traditionally, lexical databases do not include closed syntactic categories, such as pronouns. However, cyberbullying instances (messages/posts/sentences) are often explicitly directed at people and, usually, this is achieved by using certain nouns, such as names or person referring terms, but also by using personal pronouns that function as either subjects (*he, you*) or objects (*him*), or found in possessive

constructions (*your mum* or *her face*). For this reason, we also include in the database all personal pronouns, including reflexive forms, such as *himself*. Additionally, we include relative pronouns that introduce subordinate clauses, since they can be used to relate a clause to terms that denote people (*Jake who is a stupid dick* or *Jake whose sister is a whore*) and demonstrative pronouns, since they can be used anaphorically to refer to people (*Jack stood me up. That's what I call a real bastard*). Other pronouns that the cyberbullying database includes are indefinite pronouns, since they are likely to appear in negated forms of textual cyberbullying (*You deserve nothing* or *Nobody likes you*).

Apart from the type of the pronoun (personal, personal reflexive, relative, demonstrative and indefinite), another feature that is important in cyberbullying detection is gender. Like in the case of nouns, gender can play an important role in gender and sexuality discriminative forms of cyberbullying. For example, the instance *He wears his skirt pink* implies that the male person indicated by the personal pronoun *he* is gay or has effeminate characteristics, since skirt is defined as "a garment hanging from the waist; worn mainly by girls and women" (WordNet 2.1). Thus, we add gender information to pronoun entries as feminine (such as *she, her, hers, or herself*), masculine (for instance, *he, him, his, himself*), dual (for example, *you*) or neutral (for instance, *anything*).

Finally, another feature of the pronoun that we have found relevant to cyberbullying detection is person which, in fact, applies only to personal pronouns. Thus, it is important, from a detection perspective, to identify whether pronouns functioning as subjects are first person pronoun or not, in order to discriminate instances such as *I looked awful last night at the races* from instances such as *You/She looked awful last night at the races*; while the former is not a cyberbullying instance, the latter is. Additionally, it is important to identify whether pronouns functioning as objects are first person pronoun or not, so to discriminate instances such as *I really hate myself* from instances such as *I really hate you/her*. Two examples of pronoun entries are shown in (12) and (13), respectively.

(12)        **YOU<sub>1</sub>**: LEXICAL ENTRY  
               HEADWORD: you  
               SYNTACTIC CATEGORY: pronoun  
               TYPE: personal  
               GENDER: dual  
               PERSON: second  
               SENSE: n/a  
               CYBERBULLYING FUNCTION: personal marker  
               CYBERBULLYING REFERENTIAL DOMAIN: person

(13)        **ANYTHING<sub>1</sub>**: LEXICAL ENTRY

HEADWORD: anything  
 SYNTACTIC CATEGORY: pronoun  
 TYPE: indefinite  
 GENDER: neutral  
 SENSE: a thing of any kind  
 CYBERBULLYING FUNCTION: neutral  
 CYBERBULLYING REFERENTIAL DOMAIN: generalised

The first pronoun entry can be used in sentences such as *You are hardly a beauty*, while the second can be used in sentences such as *You don’t deserve anything*.

## 6. Sense

Most lexical entries in the cyberbullying lexical database encapsulate a sense typically extracted from WordNet 2.1. Sense represents some aspect of the meaning that relates the lexical entry to cyberbullying. For example, the term *cut* has several entries in the cyberbullying database, each being associated with a distinct sense. As a noun, *cut* is defined as “wound made by cutting” (*Add another cut to your arm*). As a verb, *cut* has two senses: one corresponding to the intransitive verb, being defined as “make an incision or separation” (*Do you cut?*), and one for its transitive counterpart, being defined “make an incision on something” (*I’ll fucking cut your throat!*).

As already noted previously and as shown in the examples above, syntactic category and some of their respective features, such as type, are important elements in determining the sense of an entry. Another feature that affects the entry’s sense in the case of phrasal verbs is the particle. For instance, the intransitive verb *fuck* can also be defined as “leave immediately” (WordNet 2.1) when it has attached the particle *off* (*Fuck off*).

Computationally, sense is responsible for organising the entries into conceptual networks similar to WordNet’s synsets (Miller, 1993; Miller, 1995). It allows the cyberbullying database to establish a certain level of compatibility with WordNet in order to increase the number of entries and to also obtain the senses of the initial entries. In fact, most of the entries have senses that are extracted from WordNet. The rest of the senses, most of which are associated with profanities and obscenities, are extracted from Urban Dictionary (2016), while terms pertaining to ethnic, racial and religious slurs have their senses from Wikipedia (2016).

## 7. Cyberbullying information

According to our definition, public textual cyberbullying is defined as a function of

three elements: the personal marker/pointer, the dysphemistic element and the link between them. At times these elements can be easily identified. For instance, the message/post/sentence *You are a cunt* can be classified as cyberbullying based on the fact that it explicitly contains all the elements described in our definition: the personal pronoun *you* functions as personal marker/pointer, the noun *cunt* functions as dysphemistic element, and *are* (the present tense inflected form of the verb *to be*) serves to explicitly link the two previous elements. However, other times, these elements are not explicitly present, and a single word can take on more than one function. For instance, the post *Retard!* does not contain any explicit personal markers or pointers, and consequently, it does not contain an explicit link between the personal marker and the dysphemistic element. But the term *retard*, while it explicitly constitutes the dysphemistic element, can only refer to a person and, thus, it implicitly functions as personal marker/pointer. From this perspective, we argue that the entries in the cyberbullying lexical database should specifically contain information about their primary function, as well as whether they can refer, be directed at or explicitly target people only when used on their own.

### 7.1. Cyberbullying function

The cyberbullying function determines the functional domains of the lexical database, ensuring that sense networks are further clustered around more general cyberbullying concepts for easier and more economical access. It specifies the primary role lexical entries play according to our definition, that is, whether lexical entries constitute the dysphemistic element, or whether they are personal markers/pointers. Specifically, the lexical entries are tagged using one of the following cyberbullying functions:

- (14)
- a. Personal markers/pointers such as *you, he, or Helen*.
  - b. Profane/obscene terms such as *fuck, bitch, or dickhead*.
  - c. Offensive/insulting terms such as *nigga, ugly, or horrible*.
  - d. Violent terms such as *die, cut, or hate*.
  - e. Positive terms such as *deserve, love, or beautiful*.
  - f. Neutral terms such as *look, need, or pig*.

The first function describes the personal marker/pointer element, while the next five functions describe the dysphemistic element. However, in certain instances, the dysphemistic function can overlap with that of personal pointer. For example, the cyberbullying function of the term *nigga* is explicitly offensive, but, at the same time, it can act as personal pointer when used on its own, since it can only refer to a person in an online interaction; in such cases, the dysphemistic function takes precedence over the personal marker function, and to retain the information related

to the personal marker, we introduce the cyberbullying referential domain which can identify whether or not the scope of an entry goes beyond that of people, and their character, attributes, ideas, behaviours, family, thus resolving the issue of function overlap.

## 7.2. Cyberbullying referential domain

Although reference is determined pragmatically based on the contextual information provided by the discourse (Birner, 2013), typical referents, denotata and designata can be determined semantically (Lipka, 1992), and, thus, be encoded into the list of attributes of the lexical entry. To distinguish from the conceptual domain that is defined in WordNet (Miller, 1993), we use the term cyberbullying referential domain to refer to the set of typical referents that an entry with a given sense has as a noun or it represents as a verb, or to which it can be applied as modifier (adjective, adverb, or participial – present or past). Therefore, we exclude any figurative referents, which will make it easier to identify figurative uses of certain terms in cyberbullying. Because cyberbullying is a behaviour that targets people (and their character, attributes, ideas, behaviours, family) explicitly or implicitly, we are interested in those referents, denotata and designata of entries that have a bearing on cyberbullying detection. Specifically, the role of the cyberbullying referential domain is to identify whether or not entries can be typically used to refer to people, or be applied to people and their possessions, qualities, ideas, behaviours, family. For example, the referential domain of the noun *creep* is represented by *people*, while the referential domain of one of the entries for the verb *shag* is represented by *sexual behaviour*. The referential domain of the adjective *small* is represented by generalised referential domain, since it can be applied to any physical entity, not only to people or their physical attributes; but the referential domain of the adjective *idiotic* is more restricted, since it can be applied only to people and people’s mental abilities and ideas.

Nouns and verbs have more specific and better defined cyberbullying referential domains. For instance, the noun *idiot* can only refer to a person, and, thus, its cyberbullying referential domain is explicitly that of people. On the other hand, the corresponding adjective *idiotic*, although more restricted than other adjectives, can be applied directly to a person (*idiotic man*), but also to a person’s actions or ideas (*idiotic idea*), without being used figuratively or metaphorically; however, despite its wider application, it still applies to people, directly or indirectly, and, for this reason, its cyberbullying referential domain is also that of people. Usually, adjectives have a wider referential domain, particularly, adjectives used for evaluations (*good, bad*), for expressing activities (*fast, slow*) and for expressing potency (*strong, weak*) which can be applied to almost all nouns (Fellbaum et al., 1993), and, in such cases, their

referential domain is generalised.

Adverbs also have a wider applicability range and can be applied to a wide range of verbs, adjectives and adverbs. For example, the adverb *stupidly*, although restricted to the people referential domain, it can be applied to various human behaviours and actions, while the adverb *badly* can be applied to an even wider range of verbs, adjectives and adverbs, its referential domain being generalised.

The cyberbullying referential domain of an entry is also determined by its sense and syntactic category, and related grammatical properties. For example, the lemma *cut* discussed previously has several entries associated with distinct senses, which also determine the scope of their referential domain: as a noun, *cut* is associated with the cyberbullying referential domain of result of violence, while as a verb it is associated with the cyberbullying referential domain of violent behaviour. Another example of interaction between sense, syntactic category, type, and cyberbullying referential domain is the lemma *burn* which, as a noun, it has the sense of “damage inflicted by fire” (WordNet 2.1) and, as such, is associated with the cyberbullying referential domain of result of violence (*I saw the burns on your legs*), but as an intransitive verb it has the sense of “undergo combustion” (WordNet 2.1) and, thus, is associated with the cyberbullying referential domain of violent event (*I hope you burn in hell*); moreover, as a transitive verb, it has the sense of “destroy something by fire” (WordNet 2.1) and, in this case, is associated with the cyberbullying referential domain of violent behaviour (*I’ll burn your house*). We list the types of cyberbullying referential domains that we include in our database in (15).

- (15)
- a. People/person (and their character, mental ability, ideas, behaviours), such as *creep, woman, idiotic, smart, brave, stupidly, swear, curse*.
  - b. Sexual body part, such as, *penis, vagina, dick, twat*.
  - c. Other body part, such as *face, arm, brain*.
  - d. Bodily effluvia, such as *shit, piss*.
  - e. Animal, such as *ape, snake, chicken*.
  - f. Sexual behaviour, such as *shag, fuck*.
  - g. Violent behaviour, such as *kick, kill, destroy*.
  - h. Violent event, such as *burst, burn, die*.
  - i. Result of violence, such as *cut, scar*.
  - j. Generalised – when the referential domain represents two or more of the types above, such as *bad, good, small, big, ugly*.
  - k. Other – when the referential domain is not explicitly or directly relevant to cyberbullying, for instance, *have, possess, get*.

The cyberbullying referential domain is determined by a conceptual distinction that separates people (and their character, attributes, ideas, behaviours, family) from several other domains, and, from this perspective, it can be thought as identifying

whether entries, in addition to the dysphemistic function, may serve as personal markers. In other words, it helps resolve issues associated with function overlap, such as in the case of the lemma *nigga* which can be used to offend a person, but it can also serve as personal marker/pointer; in this case, we define its cyberbullying function as *offensive* and we, also, attach the cyberbullying referential domain of *people*.

In addition, the cyberbullying referential domain contributes to the detection of the most common forms of metonymies and metaphors<sup>18</sup> that are typically found in cyberbullying, particularly those that invoke the body parts, body effluvia and animal conceptual domains. For example, the noun *snake* represents an animal entity that must be interpreted metaphorically when used to refer to a person (*you are a snake*). Computationally, it has the advantage that it quickly labels those instances that do not require additional contextual information to be qualified as cyberbullying (see discussion above on the post *Retard!*), being able to identify those entries that are used in explicit forms of cyberbullying, based on two criteria: profane or offensive dysphemistic cyberbullying function, and cyberbullying referential domain of people.

## 8. Summary

In this paper, we have characterised a novel lexical resource that aids the process of cyberbullying detection. It follows from a linguistically motivated definition of textual cyberbullying that identifies its necessary and sufficient parameters. The database is a monolingual database for English that uses both syntactic and semantic information. Morphologically, the database follows a lemma-based approach, while, syntactically, it is organised using open syntactic categories (nouns, verbs, adjectives, adverbs), as well as closed syntactic categories (pronouns).

The overall architecture of the cyberbullying lexical database is semantic, that is, the entries are organised into networks of synonyms. The synonyms networks are further grouped into cyberbullying functional domains using two cyberbullying functions that we have identified in our definition of public textual cyberbullying: the personal marker/pointer function, on the one hand, and the dysphemistic function (profane/obscene, offensive/insulting, violent, neutral, positive), on the other hand. Additional semantic information is provided by the cyberbullying referential domain.

---

<sup>18</sup> Note that, from a computational perspective, we treat cyberbullying part-whole metonymies in a similar manner to metaphors, since, like metaphors, they also constitute mappings from the body part cyberbullying referential domain to people referential domain, a design choice that was also driven by WordNet semantic separation of body part from person, each constituting a conceptual domain in its own right, with distinct semantic structure.



The cyberbullying referential domain has the role of identifying the typical referential domain of the entries, providing quick and efficient access to entries that can constitute cyberbullying on their own, as well as playing a major role in detecting common metonymies and metaphors in cyberbullying.

## **About the authors**

**Aurelia Power** is an assistant lecturer in the Department of Informatics and Creative Digital Media at the Institute of Technology Blanchardstown, Dublin. She has recently completed her PhD in computational linguistics which is entitled "A Linguistic Approach to Detecting Public Textual Cyberbullying". Her research interests include computational linguistics, linguistic theory, language acquisition, philosophy of language, artificial intelligence, data mining, cyberbullying and behaviour analysis. She is also a Board Certified Behaviour Analyst and a certified Professional Java Programmer.

**Dr. Anthony Keane** is currently the Head of School of Informatics and Engineering at the Institute of Technology Blanchardstown, Dublin. He is also a principle investigator in the Security Research Lab, located in the Learning & Innovation Centre in ITB where he has several doctoral research students working with industrial partners both SME and International. Main research areas cover Network Resilience, Cyber Security, Digital & Cloud Forensics, Cyber Bullying and Security Intelligence. Dr. Keane has multiple conference and journal publications and is a frequent invited speaker at technology conferences and industrial seminars.

**Dr. Brian Nolan** is Head of Department of Informatics and Creative Digital Media at the Institute of Technology Blanchardstown Dublin, in Ireland. His research interests include linguistic theory at the morpho-syntactic semantic interface, argument structure and valence, constructions in grammar, event structure in language, the architecture of the lexicon and computational approaches to language processing, computational linguistics. His linguistic work has been in the functional linguistic model of Role and Reference Grammar and he has published extensively internationally. Computing/computational linguistic research has concentrated on: 1) The development of a framework and supporting application suite for mobile and distributed command and control of robotic devices using speech recognition as the core enabling technology; 2) The development of a rule-based Arabic to English machine translation engine that uses Role and Reference Grammar as the linguistic model supporting an interlingua bridge; and

3) the investigation of linguistic models to underpin conversational agents. In 2012 Dr. Nolan published his book with Equinox UK on the linguistic structure of Irish in a Role and Reference Grammar account entitled: ‘The structure of Modern Irish: A functional account’. In 2013, Benjamins published his co-edited volume: ‘Linking constructions into functional linguistics – The role of constructions in grammar’ in their Studies in Language Companion series. His co-edited Benjamin volume on computational linguistics and linguistic theory, ‘Language processing and grammars: The role of functionally oriented computational models’ was published in 2014, also in their Studies in Language Companion series. He also co-edited a Benjamins book on ‘Causation, transfer and permission’ in linguistic theory, which appeared in early 2015. In January 2017, Benjamins published his co-edited book on complex predication entitled ‘Argument realisation in complex predicates and complex events: Verb verb constructions at the syntax semantic interface’. Dr. Nolan has over 40 years’ experience nationally and internationally within the computer industry, with over 2 decades in academia, in a variety of senior roles, and is also a widely published professional linguist. Dr. Nolan is a Fellow of the Irish Computer Society.

**Professor Brian O’Neill** is Director of the Research, Enterprise and Innovation Services at Dublin Institute of Technology. His research is focused on media policy and digital technologies; media and information literacy, e-safety and information society policy for children and he has an international profile as a researcher on young people’s use of new media and the internet. He is a member of the Management Group for the EU Kids Online network responsible overseeing its policy work package. He sits on Ireland’s Internet Safety Advisory Committee and also chaired the Irish government’s task force on Internet Content Governance, reporting to the Minister for Communications, Energy and Natural Resources. He has undertaken research for the European Commission, UNICEF, the Broadcasting Authority of Ireland and the ICT Coalition. He is a member of the Council of Europe’s Expert Group on Digital Citizenship Education.

### **Article history**

*Paper received: 16 February 2017*

*Paper received in revised form and accepted for publication: 30 April 2017*

## References

- Allan, K. & Burrige, K. (2006). *Forbidden words: Taboo and censoring of language*. Cambridge: Cambridge University Press.
- Babble, D. (2016). Baby Names. <<http://www.babble.com/baby-names/>> [12/02/2017].
- Birner, B. (2013). *Introduction to pragmatics*. Oxford: Wiley-Blackwell Publishing.
- Boyd, D. (2007). Why youth (heart) social network sites: The role of networked publics in teenage social life. In D. Buckingham (Ed.), *MacArthur foundation series on digital learning, youth, identity, and digital media* (pp. 1 – 26). Cambridge, MA: MIT Press.
- Chen, Y., Zhou, Y., Zhu, S. & Xu, H. (2012). Detecting offensive language in social media to protect adolescent online safety. In *Proceedings of the 4<sup>th</sup> ASE/IEEE International Conference on Social Computing* (pp. 71–80). <<http://www.cse.psu.edu/~sxz16/papers/SocialCom2012.pdf>> [12/02/2017].
- Dadvar, M., de Jong, F., Ordelman, R. & Trieschnigg, D. (2012). Improved cyberbullying detection using gender information. In *Proceedings of the 21<sup>st</sup> International Conference Companion on World Wide Web, ACM* (pp. 121–126). <<http://wwwhome.ewi.utwente.nl/~dadvarm/Maral/uploads/Main/DIR2012.pdf>> [12/02/2017].
- Dinakar, K., Jones, B., Havasi, C., Lieberman, H. & Picard, R. (2012). Common sense reasoning for detection, prevention, and mitigation of cyberbullying. *ACM Transactions on Interactive Intelligent Systems*, 2(3), 18:1-18:30, DOI 10.1145/2362394.2362400.
- Dinakar, K., Reichart, R. & Lieberman, H. (2011). Modeling the detection of textual cyberbullying. In *Proceedings of the 5<sup>th</sup> AAAI International Conference on Weblog and Social Media* (pp.11–17). <<http://web.media.mit.edu/~kdinakar/3841-16937-1-PB.pdf>> [12/02/2017].
- Fellbaum, C. (1993). English verbs as a semantic net. <<http://wordnetcode.princeton.edu/5papers.pdf>> [12/02/2017].
- Fellbaum, C., Gross, D. & Miller, K. (1993). Adjectives in WordNet. <<http://wordnetcode.princeton.edu/5papers.pdf>> [12/02/2017].
- Free Web Headers (2016). Full list of bad words and top swear words banned by google. <<http://www.freewebheaders.com/full-list-of-bad-words-banned-by-google/>> [12/02/2017].
- Hinduja, S. & Patchin, J.W. (2009). *Bullying beyond the schoolyard: preventing and responding to cyber-bullying*. Thousand Oaks, CA: Corwin Press.
- Honjo, M., Hasegawa, T., Hasegawa, T., Mishima, K., Suda, T. & Yoshida, T. (2011). A framework to identify relationships among students in school bullying using digital communication media. In *Proceedings of the 3<sup>rd</sup> IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing* (pp.1474–1479). <<http://www.proceedings.com/13745.html>> [12/02/2017].
- Hosseinmardi, H., Han, R., Lv, Q., Mishra, S. & Ghasemianlangroodi, A. (2014). Towards

- understanding cyberbullying behavior in a semi-anonymous social network. In *Proceedings of 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 244–252), <<https://arxiv.org/pdf/1404.3839.pdf>> [12/02/2017].
- Huddleston, R. & Pullum, G.K. (2005): *A student's introduction to English grammar*. Cambridge: Cambridge University Press.
- Kansara, K.B. & Shekokar, N.M. (2015). A framework for cyberbullying detection in social network. *International Journal of Current Engineering and Technology*, 5(1), 494 – 498.
- Kontostathis, A., Reynolds, K., Garron, A. & Edwards, L. (2013). Detecting cyberbullying: Query terms and techniques. In *Proceedings of the WebSci'13 Conference*. <<https://pdfs.semanticscholar.org/fee1/4e570861573f920036f2abaae119fc46a659.pdf>> [12/02/2017].
- Lagos, C. (2012). Cyberbullying: The challenge to define. *Cyberpsychology, Behavior, and Social Networks*, 15(6), 285–289.
- Li, M. & Tagami, A. (2014). A Study of contact network generation for cyber-bullying detection. In *Proceedings of the 28<sup>th</sup> International Conference on Advanced Information Networking and Applications Workshops* (pp. 431 – 437). <<http://ieeexplore.ieee.org/document/6844675/>> [12/02/2017].
- Lipka, L. (1992). *An outline of English lexicology: Lexical structure, word semantics and word-formation*. 2<sup>nd</sup> ed. Tübingen: Niemeyer.
- Livingstone S., Haddon L., Görzig A. & Ólafsson K. (2011). Risks and safety on the Internet: the perspective of European children, full findings. EU Kids Online, London School of Economics and Political Science, <<http://eprints.lse.ac.uk/33731/>> [12/02/2017].
- Luis von Ahn's Research Group (2016). Useful resources: Offensive/Profane word list. <<http://www.cs.cmu.edu/~biglou/resources/bad-words.txt>> [12/02/2017].
- Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D. & Miller, K. (1993). Introduction to WordNet: An on-line lexical database. <<http://wordnetcode.princeton.edu/5papers.pdf>> [12/02/2017].
- Miller, G.A. (1993). Nouns in WordNet: A lexical inheritance system. Available at <<http://wordnetcode.princeton.edu/5papers.pdf>> [12/02/2017].
- Miller, G. (1995). WordNet: A lexical database for English. *Communications of the ACM*, 38(11), 39–41.
- Munezero, M., Mozgovoy, M., Kakkonen, T., Klyuev, V. & Sutinen, E. (2013). Antisocial behaviour corpus for harmful language detection. In *Proceedings of the 2013 Federated Conference on Computer Science and Information Systems* (pp. 261–265). <<https://fedcsis.org/proceedings/2013/pliks/fedcsis.pdf>> [12/02/2017].
- Nahar, V., Li, X. & Pang, C. (2013). An effective approach for cyberbullying detection. *Communications in Information Science and Management Engineering*, 3(5), 238–247.
- Nitta, T., Masui, F., Ptaszynski, M., Kimura, Y., Rzepka, R. & Araki, K. (2013). Detecting cyberbullying entries on informal school websites based on category relevance maximization. In *Proceedings of the 6<sup>th</sup> International Joint Conference on Natural Language*

- Processing* (pp. 579–586). <<http://www.aclweb.org/anthology/I13-1066>> [12/02/2017].
- Nocentini, A., Calmaestra, J., Schultze-Krumbholz, A., Scheithauer, H., Ortega, R. & Menesini, E. (2010). Cyberbullying: Labels, behaviours and definition in three European countries. *Australian Journal of Guidance and Counselling*, 20(2), 129–142.
- Norvig (2007). How to write a spelling corrector. <<http://norvig.com/spell-correct.html>> [12/02/2017].
- NoSwearing (2016). List of swear words and curse words. <<http://www.noswearing.com/dictionary>> [12/02/2017].
- Princeton University (2016). WordNet: A lexical database for English. <<https://wordnet.princeton.edu/>> [12/02/2017].
- Ptaszynski, M., Dybala, P., Matsuba, T., Rzepka, R. & Araki, K. (2010). Machine learning and affect analysis against cyber-bullying. In *Proceedings of the Linguistic and Cognitive Approaches to Dialog Agents Symposium, at the AISB 2010 Convention* (pp. 7–16). <[http://arakilab.media.eng.hokudai.ac.jp/~ptaszynski/data/AISB2010\\_Cyberbullying\\_paper.pdf](http://arakilab.media.eng.hokudai.ac.jp/~ptaszynski/data/AISB2010_Cyberbullying_paper.pdf)> [12/02/2017].
- Reynolds, K., Kontostathis, A. & Edwards, L. (2011). Using machine learning to detect cyberbullying. In *Proceedings of the 10th International Conference on Machine Learning and Applications Workshops (ICMLA 2011)* (pp. 241–244). <<http://webpages.ursinus.edu/akontostathis/ReynoldsKontostathisEdwardsFINAL.pdf>> [12/02/2017].
- Sebastiani, F. (2002). Machine learning in automated text categorisation. *ACM Computing Surveys*, 34(1), –47.
- Slonje, R., & Smith, P.K. (2008). Cyberbullying: Another main type of bullying? *Scandinavian Journal of Psychology*, 49(2), 147–154.
- Sullivan, K.S. (2007). Grammar in metaphor: A construction grammar account of metaphoric language. PhD. Thesis with University of California, Berkley, <[http://linguistics.berkeley.edu/dissertations/Sullivan\\_dissertation\\_2007.pdf](http://linguistics.berkeley.edu/dissertations/Sullivan_dissertation_2007.pdf)> [12/02/2017].
- Urban Dictionary (2016). Urban Dictionary. <<http://www.urbandictionary.com/>> [12/02/2017].
- Van Valin, R.D. Jr. (2001). *An introduction to syntax*. Cambridge: Cambridge University Press.
- Vocabulary University (2016). Violence vocabulary word list. <<https://myvocabulary.com/word-list/violence-vocabulary/>> [12/02/2017].
- Wikipedia the Free Encyclopaedia (2016). List of ethnic slurs by ethnicity. <[https://en.wikipedia.org/wiki/List\\_of\\_ethnic\\_slurs\\_by\\_ethnicity#Jews](https://en.wikipedia.org/wiki/List_of_ethnic_slurs_by_ethnicity#Jews)> [12/02/2017].
- Wikipedia the Free Encyclopaedia (2016). List of religious slurs. <[https://en.wikipedia.org/wiki/List\\_of\\_religious\\_slurs](https://en.wikipedia.org/wiki/List_of_religious_slurs)> [12/02/2017].
- Wilks, Y. (1978). Making preferences more active. *Artificial Intelligence*, 11(3), 197–223.
- Wilks, Y. (2007). Making preferences more active. In K. Ahmad, C. Brewster & M. Stevenson (Eds.), *Words and intelligence I* (pp. 141–166). New York: Springer.

- Xu, J., Jun, K., Zhu, X. & Bellmore, A. (2012). Learning from bullying traces in social media. In Proceedings of 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, (pp. 656–666). <<http://pages.cs.wisc.edu/~jerryzhu/pub/naaclhlt2012.pdf>> [12/02/2017].
- Yin, D., Xue, Z., Hong, L., Davison, B.D., Kontostathis, A. & Edwards, L. (2009). Detection of harassment on Web 2.0. In *Proceedings of the Content Analysis in the WEB 2.0 Workshop*, <<http://webpages.ursinus.edu/akontostathis/harassment.pdf>> [12/02/2017].