

Dictionaries of Mexican Sexual Slang for NLP

Roberto Villarejo-Martínez

Centro Nacional de Investigación y Desarrollo Tecnológico, Departamento de Ciencias
Computacionales,
Cuernavaca, México, 62490
rvillarejo@cenidet.edu.mx

Noé Alejandro Castro-Sánchez

Centro Nacional de Investigación y Desarrollo Tecnológico, Departamento de Ciencias
Computacionales,
Cuernavaca, México, 62490
ncastro@cenidet.edu.mx

and

Gerardo Eugenio Sierra-Martínez

Universidad Nacional Autónoma de México, Instituto de Ingeniería,
Ciudad de México, México, 04510
gsierram@iingen.unam.mx

Abstract

In this paper the creation of two important relevant resources for the double entendre and humour recognition problem in Mexican Spanish is described: a morphological dictionary and a semantic dictionary. These were created from two sources: a corpus of albur (drawn from “Antología del albur”) and a Mexican slang dictionary (“El chilangonario”). The morphological dictionary consists of 410 forms of words that corresponds to 350 lemmas. The semantic dictionary consists of 27 synsets that are associated to lemmas of morphological dictionary. Since both resources are based on Freeling library, they are easy to implement for tasks in Natural Language Processing. The motivation for this work comes from the need to address problems such as double entendre and computational humour. The usefulness of these disciplines has been discussed many times and it has been shown that they have a direct impact on user interfaces and mainly in human-computer interaction. This work aims to promote the scientific community to generate more resources about informal language in Spanish and other languages.

Keywords: Double entendre, Computational humour, Sexual Slang

1 Introduction

Slang (or argot) is a linguistic modality used in specific contexts. The slang is spoken by people with something in common, e.g. occupation, career, geographic region, social status, etc. There is a wide variety of slangs for each language around the world. Sometimes these slangs are used to produce comedy and sometimes are used to hide the real meaning of a word or sentence. In Mexico, one of the most popular is the sexual slang and it is so rich that even it can have some variations depending regions. This is mainly due to the local culture or mixture of conventional Spanish with local languages or both.

There are different types of humor, some of them are more complex than others but in most it is observed that wordplay and slang are used to produce the desired effect: hilarity. In recent years, the development of systems capable of handling natural language has been emphasized; either involving human-computer communication, understanding of written narratives, information on the Web or human conversation [1]. This trend has become so

important that there even exists an experimental paradigm in which it is claimed that the human-computer relationship is fundamentally social: CASA (Computers are Social Actors) [2]. In this paradigm, it is showed that “principles drawn extant literature in social psychology, communication and, sociology are relevant to the study of human-computer interaction and have clear implications for user interface design”.

In Mexico, there is a popular verbal game called “albur” in which sexual slang and double entendre is widely used. This game consists of sexually subjugation in a verbal way to the listener, being able to realize or not depending if this knows the slang. Many linguistic devices such as wordplay, metaphor, euphemisms, corruptions, metaplasms, and sexual slang are used in “albur”. So, slang is definitely a fundamental part of language and popular culture. It is therefore necessary to form dictionaries of this kind of words. This article is specifically focused on Mexican sexual slang. The purpose is to create computational resources that can be used in NLP (Natural Language Processing) problems and to promote more works on slang for other languages.

There already exists NLP dictionaries for Spanish language but no work has been done to provide specific dictionaries of Mexican sexual slang even when it is very popular among Mexican speakers. Slang is often used with euphemistic purposes; however, it is also used for humoristic purposes like in jokes, conversations as well as other scenarios of everyday life. Resources like these could be used in more specific problems such as computational humor and double entendre.

In this work, sexual slang words were collected from two important texts written by Mexicans: “Antología del albur” [3], which is a collection of albures from several states of the Mexican Republic of different authors; and “El chilangonario” [4], which is a manual compilation of general slang used by Mexicans. Later, words were selected and organized in a morphological dictionary with their corresponding lemma and grammatical tag. For resources creation, the same format as Freeling dictionaries was used, which allows to easily implement them in programs. Freeling is a wide complete library that provides many processing language services that can be used for convenience according to the application to develop.

Furthermore, a semantic dictionary was generated as complement to the morphological dictionary. This way a representative meaning for each entry of the sexual slang dictionary is provided. Meanings from this dictionary are not explanatory like in a book but rather representative through synsets from WordNet. A synset is a unique identifier for a world concept which is used in a semantic level. A synset can be associated to one or more words, for example, 02084071-n is the synset for the concept of “dog” but either “doggy” or “dogs”. The remaining information such as number and genre should be contained in the grammatical tag of the word. In this semantic dictionary a synset is assigned to one or more lemmas from the morphological dictionary.

2 State of The Art

The usefulness of computational humor and its motivations has been discussed many times. In [1] it is mentioned that “computational humor can make computers user-friendlier, more persuasive, likeable and competent improving the human-computer interaction in general, developing better intelligent agents, improving second language learning systems, electronic advertising and e-commerce”. In [5] it is mentioned that humor detection can be used to discard irrelevant information in a web search. In [6] Rasking suggest that “an application will search for [unintended] humor, for perversion of the text, for instance, in a presidential address, diplomatic note, or any other deadly serious business”. [...] “On the other hand, the same humor detection applications can be used to determine the vulnerable spots in a text to be denigrated, e.g. in a political campaign, and then to work in conjunction with humor generation to create appropriate and effective humor”.

Humor is definitely a human complex feature. It has been studied in many fields of knowledge like psychology, philosophy, linguistics, sociology and literature [7]. However, and despite the importance in human life, only a few investigations have been performed in this in computational field. Following the most representative works in NLP about humor are presented.

In [8] a methodology for *knock*, *knock* jokes recognizing are proposed. These are short texts with a well-defined structure and consequently easier to study. It is a type of verbal humor that use wordplay to produce the hilarious effect. This verbal game is mainly based on paronymy, homonyms and homophones. Authors identified three types of *knock*, *knock* jokes and they were limited to develop the methodology of only one type. Detection is

performed using statistical language recognizing techniques without any abstract structure for world knowledge representation. The tool developed has a wordplay sequences generator which with help of a phonetic similarity table produces a phrase similar in sound to another but with different meaning.

In [9] a novel approach to solve the identification of TWSS (That's What She Said) jokes is presented: DEviaNT. This method applies some metaphor recognition techniques and explores a particular approach to solve the TWSS problem: recognizing euphemistic and structural relationships between the source domain and an erotic domain. The authors report a precision value over 71.4%, however they affirm that this measure could achieve a value of 0.995 in certain conditions. The TWSS jokes are to say the phrase "that's what she said" after someone else has enunciated a phrase that, although it was not used in a sexual context, could have been used that way. DEviaNT classified as positive a sentence if it sounds funny after adding the phrase "that's what she said". In this work, the authors identified that TWSS jokes have a similar structure with sentences of erotic domain, e.g. a sentence of the form "[subject] stuck [object] in" or "[subject] could eat [object] all day" is more likely to be a TWSS than not. TWSS jokes have some similarity with the Mexican double entendre. In Mexico, a sentence is considered funny if at the end the phrase "sin albur" is added. Thus, the double entendre is produced and the previous sentence is interpreted as if it was said in a sexual context. Albures based on lexical ambiguity (not those based on wordplay) also share common structures with sentences of erotic domain as shown below.

In [10] a method to produce test and training data sets for using in tasks for recognizing different types of human speech such as humor, sarcasm, insults and profanity. The construction process of relevant helper data sets such as lists of profanity, insults lists and list of projects with their codes of conduct is also described. To create data sets specifically focused on profanities and insults, some FLOSS (Free/Libre and Open Source Software) projects were analyzed since most of them are developed using media that are archived and transparent such as mailing lists email and chat IRC (Internet Relay Chat). Some of the FLOSS projects involved are: Apache, Debian, Ubuntu, WordPress, Joomla!, Django, Drupal, among others. Codes of conduct of these were sought to know what behavior is considered inappropriate and unacceptable.

The part of detection of insults in which this work is focused is to distinguish between code-based insults versus personal insults. To that end, they created a list of insulting gleaned from the LKML postings by Linus Torvalds from 1995-2014. They read the entirety of those postings, then created a data set of the insulting sentence or phrase. They have been tagged whether the insult refers to code, personality, or both. This with the intention of documenting incidents in which personal insults are in the LKML to learn the difference between code-based insults; and to create a list of insults that could be used to train a Linus-style insult classifier.

Finally, data sets were created for three types of gender insults: the TWSS jokes (That's What She Said) containing sexual double entendre; maternal insults, which are those in which mothers and grandmothers used as a personal insult; and others in which older women (grandmothers and older) are used to represent an unintelligent or unsophisticated person ("Even Grandma can use the software!").

In [11] some models for classify vulgar and obscene phrases are proposed for short texts in Spanish. The authors generated statistics about which are the states of the Republic that use more this type of language. With help of a mexicanisms dictionary and WEKA software they constructed two models of classification: obscenity and none and; vulgarity and none. A precision of 91.07% are reported for the first one and 98.90% for the second one using SMO algorithm in WEKA.

In [12] *DATHCE* (Detector Automático de Humor en Textos Cortos en Español) an automatic detector of humor is presented. This software identifies some types of humorous texts: *albur*, adult slang, alliteration, rhyme and combinations of these. The software process two separate files: humorous and non-humorous texts. Subsequently, each module (one for every type of humor) assigns a weight to each word. The *albur* module use a frequent terms dictionary for this type of humor with about 200 words.

3 Techniques in Mexican Sexual Slang

Slang is an extension of conventional language. The most of sexual slang words are no present in official dictionary of Spanish but rather are transmitted from speaker to speaker and they suffer modifications, adjustments and/or corruptions in process. Most words from this slang refer to genitals or sexual actions. In some cases, these words are used as a special code to hide the real meaning of a sentence and these are considered taboo. In other cases, these are

used to make more comical sentences and cause hilarity in the listener. In order to achieve these goals different techniques are used, which furthermore help to generate new words. The following describes the most popular techniques in this slang.

3.1 Euphemisms

A euphemism is a word used instead of another by considered offensive, vulgar or inappropriate. The condition for a word becomes a euphemism of another is that exist some relation between both concepts like similarity, size, usage, etc., and thereby achieve an analogy or metaphor. For example, the euphemisms to say that someone is dead in Mexican Spanish are: “colgó los guantes/tenis” (hang up the gloves/sneakers), “pasó a mejor vida” (moved to better life), “se nos adelantó” (he/she comes forward), just to name a few.

In case of sexual slang, euphemisms are generated in the same way, e.g. “huevos” (eggs) are used to refer to testicles by their similarity in form; “lavar a mano” (hand washing) as metaphor of masturbation. In some cases, a euphemistic word has become so popular that its sexual meaning becomes predominant over the original, then it is called dysphemism. Table 1 shows some examples of dysphemism in Mexican Spanish.

Table 1: Examples of dysphemism and their meanings

Dysphemisms	Original meaning	Sexual meaning
Papaya (papaya)	A fruit	Vulva/vagina
Pajarito (birdie)	A little bird	Phallus
Verga (spar)	Rod: large and thin stick	Phallus
Panocha	A bread	Vulva/vagina
Huevos (eggs)	Rounded body which contains the germ of an embryo	Testicles

3.2 Corruptions

A corruption occurs when one or more letters (and hence phonemes) are changed in a word to produce a new one, the latter keeps the meaning of the initial word. In most cases, conventional Spanish words are used as final words especially if its root is equal or very similar than initial one, e.g. “anís” (anise) as corruption of “ano” (anus). This way “anís” acquires the meaning of “ano” in a sexual context. Thus, a hidden meaning is achieved because “anís” is an existent word in conventional Spanish and listeners supposed it as the only one or as the major priority meaning.

3.3 Metaplasms

There are three types of metaplasms: by addition, by suppression and by transposition. In the case of the Mexican sexual slang the most used are paragoge and prostheses. The first one involves adding one or more letters/phonemes at the beginning of the word, e.g. “Casiano” (a proper name) or “Herculano” (a city name) to refer to “ano” (anus). The second one involves adding one or more phonemes at the ending of the word, e.g. “anófeles” (anopheles) also referring to “ano”.

Speakers apply these techniques not only to words from conventional Spanish but also to slang words. Thus, combination of these techniques is performed, e.g.: euphemisms and metaplasms, “Aniceto” (a proper name) is the result of applying paragoge to “anís” which is in turn a corruption of “ano”. Commonly these techniques and combinations are performed at phonetic level.

4 Morphological Dictionary of Sexual Slang

A collection of sexual slang words was made from two important resources: “El chilangonario” which is a descriptive dictionary of Mexican slang; and a corpus of albuers. Both are written in Spanish language by Mexicans from several states of republic. Some of the collected words are the result of applying paragoge or prostheses to others. In order to avoid a large number of entries in the dictionary, we consider that these cases must be addressed in others levels of language and therefore processed by a separate module. Below these resources are described.

4.1 El Chilangonario: A Mexican Slang Dictionary

“El chilangonario” is a Mexican slang dictionary whose content is not exclusively sexual. This book is a descriptive dictionary in which examples of use for each word are provided. That format made it easy the collecting sexual slang words because most of them was tagged as: “albur”, “adaptation”, “apocope”, “corruption”, “euphemism” or “synonym”. Some examples from this book are shown below:

Abajeño (someone from lowland): pubic hair
Anacleto (proper name): corruption and adaptation of “ano” (anus)
Aguacates (avocados): figuratively, testicles
Botiquín (kit): adaptation of “bote” (boat). Buttocks, back
Café (brown): mierda (shit)
Gayo (“gay” from English): homosexual (homosexual)
Huevo/Güevo (egg): testicle
Joyo: corruption of “hoyo” (hole)
Livais: corruption of Levi’s brand meaning lesbian.
Micrófono (microphone): metaphor for penis

4.2 Datasets

The comic effect of wordplay (like albur) usually lies in the alternative interpretation of this, i.e. it consists in make a new lexical construction based on the sound of the original phrase. Starting the phonetic representation of a phrase is possible to construct a new one. The resultant phrase is homophone or very similar in sound to the initial phrase. Hereinafter this alternative interpretation will be referred as “double entendre phrase”.

The corpus of albures was created from the book “Antología del albur” [1]. These albures were written by many people from different states of the Mexican Republic. For that reason, this book is considered an excellent representative sample of sexual slang of Mexico. They sent to the author their own albures which were organized into an anthology. Despite the title, this book contains some texts that are not albures. After making a manual selection of albures a corpus with 820 sentences was formed. The average length is four words per sentence.

Albures always have a hidden sexual meaning which makes impossible to extract a list of words of sexual slang. For that reason, it was necessary for each albur annotate its hidden meaning in a new dataset which was named “double entendre phrases”. These new phrases are not completely explicit but they have sexual slang, mostly euphemistic nouns. Table 2 shows the correspondence between an albur and a double entendre phrase.

Table 2: Albures and their hidden meanings or double entendre phrases

Albur (Apparently, names)	Double entendre phrase (Hidden meaning of albur)
A. Garráis el chico	Agarráis el chico (Grab the boy)
Berta Nates	Ver tanates (Watch testicles)
Damesio John	Dame ese hoyón (Give me that hole)
Penélope Díaz	Pene lo pedías (Penis you asked)

Afterwards, each double entendre phrase was “translated” by a human who knows the sexual slang, i.e. each sexual slang word was replaced by its corresponding correct word. This way, the dataset “Erotic Phrases” was created. In table 3 some phrases from this new dataset are shown.

Table 3: Sexual phrases generated from double entendre phrases

Double entendre phrase	Erotic phrase
Agarráis el chico	Agarráis el ano (Grab the anus)
Ver tanates	Ver testículos (Watch testicles)
Dame ese hoyón	Dame ese ano (Give me that anus)
Pene lo pedías	Pene lo pedías (Penis you asked)

Since there is a mutual correspondence between the erotic phrases dataset and double entendre phrases dataset was possible to design an automated process that generated a list of sexual slang and its correct words. Table 4 shows some words from that list. By performing this mapping between domains, it is shown that sexual slang words [right column] correspond to only 27 words of conventional Spanish.

Table 4: List of sexual slang and their corresponding correct words

Sexual slang	Correct word
Chico (small/little/boy)	Recto (rectus)
Hoyón (hole)	Recto (rectus)
Vergara (a last name)	Falo (phallus)
Tanates (testicles)	Testículos (testicles)

As shown, the number of words in the erotic domain is greater than the number of meanings in conventional Spanish; seen as sets, an element of the erotic domain is mapped to at least one (or many) elements of the "conventional Spanish" domain. E.g. in a sexual context, words like "chico", "hoyón" or "Anacleto" are used to referring to rectus. Thus, for practical purposes it can be concluded that in a sexual context all these words [sexual slang and conventional Spanish] are synonymous.

Every collected word of sexual slang (sexual slang column from table 4) were used to form the morphological dictionary. We adopted the Freeling dictionaries format to make this resource easy-to-use by other researchers. It is required to annotate all possible pairs lemma-tag of a word. The lemma is that unique form which is common to all possible variations of a word, e.g. for nouns it is used the male gender and the singular form ["dog" is the lemma of "dogs", "doggy"]; for verbs, it is used the infinitive form ["eat" is the lemma of "eating", "ate"]. Thus, the following result was produced. The label is a series of symbols that provides morphological information of the word; in this case the labels proposed by the EAGLES¹ group are used.

```
chico chico NCMS000
tanates tanate NCMP000
verga verga NCFS000
hoyo hoyo NCMS000
abajeño abajeño NCMS000
anacleto anacleto NCMS000
botiquín botiquín NCMS000
```

In total, this dictionary contains 410 words which corresponds to 350 lemmas. However, the number of words that can be analyzed increases if affixation rules (of Freeling) are used. The morphological dictionary contains all those words that are the result of the creativity of the speakers who use the techniques explained in section 3 to produce them. However, the scope of this resource is limited to the morphological level consequently it is not possible to give senses to words, i.e. a semantic analysis is not achieved.

5 Semantic Dictionary of Sexual Slang

In order to provide a meaning for each word of sexual slang, a second resource was created: the semantic dictionary, which provides that information using "synsets". A synset is a unique code used to represent a specific concept of the world. A synset is associated to lemmas, not to words, e.g. 02084071-n is the synset for "dog" concept. The lemmas associated to this word are "dog", "canis_familiaris" and "domestic_dog". Synsets are language independent, thus, in Spanish, the lemmas associated to this synset are "perro" and "can".

In table 4, sexual slang words and their equivalents in conventional Spanish were presented. Now, for the creation of the semantic dictionary, the synsets for those words (of conventional Spanish) were searched. The synsets originally come from the WordNet [13] database, however, this resource is made only for the English

¹ <http://www.ilc.cnr.it/EAGLES96/home.html>

language, so it was necessary to search in the MCR [14] (Multilingual Central Repository). This database reuses the synsets of WordNet for other languages such as Spanish, Portuguese, Catalan, Basque and Galician. Table 5 shows the assignment of synsets to their corresponding right words [conventional Spanish].

Table 5: Assigning synsets to correct words

Correct word (Conventional Spanish)	Synset
Recto (rectus)	05538016-n
Falo (phallus)	05526713-n
Testículos (testicles)	05524615-n
Vello púbico (pubic hair)	05263587-n
Nalgas, trasero (buttocks, back)	05559256-n

Since in a sexual context, all words of sexual slang are considered synonyms of these [conventional Spanish], then they can be grouped through the same synset.

05538016-n ano chico hoyo anacleto
05526713-n falo pizarro verga
05524615-n testículo tanate
05263587-n vello_púbico abajeño
05559256-n nalgas trasero botiquín

It was observed that some words, in addition to semantic changes, they suffer grammatical changes in the sexual slang, e.g. "largo" which is an adjective, it is nominalized by prefixing an appropriate article ["el largo"] and it acquires the meaning of phallus. The synset 05526713-n [concept of phallus] has the highest number of lemmas [112], followed by 05538016-n [concept of rectus] with 53 lemmas.

Both resources [morphological and semantic dictionaries] are complementary to those of conventional Spanish, i.e. in practice they are used to enrich and extend the analysis of Spanish language. Therefore, it is possible to include lemmas of conventional Spanish on these resources.

6 Implementation

These resources can be used with Freeling in two different ways: fused together with Freeling dictionaries, this way the analysis is performed for the sexual slang and for the conventional Spanish; or independent analysis, the analysis of sexual slang is performed after the analysis of conventional Spanish [with Freeling].

With Freeling, when a word is searched in dictionary, an analysis list is created. An analysis consists in a possible lemma-tag for the word. This occurs because the word can have more than one meaning depending on the adjacent words [context], can be a homographs case. Since the semantics disambiguation is performed at a later stage, at this point all possible analysis for the word are kept. When the semantic analysis is performed, all possible meanings (synsets) for each word are annotated. Fig. 1 shows the analysis performed to the word "chico", using conventional dictionaries [Freeling dictionaries].

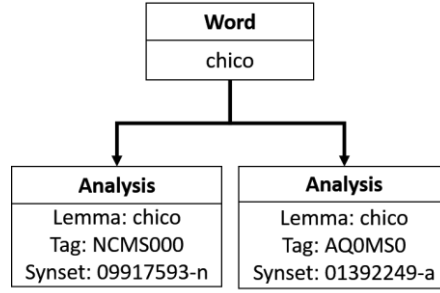


Figure 1: Analysis list of a word

Fig. 2 shows the result that would be obtained by enriching the word "chico" with Mexican sexual slang. In contrast to Fig. 1, now appears the synset 05538016-n [concept of "rectus"], this word has three meanings: as a noun refers to a boy, as an adjective refers to something small, and in the sexual slang means rectus.

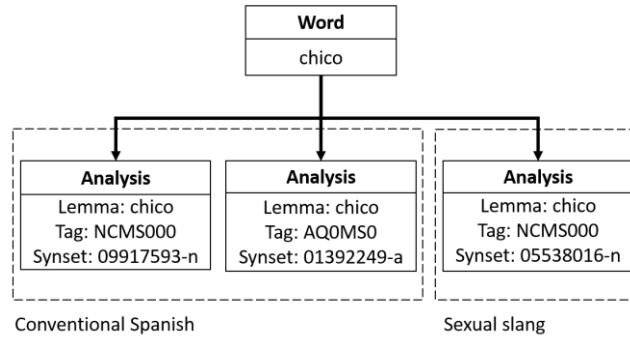


Figure 2: Sexual slang and conventional Spanish enrichment for a word

7 Future Work

As mentioned in section 3, in Mexican sexual slang, metaplasms are used to hide the intended sense of a word. Paragoge and protheses, besides they the most used, they are also the easiest to tackle. This is because they consist of adding one or more letters (and hence phonemes) at the beginning or end of a word respectively. When using these metaplasms, the resulting word takes on the meaning of the initial word, e.g. often in sexual slang "anófeles" is used instead of simply "ano".

In humor and double entendre, also phonetic ambiguity is exploited to go unnoticed, once the receiver discovers the real meaning of the phrase or word, comedy occurs. As future work, it is proposed search all those words of conventional Spanish that can be used as protheses or paragoge of erotic words. For best results, this search would be done phonetically to consider those words whose ends match in sound but not in spelling, ie, are homophones but not homographs.

To generate the phonetic form of words, a series of phonological rules created from alphabet SAMPA is proposed. The reasons for making a modification to this alphabet are: a) generalize sounds (phonemes and allophones) and b) completely ignore the accents and intonations.

A phonological rule is composed of three elements: letter(s), replacement and, context. Replacement is a symbol of the phonetic alphabet that corresponds to the sound of one or more letters (first element). The context is the condition that the first element must comply to be replaced by the second. In table 7 our phonological rules set is defined.

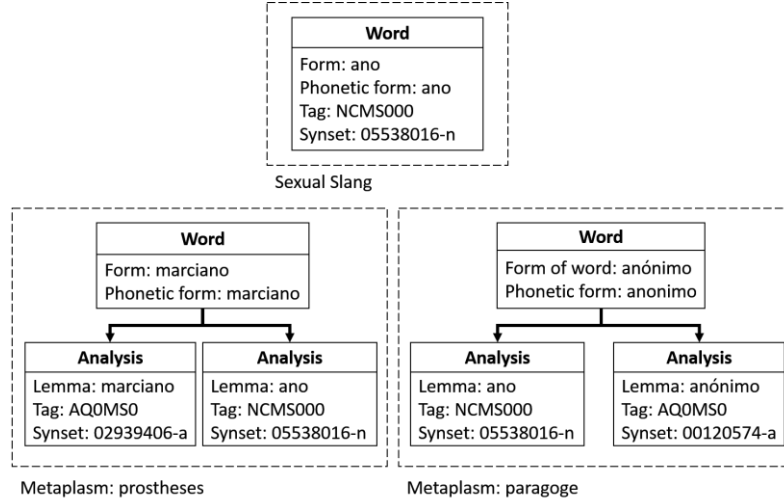
Table 7: Phonological rules set

Categories		
W=ei (vocales débiles)		
F=aou (vocales fuertes)		
B=pb (bilabiales)		
Phonological rules ²		
rr/R/_	q/k/_	z/s/_
r/R/^_	c/s/_W	h//_
ch/X/_	c/k/_	v/b/_
x/X/_	g/j/_W	y/i/_
qu/k/_W	gu_g/_W	ñ/J/_
qu/ku/_F	gü/hu/_	n/m/_B
qü/ku/_	ll/i/_	n/m/_f

Once generated phonetic forms of words, it is possible compare them each other to find out which ones can be considered parago or prosthesis of words of sexual slang, e.g. "marciano" and "anónimo" would be annotated with the lemma-tag pair of "ano", generating the following results.

marciano marciano AQ0MS0 ano NCMS000
anónimo anónimo AQ0MS0 ano NCMS000

Since a "marciano" and "ano" has been assigned the lemma "ano", then in semantic analysis, these, in addition to its appropriate synsets, would be annotated with the synset 05538016-n [concept of anus] as shown in Fig. 3.

**Figure 3:** Semantic enrichment with metaplasms at phonetic level

8 Conclusions

In this work two important resources of Mexican sexual slang are presented: a morphological dictionary and a semantic dictionary which are freely available online³. Since these are based on Freeling library, both are easy-to-implement in programs. It is possible to use these dictionaries in applications related to computational humor and double entendre but, they are also useful for other NLP tasks like analysis for texts written in Spanish by Mexicans.

² Symbol “^” indicates the first element (letter r) must appear at the beginning of a word.

³ https://github.com/robertovillarejo/mexican_sexual_slang

Since double entendre and humor are often based on lexical ambiguity, these dictionaries could be used for those purposes.

In verbal humor, various mechanisms such as lexical ambiguity, semantic ambiguity, phonetic ambiguity, etc. are used as well as rhetorical figures and wordplay. Often the humor detection problem involves the ignorance of culture. In the case of albur, a type of Mexican humor, among other abilities, one must know the sexual slang of that specific region.

In Mexican culture, albur and phrases with sexual charge, under certain conditions, are amusing, however, in other regions of the world this may be offensive or even taboo. The dictionaries presented in this paper could be used to improve NLP systems, since these could be used to help a machine to discern the meanings of words. That way, systems which have a direct interaction with humans (like bots) could avoid learn offensive language and thus keep a friendly relationship with users. Other uses for these dictionaries: a) to serve as auxiliaries in question-answer systems; b) serve as seeds to collect erotic texts and thus conform new datasets.

The most common techniques and resources were identified. Also, a process to enrich words semantically based on a phonetic approach with metaplasms and sexual slang was proposed as future work. It is possible to use the proposed process to enrich the morphological dictionary with these techniques instead of performing the metaplasms analysis on runtime.

References

- [1] J. M. Taylor, "Toward Informal Computer Human Communication: Detecting Humor in a Restricted Domain," 2008.
- [2] C. I. Nass, J. Steuer, and E. R. Tauber, "Computers are social actors," *Comput. Interact. Conf. Celebr. Interdepend.* 1994, pp. 72–78, 1994.
- [3] A. Peralta de Legarreta, *El chilangonario*, México, D.F.: Algarabía Editorial, 2012.
- [4] V. Hernández, *Antología del albur*, Toliro Multimedia and Incognita|Caja Negra, 2006.
- [5] J. M. Taylor, "Ontology-based view of natural language meaning: The case of humor detection," *J. Ambient Intell. Humaniz. Comput.*, vol. 1, pp. 221–234, 2010.
- [6] V. Raskin and J. M. Taylor, "Preface: Artificial Intelligence of Humor — Computational Humor," no. 1994, pp. 2–3, 2005.
- [7] M. P. Mulder and A. Nijholt, "Humour research: State of the art," *CTIT Tech. Reports Ser.*, 2002.
- [8] J. M. Taylor, "Computational Recognition of Humor in A Focused Domain," University of Cincinnati, 2008.
- [9] C. Kiddon and Y. Brun, "That's What She Said: Double Entendre Identification," *Comput. Linguist.*, pp. 89–94, 2011.
- [10] M. Squire and R. Gazda, "FLOSS as a source for profanity and insults: Collecting the data," *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, vol. 2015-March, pp. 5290–5298, 2015.
- [11] E. Guzmán, B. Beltrán, M. Tovar, A. Vázquez, and R. Martínez, "Clasificación de frases obscenas o vulgares dentro de tweets," *Res. Comput. Sci.*, vol. 85, no. 2014, pp. 65–74, 2014.
- [12] R. Ocampo Pólito, "Detección automática de humor en textos cortos en español," Instituto Politécnico Nacional, 2010.
- [13] G. a. Miller, "WordNet: a lexical database for English," *Commun. ACM*, vol. 38, no. 11, pp. 39–41, 1995.
- [14] L. E. and R. G. Gonzalez-Agirre A., "Multilingual Central Repository version 3.0: upgrading a very large lexical knowledge base," *Proc. Sixth Int. Glob. WordNet Conf.*, 2012.

Appendix