Article

Figurative Language in Atypical Contexts: Searching for Creativity in Narco Language

Antonio Reyes *[ID] and Rafael Saldívar [ID]

School of Languages, Autonomous University of Baja California, Mexicali 21100, Mexico; rafaelsaldívar@uabc.edu.mx
* Correspondence: areyes98@uabc.edu.mx

Abstract: Literal language is commonly defined in terms of direct meaning, i.e., any literal utterance must convey a unique meaning. Such meaning has to be the one conventionally accepted to guarantee a successful communication. Figurative language, on the other hand, could be regarded as the opposite of literal language. Thus, whereas the latter is assumed to communicate a direct and explicit meaning, figurative language is related to the communication of veiled or implicit meanings. For instance, the word *pozolero* (stewmaker), which literally refers to a person who cooks a traditional Mexican food, when it is used in a figurative utterance, it can refer to different concepts, which are hardly related to food. Therefore, it can work instead of hitman, murderer, drug dealer, and others, in such a way its literal meaning is intentionally deviated in favor of secondary interpretations. In this regard, we are focused on analyzing the use of figurative language in an atypical context: drug trafficking. To this end, a corpus about narco language in Spanish was built. This corpus was used to train a word embedding model to identify creative ways to name narco-related concepts. The results show that various concepts are commonly expressed through figurative devices, such as metaphor, metonymy, or mental imagery. This fact corroborates that figurative language is quite recurrent in our daily communication, regardless of the context. In addition, we show how this *creativity* can be recognized by applying Natural Language Processing (NLP) techniques.

Keywords: figurative language; narco language; metaphor; metonymy

1. Introduction

Traditionally, language has been described from dichotomous points of view: langue and parole, synchrony and diachrony, paradigmatic and syntagmatic, signifier and signified, oral and written, an so on. In this article, we assume another dichotomous relationship: literal language and figurative language.

Literal language is commonly defined in terms of direct meaning, i.e., any literal utterance must convey a unique meaning. Such meaning has to be the one conventionally accepted to guarantee a successful communication. In Saussure's terms, literal meaning is corresponded with a perfect dichotomy of signifier and signified (cf. [1]). Various experts have pointed out certain properties of literal language: it is direct, grammatically specified, sentential, necessary, and context-free (see Searle [2]). Hence, it is assumed that this category of language must be invariant in all contexts.

On the other hand, figurative language could be regarded as the opposite of literal language. Thus, whereas the latter is assumed to communicate a direct and explicit meaning, figurative language is related to the transmission of an implicit, or veiled, meaning. For instance, the word *pozolero* (stewmaker), which literally refers to a person who cooks a traditional Mexican food, when it is used in a figurative utterance, it can refer to different concepts, which are hardly related to food. Therefore, it can work instead of hitman, murderer, drug dealer, and others, in such a way its literal meaning is intentionally deviated in favor of secondary interpretations.

In this context, it is easy to note that a grammatical description of language is not enough to explain why the link between a word and its literal meaning broke up. On this matter, it is known that grammar regulates language in order to have a non-chaotic system, but it is also known that language is a live entity, whose pragmatic components tend to determine the meaning of what is communicated. For instance, according to Fillmore et al. [3], the idioms have a grammatical structure which is not made intelligible only by knowledge of the familiar rules of its grammar, but by inferring pragmatic information to, eventually, make sense.

In this regard, Ronald Langacker ([4]), from a cognitive perspective, states that language is a continuum of symbolic structures in which lexicon, morphology, and syntax form a continuum which differs along various parameters. This means that there are several components, not only grammar, to conform our language. Those components are embedded in a global system which, according to the cognitive linguistics foundations, depends on cognitive, experiential, and social contexts, which go far beyond the linguistic system proper (Langacker [4], Kemmer [5], Goldberg [6]).

This cognitive vision highlights the relevance of conceptualization to constitute the meaning of what is communicated. Such conceptualization is based on conventional images of the reality, in which both referential and inferential knowledge are fundamental components to shape and verbalize our reality.

In accordance with this cognitive and conceptual vision of language, in which grammar is as important as pragmatics, we will subscribe the arguments to describe, analyze and support our approach. In particular, we are focused on analyzing the use of figurative language in an atypical context: drug trafficking. On this matter, we are interested in applying NLP techniques to identify figurative devices, such as metaphor, metonymy, or mental imagery, as they are understood in Cognitive Grammar (see Langacker [4]), in a corpus about narco language in Spanish. To this end, we introduce a specialized corpus on this topic. This corpus is used to train a word embedding model to automatically identify creative ways to refer to narco-related concepts.

The rest of the article is organized as follows: In Section 2, the theoretical background about figurative language is described, emphasizing the state-of-the-art about figurative language processing. In Section 3, we introduce the context of narco language and the corpus we used in our work. The methods to carry out the experiments, as well as the human validation, are detailed in Section 4. In Section 5, we report our results and discuss the findings. Finally, in Section 6, we outline the final remarks and some pointers to address the future work.

2. Figurative Language

In the context of a dichotomous view of language, figurative language could be considered as the opposite of the literal one. This simple distinction seems to be sufficient on its own; however, figurative language is more than a straightforward deviant use of language. In this respect, Peters [7] points out that this type of language entails cognitive processes to build different layers of meaning. Therefore, it is necessary to go deeper into the mechanisms and processes that differentiate figurative from literal language.

According to the classical perspectives, the notions of literalness and figurativity are viewed as pertaining directly to language, i.e., words have literal meanings, and can be used figuratively (Searle [2]). Hence, it could be assumed that figurative language is grounded on literal meaning. This latter is disconnected from what people learn about the world [or about the words] to trigger a secondary meaning (see Bergen [8]).

By disconnecting what people know about the words, literal language becomes senseless and, then, the figurative interpretation is triggered. Let us consider Chomsky's famous example to explain this fact:

1. "Colorless green ideas sleep furiously" [9].

Example 1 highlights how grammar supports the decoding process to recognize such utterance as part of the English language. Either orthographically, phonologically or even

syntactically, Chomsky's example is understandable in terms of its linguistic constituents. However, when interpreting, its literal meaning is completely nonsensical. For instance, the bigrams [colorless green] or [green ideas] are disconnected from their conventional referents for being able to produce a coherent interpretation. Thus, in order to make the example understandable, secondary interpretations are needed. If such interpretations are successfully activated, then figurative meaning is triggered and, accordingly, a more coherent interpretation can be achieved. From this distinction, it can be argued that literal language is denotative, that is, the meaning tends to be lexicalized, whereas figurative language is connotative, hence, its meaning tends to be implicated.

In addition, Katz et al. [10] note that much figurative meaning is based on learned convention, such as with idioms, and proverbs. That means, it is pragmatically motivated. In this respect, it can be stated that figurative language plays an important role on communication due to the need of performing mental processes, such as reasoning and inferencing (Peters [7]), which require additional cognitive effort (Gibbs [11]). Moreover, Lönneker-Rodman and Narayanan [12] point out that figurative language can tap into conceptual and linguistic knowledge (as in the case of idioms, metaphor, and some metonymies), as well as evoke pragmatic factors in interpretation (as in indirect speech acts, humor, irony, or sarcasm).

Although the arguments given so far, a major issue still remains: There are no formal patterns to affirm that a word or phrase is more figurative than other. In this regard, it is necessary to stress out a crucial extra-linguistic element (with linguistic repercussion) to differentiate literalness from figurativity: *Intentionality*. According to Reyes et al. [13], the intention of maximizing a communicative success could support the identification of formal elements to differentiate both types of languages. Pragmatically, speakers are guided by a specific intention when communicating. Such intention is supposed to be consciously formalized with linguistic items (for instance, intonation, words or, even, syntactic structures) to successfully communicate something. It is likely that figurative meaning cannot be accomplished straightforwardly, but this failure will not relay on the speaker's intention, rather, on the hearer's skills to interpret what is communicated figuratively. Let us observe the following sentence to clarify this point.

2.   "La cocaína es transportada usualmente en camiones" (Cocaine is often transported by truck).
3.   "Desearía tener un poco de polvo feliz" (I wish I had some happy powder).

Whereas in Example 2 the intention is to describe how the drug is moved, in Example 3 the intention is to communicate a veiled meaning, which is understandable by adjusting literalness to a figurative interpretation, in this case, a metaphorical one. Thus, in the first example, the communicative success is based on making a precise description about the *how* (note that all the words in this context are very clear in terms of their semantic meaning), whereas in the second, the success is achieved by deliberately selecting items that entail secondary and non-literal relations: [tener—polvo], [polvo—feliz]. In addition, Example 3 highlights narco language complexity, as it involves sociocultural aspects to understand the figurative content. In this respect, we consider approaching this complexity from a cognitive linguistic view. Thus, the figurative devices we analyze, although described from a rhetorical perspective, will be approached from a cognitive standpoint. In particular, from what Lakoff and Johnson [14] introduced as conceptual metonymy and conceptual metaphor (see Section 5.3 below).

Finally, it is necessary to emphasize that language provides specific linguistic devices to express figurative content by different ways. For instance, metaphor, metonymy, similes, or irony.

Below is presented a brief review about the automatic processing of some of such figurative devices.

Figurative Language Processing

Figurative Language Processing (FLP) is a subfield of NLP, whose goal is on modeling figurative usages of natural language to be implemented computationally. Despite enormous challenges, there are many approaches to deal with FLP: From the automatic differentiation between literal and figurative language, to the development of models to recognize certain figurative devices in large collections of texts. For instance, Bogdanova [15] bases her approach to figurative language detection on the fact that the sense of a word significantly differs from the sense of the surrounding text. To her, this is an insight about how a word is used figuratively. In the same vein, Li and Sporleder [16] use gaussian mixture models to detect figurative language. They assume that figurative language manifests less semantic cohesive ties with the context than literal language. Rentoumi et al. [17], in turn, propose a methodology for sentiment analysis of figurative language which applies word sense disambiguation and Hidden Markov Models. On the other hand, Sikos et al. [18] use experimental techniques to understand the cognitive and neural mechanisms that underlie figurative interpretations. According to the authors, figurative language forces the two cerebral hemispheres to work in parallel in order to reach an appropriate interpretation.

With respect to specific figurative devices, one of the most analyzed in the past few years is irony. This device is a conscious use of language in which a speaker implies the opposite of what is literally said (Colston and Gibbs [19]). Some of first works on irony are reported by Utsumi [20], Veale and Hao [21,22], as well as by Reyes et al. [13]. More recently, Bosco et al. [23], and Tang and Chen [24] have developed ad hoc corpora for the task in languages beyond English. Likewise, some other researchers have addressed the task by setting a social media scenario in which it is quite common to find ironic statements about anything. For instance, Buschmeier et al. [25] focus their approach on product reviews. Whalen et al. [26], on personal blogs. Fersini et al. [27] or Karoui et al. [28] on microblogs. Wallace et al. [29], in turn, investigate irony in broader scenarios such as online communities.

On the other hand, idiomatic expressions are good examples regarding the use of figurative content. Unlike metaphoric, metonymic, or ironic expressions, whose ways of expression have no limits, idioms can be recognized by people effortlessly; i.e., an idiomatic expression is finite in terms of possibilities of expression. In this context, Li and Sporleder [16] analyzed the degree of fixedness in syntactic structures to identify idioms. In turn, Feldman and Peng [30] pointed out that idiomatic expressions have low semantic proximity, therefore, they are likely to be easily identified.

With respect to similes, Veale and Hao [31,32] have studied how the markedness of similes allows for a large case-base of illustrative examples to be easily acquired from the web. In their first work, the authors presented a system that used these examples both to understand property-attribution metaphors and to generate apt metaphors for a given target on demand. In the latter, they retrieved explicit similes with the pattern "as ADJ as a | an NOUN" to demonstrate how web users often use similes to post ironic comments.

Defined as a simple comparison, the approaches regarding metaphor involve different tasks. For instance, Rentoumi et al. [17] addressed metaphor processing with a sentiment analysis task. They argued that metaphors can be used as expressive subjective elements since they display polarity implicitly. Saygin [33], in turn, analyzed the role that metaphor plays in translation, specifically, with Turkish and English. According to her results, when people translated sentences to and from their native language and their second language, both the metaphoric and the literal meaning tend to be active, even though it is clear from the context which meaning is intended. From a different perspective, Veale [34] found some patterns about how people express metaphorical comparisons in an information retrieval task. In addition, Zinken [35] noted in a corpus of newspaper texts how metaphors have a regular pattern to make the comparisons. Furthermore, he found that such patterns are form-specific, which means that they are connected to particular lexical items. On the other hand, Pierce et al. [36] showed that metaphor processing tends to be automatic at

cognitive level, due to violations of semantic expectancies. This makes people consider a wider semantic neighborhood and, accordingly, to recognize the metaphoric meaning.

In the case of metonymy (referred as the representation of the whole for the part, or the use of one entity instead of another (Lönneker-Rodman and Narayanan [12])), the approaches are fewer. This is mainly due to the common frames that metonymy and metaphor share. Then, most of the research has focused on processing the latter. The former, however, has been faced from different perspectives. Those involve the use of specific linguistic resources such as thesaurus and corpus-based analysis. For instance, Peters and Wilks [37] describe the use of taxonomic information from WordNet to select instances of metonymy. With the same lexical resource, Peters [7] uses WordNet's hierarchical structure to identify instances of metonymy. In contrast, Markert and Nissim [38] addressed the task of metonymy resolution by employing corpus information to be able to discriminate between literal and metonymic usages of a word.

Finally, it is worth mentioning that not all the previous devices appear in the context of narco language, at least, according to the texts gathered for our corpus, which is described below.

3. Narco Language

At first glance, narco language can be classified as slang, insofar as it is a body of non-standard vocabulary used by some isolated social groups. However, in Mexico (and likely in others countries) this type of language has gone from marginality to common speech, by describing a reality in which violence, corruption, and institutional collapse predominate; everything derived from the phenomenon of drug trafficking. In this context, narco language is not used only by drug traffickers or by the police, but it has reached all social strata and is used by different actors.

It is necessary to specify that this term does not represent a language properly, but a lexicon to refer to narco-related concepts, i.e., as far as it has been reported, narco language has no particular morphological, syntactic, or even, phonological rules. It is featured, on the contrary, by a compendium of lexical items (some of them neologisms) and phrases that are often completely obscure for most people. In this respect, narco language has compiled an interesting linguistic inventory, which has been fed from different sources, such as mass media, literature, political discourse, or popular folklore.

Finally, as reported by Saldívar [39], some of the most frequent mechanisms to characterize narco concepts come from figurative language. For instance, metaphors, metonymies and analogies, either for cryptic purposes, attenuation, or as a simple exercise of creativity.

3.1. Figurative Language in Narco Language

According to Saldívar [39], one of the notable characteristics of the drug lexicon is its figurative nature, due to its crypticity, but also as a reflection of the creativity of the subjects involved. Although this type of lexicon has been of interest to mass media, linguistics and literature, such lexicon is elusive because of the speed with which it is transformed. Despite this fact, there have been some approaches from different angles to describe this type of language. In the following lines, we summarize some of them in order to understand the phenomenon of narco language with respect to figurative language.

Firstly, lexicographical studies on criminal slang and drug use have been of great interest in the field of applied linguistics, due to the way in which language is linked to a social problem that in recent years has been important in Latin America (see Acosta and Mora [40], Sanmartín [41]). Likewise, in his book about the phenomenon of drug trafficking as a phenomenon that generates changes in Mexican speech, Saldívar [42] described that two of the most productive mechanisms identified to create terms for drug names are metaphors and metonymies.

In their text about the prison lexicon of Colombia, Acosta and Mora [40] showed how the criminal jargon is characterized by the use of metaphors and metonymies in an ingenious way to refer to, for instance, the brand of an object. This seems to be repeated in

the language of drug trafficking in Mexico, since the technical lexicon associated with crime is saturated with metaphorical language (see Mattiello [43]). Torregrosa and Sánchez [44], in turn, suggested the study of conceptual metaphors in the use of drugs for educational purposes in the training of lawyers.

3.2. Corpus

In order to train a model for identifying figurative language in the context of narco, we built a specialized corpus on this topic in Spanish. The corpus contains documents from different sources to cover, as much as possible, a broad scenario about this language. Eventually, this coverage could shed light on how this social phenomenon has linguistically permeated our society.

The corpus is divided in six categories: blogs, political discourse, essays, literature, narcocorridos, and press. The documents for each category were retrieved both manually and automatically. In this respect, for some categories, the content about narco is lesser. For instance, the documents from press contain information about narco, but also about unrelated news. Same fact for political discourse. The rest of categories are more focused on the topic.

Some statistics about the corpus are given in Table 1.

Table 1. General statistics about the corpus.

	Tokens	Types
Blogs	229,338	29,585
Political	399,006	20,891
Essays	543,718	43,601
Literature	370,794	38,726
Narcocorridos	79,664	12,554
Press	1,728,165	83,514

The corpus is available upon request for academic purposes.

4. Model

In this section, we describe the word embedding model for representing our data, which are further analyzed in order to identify how figurative language is used to verbalize some narco-related concepts.

4.1. Word Embeddings

In the past few years, one of the most recurrent learning techniques used in NLP and Machine Learning (ML) to model syntactic and semantic information is word embeddings. In simple terms, the word embeddings could be defined as representations of words in a vector space by grouping similar items (Mikolov et al. [45]). More specifically, as described in [46], word embeddings are real-valued word representations, in which the co-occurrence of words, trained on natural language corpora, allows the prediction of syntactic and semantic properties. According to Almeida and Xexéo [47], word embeddings are divided in two main types: Prediction-based models and count-based models. The former relies on local data, i.e., the word's context, and is commonly linked to Neural Networks Models (NNM). On the other hand, the count-based models rely on global data, i.e., word's context considering the co-occurrence in the whole corpus.

Although the implementation of word embeddings in NLP tasks has highlighted some problems, such as there is no clear differentiation between semantic relatedness and semantic similarity (Bakarov [46]), their use in FLP is growing. For instance, the works described by Liu et al. [48], Aggarwal and Singh [49] regarding metaphor detection, as well as the ones devoted to irony detection (Nozza et al. [50], González et al. [51]), and idiom recognition (Peng and Feldman [52]).

4.2. Word2Vec

Given the efficiency for representing linguistic properties, several word embeddings implementations exist. For instance, Word2Vec, FastText, or GloVe. In this article, we adopted the Word2Vec algorithm, as described by Mikolov et al. [45,53].

This algorithm stresses the meaning and semantic relations among the words by computing the co-occurrence of words in different documents. In this respect, the word vectors that share some regularities are located nearly in the vector space. According to Dessì et al. [54], "the idea behind this algorithm is to model the context of words by exploiting ML and statistics to represent and come up with a vector representation for each word within the corpus. The resulting word vector representations allow the recognition of relatedness between words". This algorithm can be trained in two ways: By considering a Continuous Bag-Of-Words (CBOW) model or a Skip-gram model.

As mentioned previously, there are some FLP approaches based on implementing word embeddings; however, there is no pre-trained model to be used in the narco domain. Therefore, we used our corpus to train different embeddings representations with the Skip-gram model (it is worth noting that in the preliminary setup, the CBOW model was also used, but the results were not as interesting as with the skip-grams). The main differences in the embeddings are: Vector dimension, windows distance, and word frequency. Finally, in order to tune the vectors and get a unique model, we trained a final average model by finding the centroid of all the models. This final average model was used for the experiments reported below.

4.3. Identifying Triggers of Figurative Content

As stated in Section 3, apart from its crypticity, narco language is elusive because of the speed with which it is transformed. This impacts directly on the possibilities of focusing the search of figurative content on specific paths. For instance, on spatial or orientational paths (Lakoff and Johnson [14], Lakoff [55]). That is why we decided to use some words reported in the literature as prototypical of this domain to build a dictionary. Hypothetically, this dictionary could allow us to reduce the search space to locate figurative devices by focusing on documented narco-related words, i.e., instead of performing a blind search for all the words in our corpus, we concentrated on those words, which according to the experts, are often used in the narco context.

The dictionary contains 439 entries in Spanish. Some examples are drug, gun, salt, snow, ice, crystal, fuel, candy, chocolate, poison, line, nose, and others.

According to Sikos et al. [18], understanding figurative language often involves an interpretive adjustment to individual words, i.e., not all the words in a given sentence trigger the figurative meaning on their own; rather, this meaning is usually triggered by manipulating individual words. In this respect, by searching the most similar words for all the words in the dictionary, we aimed to identify recurrent ways for expressing narco-related concepts. For instance, if the words *joy* and *cocaine* share some distributional patters in the vector space, then the former could be a figurative means to map the latter. It is worth noting that the underlying assumption, at the moment, is to identify a set of words to trigger a figurative interpretation, not a figurative device in particular.

Of the 439 entries, only 183 appeared in our corpus. To reduce noise, for each word we retrieved their 10 most similar words. This means a total of 1830 words to be analyzed in search of triggers of figurative content. In Figure 1, we show the set of most similar words for an entry from the dictionary.
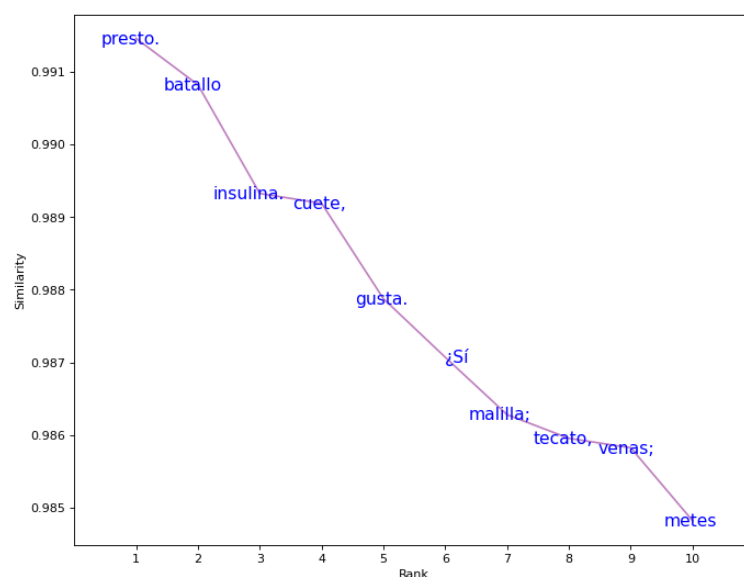
Figure 1. 10 most similar words for *insulin*.

Despite not being an expert in drugs or drug trafficking, the information depicted in this figure is very interesting. As everybody knows, insulin is a legal substance (hormone) for diabetes control. However, some of the related words do not refer to a disease domain at all. The words *gusta* and *metes* (inflectional forms for the verbs to like and to do drugs, respectively) are never used in the context of diabetes. Nobody says something like "voy a meterme insulina (I am going *to do insulin*?)". Another interesting word is *venas* (veins). Although insulin is commonly injected into the patient, it is unusual to be injected in the veins, unless the patient is at the hospital. On the contrary, there are several drugs to be injected in the veins to enhance the effect.

Although figurative language is a regular mechanism in our daily communicative interactions, it is totally unlikely that all of 1830 words could be triggers of figurative content. That is why we performed a cleaning to reduce the spectrum of possible triggers. The cleaning consisted of (i) eliminating duplicated words, (ii) joining plural forms to a root form, and (iii) removing synonyms and close-related words due to they are registered forms known for most people, i.e., they are not veiling anything. After the cleaning, the total number of potential triggers was 505.

5. Results

In this section, we describe the results of applying the word embedding model and discuss the implications. Finally, we provide some linguistic explanations to support our results.

5.1. Human Validation

Given the risky nature of the language we are investigating, as well as the consequent lack of available material to assess our findings, we contacted an expert on the matter in Mexico. This expert has published several academic papers and some books about narco language since 2009.

We asked him to revise the 505 words to (i) determine whether or not they are terms used in the narco context, (ii) identify, as far as he knows, whether or not they are used to refer or name any narco-related concept cryptically. The first task intended to validate that, at least, the words belong to the narco domain. The second one, on the other hand, to recognize words that are likely connoting a secondary meaning, i.e., possible triggers of figurative content. It is necessary to mention that no context was provided to the expert, just the words.

After analyzing the set of words, the expert gave us the following feedback: Regarding the first task, he validated all the 505 words as part of the narco language. With respect to the second task, he marked 151 words as items that are used to name narco-related concepts in a cryptic manner. Finally, he translated some of these words into their known terms. For instance, considering the information given in Figure 1, words such as *cuete* and *insulina*, or *yongo* were translated to syringe and place to do drugs, respectively. In Table 2, we provide some examples of the second task.

Table 2. Sample of cryptic words to name narco-related concepts.

adormidera	calentar plaza	falopero	perico
aguaje	churro	ice	petate
aspirina	clika	lancha	plaza
bajón	cuerear	levantar	pozolero
blanca	engordar vena	madrina	tlacuache
cachucha	enteipar	manguerear	vidrio

5.2. Narco Language Categories

Prior to analyzing the words marked by the expert, we decided to group them in categories with the purpose of finding patterns about their semantic relatedness, in such a way we could identify and explain what figurative device underlies their usage. To this end, we ran a similarity analysis for all these words considering their co-occurrences in the whole corpus. Figure 2 shows a sample of the semantic relatedness for some of the 151 triggers of figurative content.
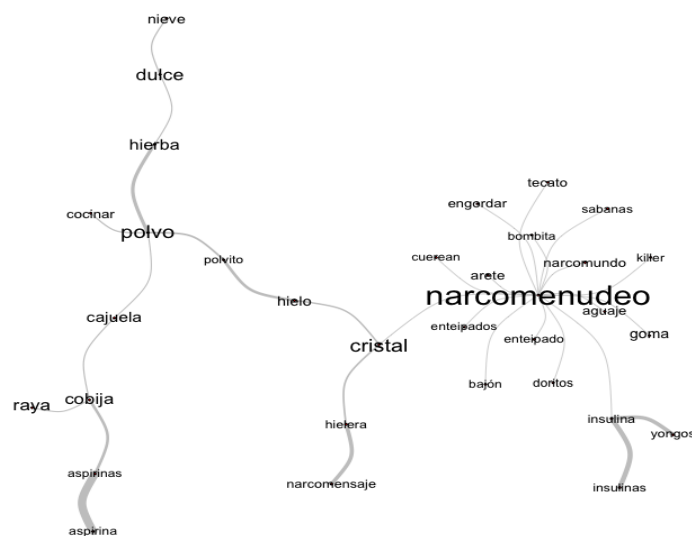

Figure 2. Sample of semantic relatedness for some triggers of figurative content.

By grouping these words according to their semantic relatedness, we could perceive some interesting patterns about common communicative topics from the linguistic contexts that such words share. This can be observed more clearly in Figure 3 below.
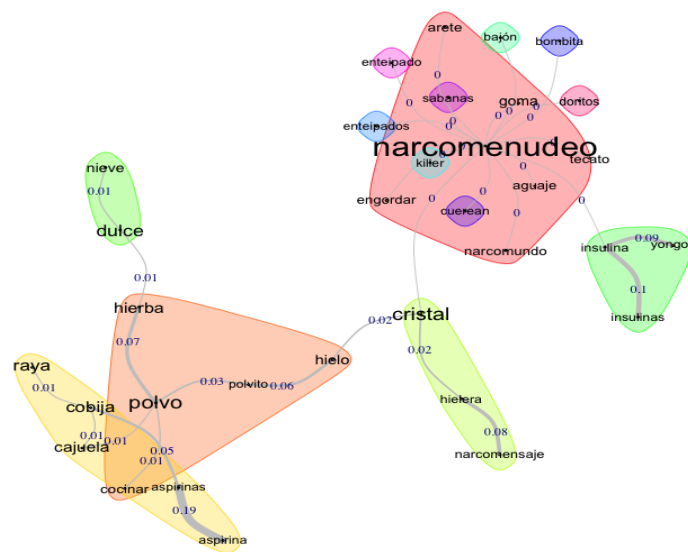
Figure 3. Sample of common communicative frames.

From the information depicted in both figures, we can note how some general words are linked to narco-related concepts by means of figurative projections. For instance, the word *nieve* (snow) is mapped to *polvo* (powder) and then to *cristal* (crystal) metaphorically and metonymically. First, *nieve* is compared to the drug called crystal by profiling its white color. Then, a component (*polvo*) of the whole (crystal) is used to refer to the same drug by profiling its rock-like appearance. In addition, we can distinguish some natural clusters about common communicative topics. For instance, note in Figure 3 how the words *hierba*, *polvo*, *polvito*, and *hielo* share a common frame. This natural grouping could help us identify some recurrent categories for expressing narco-related concepts. On this matter, we summarize the categories we could derive from the similarity analysis in Table 3.

Table 3. Categories for narco-related concepts.

Substances	Violence	Production	Corruption	Persons
perico	cajuela	cortar	lavar	arete
aspirina	carne asada	espolvorear	sembrar	chacuil
borrego	pozolero	escarchar	placa	tacuache
chiva	encobijar	mesero	cachucha	halcón
dulce	suicidar	puchador	madrina	operador
tecata	calentar	aditivo	prostituir	regente

It is worth mentioning that the categories are not balanced. For instance, the category *Persons* hardly contains 14 words, whereas the category *Substances*, as expected, contains more than 40 items.

Finally, to understand the results described so far, below we provide some explanations about the figurative devices found in our analysis.

5.3. Discussion

After manually analyzing the contexts in which the 151 words appeared in our corpus, we noticed that two figurative devices are the most common to refer to narco-related concepts: Metaphor and metonymy.

With respect to the former, the type of metaphor that tends to appear is the conceptual. According to Lakoff and Johnson [14] and Lakoff [55], the conceptual metaphors are part of the cognitive subconscious and are learned and used automatically. In fact, the authors state that metaphorical expressions in language are linked to concepts, in a systematic way,

we can use linguistic expressions to study the nature of metaphorical concepts and to gain a better understanding of the metaphorical nature of our activities. In addition, Panther and Thornburg [56] outline that conceptual metaphors involve the schematization of a concrete source domain to a more abstract target domain, where the target domain is (at least partially) structured by the target domain. Thus, in the conceptual metaphor, PEOPLE ARE ANIMALS, we can group specific realizations such as calling an informer *rata* (rat), or a vigilant *halcón* (falcon). These are a set of systematic conceptual correspondences between two different conceptual domains, where some of the properties of the source domain are transferred to the target domain (Ibarretxe-Antuñano [57]). It is worth clarifying that the source domain (origin) is the conceptual domain from which we extract metaphorical expressions to understand another conceptual domain, while the conceptual domain that is understood in this way is the destination domain (goal) (Kovecses [58]).

From the previous explanation, we can understand some conceptual metaphors in our data. For instance, in the case of the conceptual metaphor, DRUGS ARE JOY, (which is probably subordinate to the general metaphor, JOY IS UP), we can derive examples of the use of words such as *feliz* (happy), *alegría* (joy), or *risa* (laughter) to refer to narco-related concepts. Surely, in some cases, there is a relationship with the experience of consumption; it is known that marijuana produces laughter, and opioids initially produce a feeling of euphoria. Consider the following segments to exemplify this kind of conceptual metaphor: *hierba de la risa* (giggle weed) or *polvo feliz* (happy powder).

In the same vein, the conceptual metaphor, GOOD IS UP, can be observed in our data. *Elevarse* or *ponerse/andar hasta arriba* (to be high) are figurative forms to express that a person who uses drugs finds desirable to feel a high effect; thus, there is a reference to high places: *Les dio un bajón horrible* (they got a slump).

Others types of metaphor found in our data are the so-called experiential and justifying metaphors (Torregrosa and Sánchez-Reyes [44]). Experiential metaphors have to do with perception and the senses, i.e., the projections they take from the source semantic field have to do with the sensations that these drugs produce or with the appearance they have, which is one of the characteristics of primitive metaphors. One of the most frequent types of experiential metaphors are those that originate from the appearance of two related entities, that is, the feature that is projected is the image generated by the sign. In this respect, some names of drugs have originated from the objects they resemble. For instance, *pasto* (grass), *vidrio* (crystal), or *harina* (flake). In these examples, we could assume that the mapping between the source and target domain is originated in a metonymic process, in which the whole (marijuana, crystal, cocaine) is named for one fuzzy part (pasto, vidrio, harina, respectively). The experiential metaphor, on the other hand, is related to the effect that drug users experience, which refers to stimulating, relaxing, or hallucinogenic sensations. Among these metaphors is the well-known relationship between intensity and heat, which, according to Lakoff and Johnson [14], served to support the hypothesis of embodiment. These three effects (stimulants, relaxants or hallucinogens) produce different metaphoric relationships that could be grouped as follows: stimulants refer to intensity and potency, relaxants to low places, and hallucinogens to spaces outside of reality, or inaccessible, like outer space.

In regard to the second figurative device, metonymy, Kovecses [58] states that, unlike metaphors, in metonymies there are not necessarily schematizations between the source and target domain; rather, one gives access to the other, as a reference. In this respect, it is usually understood that between both conceptual domains there is a certain relationship of contiguity (Panther and Thornburg [56]), based mostly on our experiences. Within this category, there are two types of expressions, the ones that use the name of the place only, and those that use the name accompanying another element or the name of the place as a determining factor. For instance, Mexican green (marijuana), California sunshine (LSD), or China cat (heroine).

Finally, another metonymic mechanism that occurs in this context is to refer to the part for the whole. Usually in this type of metonymies, there are examples in which a physical

part of an element is used to name the whole. In our data, we found interesting forms to verbalize such relationship: A body part used for the drug consumption process (*narizaso* (nose) to name the fact of inhaling cocaine), or the instrument used for its consumption (*cucharita* (spoon) to refer to a component for boiling heroine to be further injected).

To conclude this section, it is important to stress out two facts: (i) both metaphor and metonymy, and figurative language itself, are part of our creative faculty of making sense, known as mental imagery, which according to Langacker [4]), is our capacity to structure or construe the content of a domain in alternate ways, and (ii) such devices are not only rhetorical tools, but cognitive instruments to permeate our way of speaking and acting in the world. Thus, they help us to shape and structure several aspects of the conceptual system of any culture (see Nerlich [59]).

6. Conclusions and Future Work

In this article, we approached figurative language in an unusual context: Narco language. Our objective was to identify figurative devices to name narco-related concepts in Spanish. By training a word embedding model with a specialized corpus, we could identify a set of triggers of figurative content. We then formed groups with those triggers for better understanding what figurative devices were used, as well as for explaining how such devices were working. In this respect, as shown in this work, most of the triggers found can be explained in terms of figurative devices such as metaphor and metonymy, which according to the Cognitive Linguistics foundations, are part of our conceptual structure.

As future work, we are planning some experiments to (i) identify specific figurative devices, for instance, irony or metaphor, by applying a set of reported features to assess their applicability in the narco context, and (ii) expand the triggers by combining registered words with their figurative equivalents in the word embeddings. For instance, positive = [*droga* (registered word)] + negative = [*mujer* (figurative equivalent)] → cocaína, heroína mariguana, maleante, envoltorio. Finally, in addition to the scientific contributions presented here, we consider that there could be an application in the field of law enforcement. Since, through this methodology, it would be possible to identify veiled references regarding other areas of crime, such as the use of drugs, illicit activities, or human trafficking.

Author Contributions: All authors equally contributed to the research performed in this paper. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

NLP	Natural Language Processing
FLP	Figurative Language Processing
ML	Machine Learning
NNM	Neural Networks Models
CBOW	Continuous Bag-Of-Words

References

1. de Saussure, F. *Course in General Linguistics*; Fontana: London, UK, 1974; p. 240.
2. Searle, J.R. Literal Meaning. *Erkenntnis* **1978**, *13*, 207–224. [CrossRef]
3. Fillmore, C.; Kay, P.; O'Connor, M. Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language* **1988**, *64*, 501–538. [CrossRef]
4. Langacker, R. *Concept, Image and Symbol. The Cognitive Basis of Grammar*; Mounton de Gruyter: Berlin, Germany, 1990.

5. Kemmer, S. About Cognitive Linguistics: Historical Background. 2010. Available online: http://www.cognitivelinguistics.org/cl.shtml (accessed on 25 August 2011).

6. Goldberg, A. Construction Grammar. In *Concise Encyclopedia of Syntactic Theories*; Brown, E., Miller, J., Eds.; Elsevier Science Limited: Amsterdam, The Netherlands, 1996.

7. Peters, W. Detection and Characterization of Figurative Language Use in WordNet. Ph.D. Thesis, University of Sheffield, Sheffield, UK, 2004.

8. Bergen, B. Mental Simulation in Literal and Figurative Language Understanding. In *The Literal And Nonliteral in Language and Thought*; Coulson, S., Ed.; Peter Lang Publishing: Berlin, Germany, 2005; pp. 255–280.

9. Chomsky, N. *Syntactic Structures*; Mouton and Co: The Hague, The Netherlands, 1957.

10. Katz, A.; Turner, M.; Gibbs, R., Jr.; Cacciari, C. Counterpoint Commentary. In *Figurative Language And Thought*; Marschark, M., Ed.; Oxford University Press: Oxford, UK, 1998; pp. 158–192.

11. Gibbs, R. Evaluating Contemporary Models of Figurative Language Understanding. *Metaphor. Symb.* **2001**, *16*, 317–333. [CrossRef]

12. Lönneker-Rodman, B.; Narayanan, S. *Computational Approaches to Figurative Language*; Cambridge University Press: Cambridge, UK, 2008.

13. Reyes, A.; Rosso, P.; Veale, T. A Multidimensional Approach For Detecting Irony in Twitter. *Lang. Resour. Eval.* **2013**, *47*, 239–268. [CrossRef]

14. Lakoff, G.; Johnson, M. *Metaphors We Live by*; University of Chicago Press: Chicago, IL, USA, 1980; p. 242.

15. Bogdanova, D. A framework for figurative language detection based on sense differentiation. In *Proceedings of the ACL 2010 Student Research Workshop*; Association for Computational Linguistics: Morristown, NJ, USA, 2010; pp. 67–72.

16. Li, L.; Sporleder, C. Using Gaussian Mixture Models to Detect Figurative Language in Context. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*; Association for Computational Linguistics: Los Angeles, CA, USA, 2010; pp. 297–300.

17. Rentoumi, V.; Giannakopoulos, G.; Karkaletsis, V.; Vouros, G. Sentiment Analysis of Figurative Language using a Word Sense Disambiguation Approach. In *Proceedings of the International Conference RANLP-2009*; Association for Computational Linguistics: Borovets, Bulgaria, 2009; pp. 370–375.

18. Sikos, L.; Brown, S.W.; Kim, A.; Michaelis, L.; Palmer, M. Figurative Language: "Meaning" is often more than just a sum of the parts. In *Proceedings of the AAAI 2008 Fall Symposium on Biologically Inspired Cognitive Architectures*; Association for the Advancement of Artificial Intelligence: Washington, DC, USA, 2008; pp. 180–185.

19. Colston, H.; Gibbs, R. A Brief History of Irony. In *Irony in Language and Thought*; Gibbs, R., Colston, H., Eds.; Taylor and Francis Group: New York, NY, USA, 2007; pp. 3–24.

20. Utsumi, A. A unified theory of irony and its computational formalization. In *Proceedings of the 16th Conference on Computational Linguistics*; Association for Computational Linguistics: Morristown, NJ, USA, 1996; pp. 962–967.

21. Veale, T.; Hao, Y. Support Structures for Linguistic Creativity: A Computational Analysis of Creative Irony in Similes. In Proceedings of the CogSci 2009, the 31st Annual Meeting of the Cognitive Science Society, Amsterdam, The Netherlands, 29 July–1 August 2009; pp. 1376–1381.

22. Veale, T.; Hao, Y. Detecting Ironic Intent in Creative Comparisons. In *Proceedings of 19th European Conference on Artificial Intelligence—ECAI 2010*; IOS Press: Amsterdam, The Netherlands, 2010; pp. 765–770.

23. Bosco, C.; Patti, V.; Bolioli, A. Developing Corpora for Sentiment Analysis: The Case of Irony and Senti-TUT. *Intell. Syst. IEEE* **2013**, *28*, 55–63. [CrossRef]

24. Tang, Y.j.; Chen, H.H. Chinese Irony Corpus Construction and Ironic Structure Analysis. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics*; Technical Papers; Dublin City University and Association for Computational Linguistics: Dublin, Ireland, 2014; pp. 1269–1278.

25. Buschmeier, K.; Cimiano, P.; Klinger, R. An Impact Analysis of Features in a Classification Approach to Irony Detection in Product Reviews. In *Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*; Association for Computational Linguistics: Baltimore, MD, USA, 2014; pp. 42–49. [CrossRef]

26. Whalen, J.M.; Pexman, P.M.; Gill, A.J.; Nowson, S. Verbal irony use in personal blogs. *Behav. Inf. Technol.* **2013**, *32*, 560–569. [CrossRef]

27. Fersini, E.; Pozzi, F.; Messina, E. Detecting irony and sarcasm in microblogs: The role of expressive signals and ensemble classifiers. In Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Paris, France, 19–21 October 2015; pp. 1–8.

28. Karoui, J.; Benamara Zitoune, F.; Moriceau, V.; Aussenac-Gilles, N.; Hadrich Belguith, L. Towards a Contextual Pragmatic Model to Detect Irony in Tweets. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*; Association for Computational Linguistics: Beijing, China, 2015; pp. 644–650. [CrossRef]

29. Wallace, B.C.; Choe, D.K.; Charniak, E. Sparse, Contextually Informed Models for Irony Detection: Exploiting User Communities, Entities and Sentiment. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*; Association for Computational Linguistics: Beijing, China, 2015; pp. 1035–1044. [CrossRef]

30. Feldman, A.; Peng, J. An approach to automatic figurative language detection: A pilot study. In Proceedings of the Corpus-Based Approaches for Figurative Language Colloquium, Liverpool, UK, 20–23 July 2009.

31. Veale, T.; Hao, Y. Comprehending and generating apt metaphors: A web-driven, case-based approach to figurative language. In *Proceedings of the 22nd National Conference on Artificial Intelligence*; AAAI Press: Vancouver, BC, Canada, 2007; Volume 2, pp. 1471–1476.

32. Veale, T.; Hao, Y. Learning to Understand Figurative Language: From Similes to Metaphors to Irony. In Proceedings of the Irony CogSci 2007: The 29th Annual Meeting of the Cognitive Science Society Nashville, Nashville, TN, USA, 1–4 August 2007.

33. Saygin, A.P. Processing figurative language in a multi-lingual task: Translation, transfer and metaphor. In Proceedings of the Corpus-Based and Processing Approaches to Figurative Language Workshop, Lancaster, UK, 29 March 2001.

34. Veale, T. Creative language retrieval: A robust hybrid of information retrieval and linguistic creativity. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2011; Volume 1, pp. 278–287.

35. Zinken, J. Discourse metaphors: The link between figurative language and habitual analogies. *Cogn. Linguist.* **2007**, *18*, 445–466. [CrossRef]

36. Pierce, R.; MacLaren, R.; Chiappe, D. The role of working memory in metaphor production and comprehension. *J. Mem. Lang.* **2007**, *56*, 172–188.

37. Peters, W.; Wilks, Y. Data-Driven Detection of Figurative Language Use in Electronic Language Resources. *Metaphor. Symb.* **2003**, *18*, 161–173. [CrossRef]

38. Markert, K.; Nissim, M. Corpus-based Metonymy Analysis. *Metaphor. Symb.* **2003**, *18*, 175–188. [CrossRef]

39. Saldívar, R. Metáforas y metonimias conceptuales en nombres de drogas en inglés y en español. *Forma FunciÓn* **2022**, *35*. [CrossRef]

40. Acosta, D.; Mora, C. Subcultura carcelaria. Diccionario de la jerga canera. *Esc. Penit. Nac.* **2008**.

41. Sanmartín, J. *Lenguaje y Cultura Marginal. El Argot de la Delincuencia*; Universitat de Valencia: Valencia, Spain, 1998.

42. Saldívar, R. *Análisis Lexicológico del Narcolenguaje en Baja California*; UABC: Mexicali, Mexico, 2014.

43. Mattiello, E. An introduction to English slang: A description of its morphology, semantics and sociology. *Polimetrico* **2008**, *2*, 2–10.

44. Torregrosa, G.; Sánchez-Reyes, S. Raising metaphor awareness in English for law enforcement. *Procedia Soc. Behav. Sci.* **2015**, *212*, 304–308. [CrossRef]

45. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*; Curran Associates Inc.: Red Hook, NY, USA, 2013; pp. 3111–3119.

46. Bakarov, A. A Survey of Word Embeddings Evaluation Methods. *arXiv* **2018**, arXiv:1801.09536.

47. Almeida, F.; Xexéo, G. Word Embeddings: A Survey. *arXiv* **2019**, arXiv:1901.09069.

48. Liu, J.; O'Hara, N.; Rubin, A.; Draelos, R.; Rudin, C. Metaphor Detection Using Contextual Word Embeddings From Transformers. In *Proceedings of the Second Workshop on Figurative Language Processing*; Association for Computational Linguistics: Cambridge, MA, USA, 2020; pp. 250–255. [CrossRef]

49. Aggarwal, S.; Singh, R. Metaphor Detection using Deep Contextualized Word Embeddings. *arXiv* **2020**, arXiv:2009.12565.

50. Nozza, D.; Fersini, E.; Messina, E. Unsupervised irony detection: A probabilistic model with word embeddings. In Proceedings of the International Conference on Knowledge Discovery and Information Retrieval, Porto, Portugal, 9–11 November 2016; Volume 2, pp. 68–76.

51. González, J.Á.; Hurtado, L.F.; Pla, F. Transformer based contextualization of pre-trained word embeddings for irony detection in Twitter. *Inf. Process. Manag.* **2020**, *57*, 102262. [CrossRef]

52. Peng, J.; Feldman, A. Automatic idiom recognition with word embeddings. In *Information Management and Big Data*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 17–29.

53. Mikolov, T.; Chen, K.; Corrado, G.S.; Dean, J. Efficient Estimation of Word Representations in Vector Space. In Proceedings of the International Conference on Learning Representations ICLR, Scottsdale, AZ, USA, 2–4 May 2013.

54. Dessì, D.; Recupero, D.R.; Sack, H. An Assessment of Deep Learning Models and Word Embeddings for Toxicity Detection within Online Textual Comments. *Electronics* **2021**, *10*, 779. [CrossRef]

55. Lakoff, G. The neural theory of metaphor. In *The Cambridge Handbook of Metaphor and Thought*; Gibbs, R.W., Jr., Ed.; Cambridge University Press: Cambridge, UK, 2012; pp. 17–38.

56. Panther, K.U.; Thornburg, L.L. The roles of metaphor and metonymy in English -er nominals. In *Metaphor and Metonymy in Comparison and Contrast*; DE GRUYTER: Berlin, Germany; Boston, MA, USA, 2012.

57. Ibarretxe-Antuñano, I. Metáforas de la percepción: Una aproximación desde la lingüística cognitiva. In *Recorridos de la Metáfora: Mente, Espacio y Diálogo*; Yáñez, C.S., Baeza, J.O., Eds.; Cosmigonon: Concepción, Chile, 2011; pp. 141–166.

58. Kovecses, Z. *Metaphor: A Practical Introduction*; Oxford University Press: Cary, NC, USA, 2002.

59. Nerlich, B. Metonymy. In *Encyclopedia of Language and Linguistics*; Brown, K., Ed.; Elsevier: Amsterdam, The Netherlands, 2006; pp. 109–113.