Trends in the Use of Lexical Indigenisms in Mexican Texts: Contributions from the
Corpus de Referencia del Español Actual (CREA)

Abstract

The ways in which indigenous lexical items have been incorporated into the Spanish of America have been investigated by different means: through dictionaries, dialectal surveys and searches in different text types (fiction compared with non-fiction). This article investigates texts to detect the presence of indigenisms and to determine the possible trends by using the *Corpus de Referencia del Español Actual* (*CREA* 2015-0.1). To this end, an empirical research study was designed to complement Lope Blanch's investigation (1979) to recognize how indigenisms appear in the texts and to study the suitability of the *CREA* 2015-0.1 for this kind of research. The results reveal that 13.14% of the indigenisms are used in fiction, 7.05% of the indigenisms are used in non-fiction, 47.76% of the indigenisms are documented in both categories, and the rest do not appear. Our conclusions show the possible contribution of this database in determining the propensity of indigenisms in Mexican texts between 1975 and 2004.

1. Introduction

The trends that are followed by the Amerindian lexicon that appears in fiction and non-fiction texts in the twentieth-century Spanish language have not yet been investigated in depth, according to the interest that lexicographers have in collecting authentic samples of language use in different text typologies. The main research studies have investigated the frequency of this vocabulary in a certain community of speakers and have attempted to verify whether it is considered obsolete and is no longer used, which may have initiated its disappearance (Morínigo 1964, Alvar 1970, Lope Blanch 1979, Haensch 1987, López Morales 2006, Luna Traill 1999, Cáceres Lorenzo 2015). In addition, studies have explored written texts as a possible way to obtain dialectological data, since each type of text provides different information to the researcher. For example, it is possible to find a better reflection of more informal language in fiction than in non-fiction through certain characters or situations, although the local press can also use indigenisms to gain the acceptance of a certain audience. Fajardo Aguirre (1991: 8) managed to count one hundred and twenty-nine Amerindian words in the Argentine fiction texts that he examined. Lévêque (2011: 99) made a similar contribution when he documented sixty-five Amerindian words in his analysis of thirty years of the discourse of the Central American novel. In non-fiction, Prieto (2006: 188) showed that certain presses use a greater number of Amerindian terms, with two hundred and six Quechuan lexical items obtained in his analysis of thirty years of Chilean newspapers. San Martín Núñez (2009: 183) made a similar contribution when he found one hundred and thirty-two indigenous words that manifest a semantic adaptation in his investigation of a Chilean newspaper. This lexicon must be known by the speaking community

because, as explained by Ávila (2003) and Ueda (1996), the media only chooses the most comprehensible terms.

Moreover, in fiction, although the presence of the Amerindian vocabulary can represent spoken and informal language, making regional literature a source of dialectological studies, it does not always guarantee that this vocabulary is used in common speech. Alba (1976) and Béjar and Zamora (1987) remind us in their work that in the discursive strategies of the Cuban poetic movement, Siboney, and in certain literary works of the Dominican Republic, a large number of indigenisms are frequently present, which is a consequence of the ethnographic erudition of their authors. However, Aleza Izquierdo (1995), Enguita Utrilla (1998) and Navarro Carrasco (2000) indicate that an analysis of the vocabulary in literary texts can often be a necessary complement to researchers despite the possible limitations. Lévêque (2011) has called the regional literary discourse a source of data for the dialectological classification of a certain geographical enclave. Hediger (1977: 19), in turn, refers to regional literary discourse as a material that identifies its own particularities, and Perna (2015: 197), in his analysis of the gaucho's literature, insists that the specialist can find a great variety of data in fiction that is socially recognized.

This research is based on the idea that texts reflect part of the use of a vocabulary in different communication situations. The trends of indigenous words that are used in fiction and non-fiction texts as a research problem have not been examined with the databases that are currently available. Accordingly, our research proposes the analysis of a glossary of text-based indigenisms as a complement to what we know as the Amerindian vocabulary in the different language varieties of American Spanish. A researcher of the lexical loans of the Spanish language cannot ignore the different materials and their limitations when concluding what indicates a tendency in the knowledge and use of the vocabulary. In the existing bibliography, it is indicated that all individuals register an active or productive vocabulary and a passive or receptive vocabulary. These terms are not contradictory but complementary and can vary from one generation to the next. This tendency is related to the ability of certain lexical units to move from productive use to receptive knowledge. In this line of research, Haensch (1987: 562) realized in a field study that was conducted in 1982 for the Augsburg Project that the Americanisms that were compiled from 1925-1975 in Colombia had experienced changes so significant that 25-30% of the previously registered entries did not appear in their compilation, and 15-20% were used restrictively in a non-urban context and by the generation of people over 50 years of age. This conclusion coincides with Rosenblat in an analysis of the Venezuelan lexicon as well as the research of Lope Blanch in Mexico, López Morales in Cuba, Vaquero in Puerto Rico, Alba in the Dominican Republic and Montes Giraldo in Colombia.

In the specific case of the indigenisms in Hispanic America, researchers have often obtained information from dialectological dictionaries in previous years. Since the 1960s, many scholars have expressed their conviction that the inventory

of words that appears in a dialectological dictionary is not always representative of these words' actual use among speakers. This contention explains the importance of accessing other sources, as Lara (1990: 67) explains, to evaluate the linguistic reality of the Amerindian loan words in the Spanish language. That is, these dictionaries present biased information, because the lexicographers repeat, for reasons of erudition, dying terms that nobody knows or that people only know passively, as Morínigo (1964: 218) indicates. Fajardo Aguirre (2010: 318) adds that the dialectal lexicographers, with a certain frequency, care much more about following previous works that they consider highly prestigious and do not wish to contradict. This emphasis distorts the results of the research that is conducted with dictionaries. *Diccionario de Americanismos* [The Dictionary of Americanisms] (2010), which was published in the 21st century, attempts to overcome these limitations with Pan-Hispanic information.

As a way to obtain data on lexical indigenisms, surveys also present certain limitations. For example, according to Alvar (1960: 59), the surveys to determine the use of the Amerindian words that were used to produce the *Atlas lingüístico de Hispanoamérica* [*Linguistic Atlas of Latin America*] were obtained from rural informants through the question-and-answer mechanism. This technique produced a single specific response so that possible synonyms, emotional terms, etc., were lost. Thus, surveys of this type provide conclusions that are similar or complementary to the conclusions that are recorded by dictionaries (García Moutón 2015: 76). Surveys have also been used to conduct oral recordings with spontaneous dialogues, either directed or free, in formal situations, with other types of subjects (urban men and women of different generations and based on other sociocultural factors) in an attempt to overcome the predilection of some previous scholars for a type of rural informant that Chamber and Trudgill (1980) referred to as non-mobile, elderly, rural, and male (NORM). In this line of inquiry, Lope Blanch (1979: 98) presents a coordinated study of the linguistic norm of Spanish that is spoken in the main cities of the Hispanic world; this research is based on the idea of conducting Pan-Hispanic research through surveys that use sociolinguistic criteria in the cities where Spanish is spoken as an official or co-official language. In later years, different studies have appeared that always focused on urban subjects. In contrast to the traditional dialectology, which was influenced by the neogrammatic theory that prioritized uncultivated speech, the sociolinguistic programme of Lope Blanch assumed that only the features that provide diastratic information configure the main norms of the language of the current speakers who live mostly in cities. This perspective has been analysed by Rabanales (2004: 80) who explains the limitations that are evident in the absence of some autochthonous voices that designate rural realities.

In addition, Lope Blanch (1979) and Sala *et al.* (1977) propose using the concept of lexical vitality to explain the frequency of word use with which they indicate that if a word has a great geographic diffusion, it can create derivatives and new

meanings and is a word with significant vitality. Lara (1990: 84) also describes a corpus of references and presents a frequency index for the entire Mexican Republic in a stage that was previous to the work of the Spanish Royal Academy that is known as the *Corpus de Referencia del Español Actual* [*Corpus of Reference of the Present Spanish*] (*CREA*), which collects texts of different origins from 1975-2004.

Finally, obtaining the results regarding the use of an indigenous vocabulary by searching different text types is also presented as a novel option that is associated with empirical research with a corpus. This proposal is not really new because the first academic dictionary, the *Diccionario de Autoridades* [*Dictionary of Authorities*] (1726-1739), used texts as support to explain a given word (Álvarez de Miranda 2000: 88). In recent years, different linguistic and digital corpuses have been created and are databases that are available to vocabulary scholars, who are asked to radically change the working methods, as explained by Rojo (2016: 198). In this line of inquiry, in 2015, the Royal Spanish Academy published a renewed version of the *CREA* that has become an important source of examples of real texts, both fiction and non-fiction, despite the limitations that were shown by Molina Salinas and Sierra Martínez (2015).

The general objective of this work is to analyse the state of the terms that were taken from indigenous languages and have been used in the texts that were produced in Mexico during a specific chronological period that complements other investigations. This purpose relates to the fact that the suitability of the *CREA* for this type of research is also being evaluated. We refer to the monograph by Lope Blanch (1979: 35, 37), who analysed the Amerindian vocabulary that was obtained in 1969 and revised his work ten years later in his publication *Léxico indígena en el español de México* [*Indigenous Lexicon in the Spanish of Mexico*] (*LIEM*). This monograph provides a scale that describes the knowledge of this vocabulary in one hundred interviewed subjects and allows us to design an investigation with the parameters of the fiction and non-fiction texts from Mexico; it compares its results with the documentary information of the new *CREA* 2015 (0.1). This article does not intend to inspect what Lope Blanch published, although we know that Alvar (1970: 324) questions the registration of some terms as scarcely used in the first edition of 1969, although this problem was solved in the 1979 version. Our research considers that the *LIEM* provides a valid result according to its own design, and we use it as our initial search material because we attempted to add data through *CREA* 0.1 for a given period to obtain an overview of the problem of researching the frequency of lexical indigenisms in Mexico.

The procedure that is followed by Lope Blanch begins with the development of a list that includes indigenisms that come from oral speech (26.28%), from written texts (21.15%), from speech and writing (36.53%) and from personal additions of the author (16.02%). With this list, he also prepares a dialectal questionnaire. These indigenisms are Nahua words, with the exception of the Mayan terms (*canán, cenote, chilango, henequén, ixtabetún, maquech, papadzul, pibil* and *salbute*) and those from

Tarasco (*cuacha, charal, huarache, sambache, tambache* and *uchepo*), Otomi (*naco*), Zapotec (*guelatza*), and Cahita (*guare*). Further research on Nahua loan words has corroborated these results in Mexican Spanish and in other countries in the Americas (Moreno de Alba 1992, Luna Traill 1999 and Cáceres-Lorenzo 2015). The Nahua voices that were collected by Lope Blanch include examples of words that were formed by derivation, which is an indicator that these words are heavily used, as explained by Lope Blanch (1979) and Sala *et al.* (1977: 144) in their analysis of the American vocabulary. Lope Blanch himself (1979: 54) also stated that written language shows a much greater capacity than spoken language to form neologisms through the use of suffixes compared with the use of oral discourse, which prefers the periphrastic procedures of spoken language. The next step in the methodology that was designed by Lope Blanch envisages an interview with one hundred individuals with different diastratic indicators in Mexico City who were asked about their knowledge of three hundred and twelve indigenisms.

The result is the classification of six groups of indigenous terms that are characterized by the passive vocabulary (which is known but little used) that each individual claim to know to different degrees, which explains the presentation of the classification with the groups that are mentioned above as a scale of knowledge. A lexical unit can be a part of a passive vocabulary because it is outdated or because it has a restricted use in certain specific or ethnographic languages that appear in the texts. However, it can also indicate that a certain word has embarked on a process of lexical mortality. As indicated by Alvar (1960: 62), this situation can be modified for stylistic reasons, which is what López Morales (2006: 907) called "lexical resurrection", where the term can acquire new meanings years after a researcher indicates the meanings that are already used.

Considering the above discussion, we propose the following research questions. Is the *CREA* suitable to record the data on the number of Amerindian voices in each type of text? What derivatives are recorded in the *CREA* texts as indicators of lexical vitality? Finally, is it possible to detect some type of tendency in the use of lexical indigenisms? The working hypothesis that we propose as a provisional answer to these questions is that it is feasible to recognize the lexical tendency that is used in the written texts through the investigation of indigenisms in written sources in a resource such as the *CREA* 0.1 (2015), which we consider appropriate for obtaining information regarding actual use.

We hope that this work makes an empirical contribution to the knowledge of indigenisms in Mexico for the historical period from 1975-2004 based on the textual support of the new annotated version of the *CREA*. The sum of the data that were obtained from dictionaries, surveys and textual typology can yield more reliable information on a polyhedral reality such as the use of words.

2. Materials and method

To answer the questions that are raised, an investigation was designed that begins with the analysis of the different terms that Lope Blanch obtained in his work on indigenisms in Mexico - *LIEM*. A total of three hundred and twelve terms, classified into six unequal groups, are presented in Appendix. The description of the entries in each group, according to Lope Blanch, is as follows:

> I. Absolutely known in a general way (99-100%);
> II. Generally known but with some uncertainties (85-98%);
> III. Average recognition (50-85%);
> IV. Little known or has an imprecise meaning for most informants (25-50%);
> V. Virtually unknown (2-25%); and
> VI. Completely unknown, at least to the inhabitants of Mexico City (0-1%).

The *CREA* 2015-0.1 offers statistical data on the texts that were written in the period from 1975-2004. This linguistic and digital corpus allows searching for terms and provides statistical information on the number of documents in which the terms appear and their absolute frequency, that is, the number of times that each analysed word appears in this corpus; it also distinguishes between fiction and non-fiction texts for each term. In the search for the pattern that is followed by these indigenisms, the different meanings have not been considered; Lope Blanch also did not consider them, since these meanings are possible to find in all the groups' voices that designate historical concepts, elements of popular culture, plants, animals, professions, etc. (Buesa Oliver and Enguita Utrilla 1992).

We use this information to calculate two new indexes that provide data on the use of these words. These two indexes have been developed to overcome the possible limitations of the *CREA* according to the research of Molina Salinas and Sierra Martínez (2015). To this end, we calculate the percentage of the texts in which indigenisms of each type appear by dividing the number of documents in which the analysed term appears by the total number of documents in the database (81 fiction and 5,558 non-fiction). In addition, it is possible to obtain the relative frequency of the occurrence of these terms by first dividing the absolute frequency of each word by the total number of words in the database (3,803,442 in the fiction section and 8,514,877 in the non-fiction section), calculated by one million words.

The two indexes explain the frequency of the use of terms, but neither index is more important than the other. The first index shows us the most used terms in the documents, and the second index shows the number of times these words are repeated. We believe that an analysis is more feasible with these indexes because there could be terms that are infrequently used but, in many documents, and other terms that are repeated often but only in a few texts. To unify the two aspects, a unified index (Iu) is calculated by multiplying the previous two indexes. The Iu calculates the unified use rates for the three hundred and twelve analysed terms and distinguishes their use by the type of fiction and non-fiction texts.

In our procedure, the Iu is used to develop an order for each indigenism. This order is different for each type of text, as shown in Appendix. In this way, the analysed terms are grouped in the following six new groups: words that are only present in one type of text (a), either fiction (group a1) or non-fiction (group a2); entries with a presence in both types of text (b), with a similar importance in both texts (group b1), and a greater presence in fiction (group b1) or non-fiction (group b2) texts; and finally, terms that are not found in the *CREA* (group c).

3. Results and discussion

Appendix presents the quantitative analysis that was conducted with the three hundred and twelve indigenisms that were proposed by Lope Blanch, presented in alphabetical order. In this Appendix, the position of each word in each of the typologies is specified. For example, *pulque* holds position 14 in fiction and has a greater use in this textual type than in non-fiction, where it occupies position 20. Some words have very close positions with respect to each typology, such as *pulquería*, with positions 49 and 50, which indicates that such words are indigenisms that are widely used in the Mexican texts of the *CREA*. However, the *pulcazo* and *pulquería* derivatives hold the position of the most used voices within this group. Something different occurs with *ixtle*, which occurs at position 77 in fiction and 33 in non-fiction, and it indicates a greater use than *ixtlero*, which only appears in non-fiction at the 102nd position.

Our results partially confirm the information on the voices that were registered in the *LIEM* (Lope Blanch 1979) as almost exclusive to the spoken language. For example, *guacamole* (position 148 in fiction and 94 in non-fiction) and *pepenar* (position 78 in fiction and 118 in non-fiction) are voices in written texts, although their position is less significant, despite the proposal by our researcher. In contrast, *cacle* and *pagua* ratify Lope Blanch's assertion (1979) that they are seldom used. The data of each group that were obtained from the *CREA* 2015-0.1 are shown in Table 1.

Classification		Nº indigenous	%	Total %
a. Exclusively in one type	**a.1. Fiction**	41	13.14%	13.14%
	a.2. Non-fiction	22	7.05%	7.05%
b. Appears in both types	**b.1. Similar presence**	80	25.64%	
	b.2. More often in fiction	26	8.33%	
	b.3. More often in non-fiction	43	13.78%	47.76%
c. Not documented		100	32.05%	32. 05%

Table 1. Classification of indigenisms in the *CREA*

An analysis of Table 1 indicates that we have been able to register 67.95% of the vocabulary that we are investigating, although we also show many undocumented voices (32.05%). These undocumented voices may indicate the partial suitability of the *CREA* to obtain data on words that have already disappeared from written texts (fiction and non-fiction). By contrast, in literary fiction, thirty-two instances of the indigenous vocabulary of Lope Blanch's first three groups are recorded against the greater number that appears in non-fiction. This question seems to coincide with the conclusions that were provided by Ueda (1996: 93) regarding the vocabulary of television in which the news uses a smaller number of indigenous voices compared with other types of programming. In this line, Ávila (2003: 13) determines a tendency in non-fiction media to use the most comprehensible vocabulary for all sociocultural groups, which affirms that many indigenous words do not fulfil that communicative function.
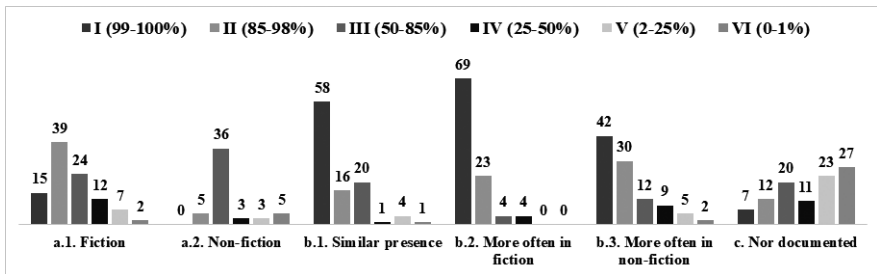
Figure 1. Results of the documentary support

The first three groups of greater knowledge have a higher frequency of use (between 50 to 100%); the words in these groups are recorded in the typologies (b.1, b.2, b.3), although they are also recorded in a.1 and a.2, as indicated by the percentages in Figure 1. That is, it seems that these groups are very well represented in the *CREA* texts. What is found is that in general, group b shows a higher frequency of use in Mexico. In fact, it is the most numerous group and, according to our research, presents three possibilities, which are reflected in Table 1 and the Figure and Appendix of this paper. The first group consists of Amerindian voices that have a similar presence in the two typologies (b.1). With eighty words, the first group is the repertoire with greater frequency. The lowest numbers are the following: *cuate* (ranked 1 in fiction and 17 in non-fiction), *chocolate* (ranked 2 in fiction and 6 in non-fiction), *tequila* (ranked 3 in fiction and 4 in non-fiction), *chile, mole, tamal, petate, chicle, guajolote, nopal, pulque, coyote, atole, chihuahua, huarache, nahua, milpa cacahuate, hule* and *aguacate*. Some of these examples are Pan-Hispanic voices that refer to natural products and animals, which are used in a general way together with other more specific words, as evidenced by the investigations of

Buesa Oliver and Enguita Utrilla (1992), Luna Traill (1999) and Cáceres-Lorenzo (2015). The fact that not all the words appear indicates the aforementioned partial suitability of the *CREA* to obtain the data on words' registration in written texts. We obtain other information from the analysis of *charal* (105 in fiction and 118 in non-fiction), *chiche* (111 in fiction and 135 in non-fiction), *mixiote* (122 in fiction and 117 in non-fiction), *chicloso, toloache, chimal, tocayo, huisachal, chichicuilote, molote, piocha, tatemar, tepache, malinchista, topil, jicote, pilmama, teocalli, tequesquite, totol* and *zacatón*, which Moreno de Alba (1992) indicates are words that are used only in Mexico and sometimes in Central America.

The next group consists of the voices that are registered preferentially in Mexican fiction texts (b.2). These words, as seen from the analysis in Appendix, belong to Lope Blanch's first two groups, that is, they are widely used terms according to their appearance in the texts. More than 90% of the terms in group b.2 are included in Lope Blanch's categories I and II (1979), which is not the case for the third group (b.3), which includes the Amerindian voices of common use in non-fiction. Our data confirm that seven words, namely, *amate, chaquiste, coyol, huehuenche, peyote, pizote* and *tecali*, come from the groups that are less known by the respondents and that almost 20% of group b.3 provides voices that were integrated into classes IV, V and VI.

Finally, there is the group that is formed by the one hundred terms that we have not been able to document in the *CREA*, which represent voices in the process of obsolescence or that have not yet been documented.

Regarding the seventy-five derivative voices that indicate the possible vitality of the word (Lope Blanch, Sala *et al.*), our inquiry shows different behaviours with respect to the primitive voice, as shown in Appendix, where the greater number of derivatives are c, the terms that are not found in the *CREA*, with 38%, and a.1, the terms that are exclusive to fiction, with 34.15%. The groups with the fewest derivatives are b.1 (8.75%) and b.2 (3.85%). The remaining two are groups b.3 (20.93%) and a.2 (27.27%).

A specific mention is the cases of *coyotaje* and *coyotera* in a.1, with a very low register (position 166) with respect to *coyot*e in b.1, with position 15 in fiction and 30 in non-fiction. Something different occurs with *tamal* (position 9 in fiction and 14 in non-fiction) and *huarache* (position 20 in fiction and 22 in non-fiction), which are recorded in group b.1, between the lexical units of greater knowledge and their derivatives with a very low position, namely, *nacatamal* and *quilotama* and *huaracheo* in group c. These words which are used in American Spanish may perhaps be accompanied by an obsolescent type that has stopped being used. In addition, *chicle* is recorded in group b.1, and *chiclero* is recorded in b.3 with different frequency indexes. This result coincides with the findings that were provided by Sala *et al.* (1977: 135) when affirming that a primitive word and its derivative do not always have the same frequency index.

4. Conclusions

This paper presents an empirical analysis with *CREA* data that extends the research that we have taken as preliminary material. These results are one more piece of the information that students must know and that complement what was exposed by Lope Blanch in *LIEM* (1979). The answers to the research questions that guided our study are as follows.

Is the *CREA* suitable to record the data on the number of Amerindian voices in each type of text? The numerical results indicate that the most well-known Amerindian voices are recorded in both fiction and non-fiction. These data are not synonymous with the fact that these words have a common use in everyday communication, but it is possible to determine that these words are part of the vocabulary that is known by readers, the recipients of the texts. The differences among the typologies open new avenues of investigation because it seems to be feasible that some words are specialized to fiction and other words are specialized to non-fiction. If this were true, we would have an indicator that could identify the lexical units of the vocabulary. Simultaneously, the fact that 32% of the words that were collected by Lope Blanch do not appear in the *CREA* can have several interpretations: it may be an indicator of lexical mortality or representative of certain limitations of this corpus that must be resolved in future research.

What derivatives are recorded in the *CREA* texts as an indicator of lexical vitality? In the specialized bibliography, derivatives have been used as indicators of a word's use, but our investigation shows the different degrees. Thus, we propose the need to analyse the derivatives from the perspective of use in different materials to differentiate what occurs with the derivative with respect to its primitive word.

Is it possible to detect some type of tendency in the use of lexical indigenisms?

A trend that our work reveals has allowed us to establish a different typology to Lope Blanch's typology, which complements its results with the *CREA*. Simultaneously, a limitation is detected that comes from the same investigation of Lope Blanch, who conducts his research in the city, which was a novelty at the time but may perhaps restrict the real results to a certain chronological period, as if the speakers of the capital of Mexico were the only norm of this country. In addition, the difference in the positioning of our indigenisms confirms the need to track different empirical material to determine the vocabulary that is used by speakers in terms of their active and passive vocabulary.

We conclude the need to investigate the vocabularies not only with Mexican texts but also with the texts of other countries of the Mesoamerican and Central American area to verify if they are really terms that are used as active or passive vocabulary in the dates that we analyse or in later texts. The existence of other corpuses is an invitation to continue the search in different years up to the present. The construction of a textual and diachronic map of each word is possible thanks to the databases.

Universidad de Las Palmas de Gran Canaria   María-Teresa Cáceres-Lorenzo

References

Alba, Orlando 1976. Indigenismos en el español hablado en Santiago. *Anuario de Letras* 14: 71-99.

Aleza Izquierdo, Milagros 1995. La reivindicación de una cultura: la obra de José M. Arguedas. *Quaderns de Filologia. Estudis literaris.* 1.1: 25-104.

Alvar, Manuel 1960. Los dialectalismos en la poesía española del siglo XX. *Revista de Filología Española* 43: 57-59.

Alvar, Manuel 1970. *El léxico indígena en el español de Méjico. Revista de Filología Española* 53: 323-325.

Álvarez de Miranda, Pedro 2000. Palabras y acepciones fantasmas en los diccionarios de la Academia. *La fabrique des mots. La néologie ibérique*. Ed. J. C. Chevalier. Paris: Presses de l'Université de Paris-Sorbonne. 87-106.

Ávila, Raúl 2003. La lengua española y sus variantes en los medios de comunicación masiva. *Pautas y pistas en el estudio del léxico hispano (americano)*. Ed. G. Wotjak. Madrid: Iberoamericana Vervuert. 11-26.

Béjar, Eduardo C. & Juan C. Zamora 1987. El léxico indígena en la poesía siboneísta: signo, símbolo, sentido. *Actas del I Congreso Internacional sobre el español de América.* Eds. H. López Morales & M. María Vaquero. Madrid: Academia Puertorriqueña de la Lengua Española. 935-945.

Buesa Oliver, Tomás & José María Enguita Utrilla 1992. *Léxico del Español de América. Su elemento patrimonial e indígena*. Madrid: Fundación MAPFRE.

Cáceres-Lorenzo, María-Teresa 2015. Diffusion trends and Nahuatlisms of American Spanish: Evidence from dialectal vocabularies. *Dialectología y Geolingüística* 23: 50-67.

Chambers, Jack & Peter Trudgill 1980. *Dialectology*. London: Cambridge University Press.

Corpus de referencia del español actual. (*CREA* annotated version) 2015. September 1, 2017. Real Academia Española.

Diccionario de Americanismos 2010. September 1, 2017. Asociación de Academias de la Lengua Española.

Enguita Utrilla, José 1998. Americanismos léxicos y textos literarios. *La Torre: Revista de la Universidad de Puerto Rico* 3: 381-397.

Fajardo Aguirre, Alejandro 1991. *Americanismos léxicos en la narrativa argentina contemporánea*. Madrid: Universidad Complutense.

Fajardo Aguirre, Alejandro 2010. La lexicografía del español de América. Evolución, tipología y metodologías. *La lengua española en América: normas y usos actuales*. Eds. M. Aleza Izquierdo & J. Enguita Utrilla. Valencia: Universitat de Valencia. 317-354.

García Moutón, Pilar 2015. Dialectología y geografía lingüística. *Enciclopedia de Lingüística Hispánica*. Ed. J. Gutiérrez Rexach. London: Routledge. 30-40.

Haensch, Günther 1987. La lexicografía hispanoamericana entre la teoría y la práctica. *Actas del I Congreso Internacional sobre el español de América.* Eds. H. López Morales & María Vaquero. Madrid: Academia Puertorriqueña de la Lengua Española. 555-579.

Hediger Helga 1977. *Particularidades léxicas en la novela hispanoamericana contemporánea*. Frankfurt: Peter Lang.

Lara, Luis Fernando 1990. *Dimensiones de la lexicografía: a propósito del Diccionario del español de México*. Mexico: Colegio de México.

Lévêque, Daniel 2011. *Hablantes y hablas en la novelística social centroamericana (1940-1970). Para una tipificación dialectológica del discurso literario regional*. Germany: LINCOM Linguistic.

Lope Blanch, Juan M. 1979. *Léxico indígena en el español de México.* Mexico: Colegio de México.

López Morales, Humberto 2006. La obsolescencia léxica. A propósito del Diccionario usual del español de Puerto Rico. *Corcillvm: Estudios de traducción, lingüística y filología dedicados a Valentín García Yebra*. Ed. R. Rosario Gonzalo. Salamanca: Universidad de Salamanca. 907-916.

Luna Traill, Elizabeth 1999. Nahuatlismos en el léxico culto de seis ciudades hispánicas. *El Centro de Lingüística Hispánica y la lengua española*. México: UNAM. 395-418.

Molina Salinas, Claudio & Gerardo Sierra Martínez 2015. Hacia una normalización de la frecuencia de los corpus CREA y CORDE. *Revista Signos. Estudios de Lingüística* 48. 89: 307-331.

Moreno de Alba, José 1992. *Diferencias léxicas entre España y América*. Madrid: MAPFRE.

Morínigo, Marcos 1964. La penetración de los indigenismos americanos en el español. *Presente y futuro de la lengua española*. Madrid: Cultura Hispánica. 217-226.

Navarro Carrasco, Ana 2000. *La Academia y los americanismos de La tía Julia y el escribidor*. Alicante: Publications Service of the University of Alicante.

Perna, Carlos Gabriel 2015. Fuentes históricas para el estudio de la lengua en la pampa en el siglo XIX (I): El Martín Fierro y la tradición de la gauchesca. *Signo y seña* 27: 197-222.

Prieto Vera, Luis 2006. Quechuismos en el léxico de la prensa de Santiago de Chile. *Boletín de Filología* 41: 97-196.

Rabanales, Ambrosio 2004. La encuesta como técnica dialectológica. *Onomázein: Revista de lingüística, filología y traducción de la Pontificia Universidad Católica de Chile* 9: 75-93.

Rojo, Guillermo 2016. Citius, maius, melius: del CREA al CORPES XXI. *Lingüística de corpus y lingüística histórica iberorrománica*. Ed. J. Kabatek. Berlin: de Gruyter. 197-212.

Sala, Marius, Dan Munteanu, Valeria Neagu & Tudora Sandru-Olteanu 1977. *El léxico indígena americano. Apreciaciones sobre su vitalidad*. Mexico-Bucarest: Academia Mexicana-Editura Academiei Române.

San Martín Núñez, Abelardo 2009. Influencia de lenguas indígenas en el registro festivo del diario chileno La Cuarta. *Boletín de Filología* 44. 2: 135-185.

Ueda, Hiroto 1996. Análisis demolingüístico del léxico variable español. *Lingüística Hispánica* 19: 63-98

Appendix
F: position in fiction texts; **Iu (F)**: unified index for fiction texts; **NF**: position in non-fiction texts; **Iu (NF)**: unified index in non-fiction texts; **G**: our classification, **a1:** Fiction: **a2:** non-fiction; **b1:** similar presence in both kinds of text; **b2:** better position in fiction; **b3:** better position in nonfiction; **c:** not documented; **LP**: classification of indigenisms according to Lope Blanch (1979): **I:** Absolutely generally known (99-100%); **II:** generally known (85-98%); **III:** medium recognition (50-85%); **IV:** scarcely known (25-50%); **V:** practically unknown (2-25%); and **VI:** completely unknown (0-1%). Other classifications, according to Lope Blanch: **C**: located in spoken and written form; **D**: derivative; **E**: recorded in written text; **H**: located in the spoken language; **ad**: additions (words not documented in recordings or texts).

Indigenisms	F	Iu (F)	NF	Iu (NF)	G	LP	Indigenisms	F	Iu (F)	NF	Iu (NF)	G	LP
achahuisclarse D, ad	191	0	172	0	c	IV	chapulín H	97	7,8	46	0,22	b3	I
achichinar H	166	0,32	172	0	a1	IV	chaquiste E	166	0,3	139	0	b3	V
achichincle H	61	23,37	172	0	a1	II	charal C	105	6,8	118	0,01	b1	II
achiote H	191	0	82	0,05	a2	V	chayote C	130	2,9	37	0,44	b3	I
acocil H	166	0,32	172	0	a1	IV	chayotera D, E	166	0,32	172	0	a1	II
acocote E	191	0	172	0	c	V	chía C	94	8,1	118	0,01	b1	II
aguacate C	27	113,6	12	2,2	b1	I	chiche H	111	5,8	135	0	b1	I
aguate E	191	0	172	0	c	V	chichicascle H	191	0	172	0	c	IV
ahuaucle H	191	0	172	0	c	V	chichicuilote H	130	2,9	139	0	b1	II
ahuehuete C	44	54,5	90	0,04	b2	II	chichile D, H	191	0	172	0	c	VI
ahuizotear E	191	0	172	0	c	VI	chicle C	11	383	19	1,26	b1	I
ahulado D E	166	0,3	118	0,01	b3	II	chiclero C, D	162	0,9	64	0,1	b3	I
ajolote C	94	8,1	42	0,3	b3	II	chicloso D, H	122	3,97	139	0	b1	I
amate ad	115	5,1	47	0,21	b3	V	chicozapote D	166	0,3	101	0,03	b3	I
amole E	191	0	172	0	c	V	chihuahua C	19	194,7	2	99,19	b1	I
apapachar C	115	5,1	90	0,04	b3	I	chilacayote C	61	23,3	107	0,02	b2	II
apipizca ad	166	0,32	172	0	a1	II	chilango ad	122	3,9	48	0,19	b3	II
atemole D, H	191	0	172	0	c	VI	chilaquile H	74	14,3	118	0,01	b2	II
atole C	17	262,9	23	0,9	b1	I	chile C	4	1328,2	1	455,3	b1	I
áxcale E	191	0	172	0	c	V	chilmole D, H	191	0	139	0,002	a2	IV
ayacahuite E	191	0	172	0	c	V	chilpachole C	191	0	172	0	c	III

Entry							Entry						
ayate C	122	3,9	61	0,1	b3	II	chilpayate H	35	82,12	172	0	a1	II
biznaga ad	166	0,3	139	0	b3	II	chiltepín ad	191	0	172	0	c	IV
cacahuacincle ad	191	0	172	0	c	II	chimal ad	128	3,2	109	0,02	b1	V
cacahuate C	24	145,4	25	0,9	b1	I	chimolero D, E	191	0	172	0	c	III
cacao C	40	62,3	10	3,3	b3	I	chinaco C	79	11,6	139	0	b2	III
cacascle ad	191	0	172	0	c	V	chinampa C	71	15,5	41	0,32	b3	II
cacle H	191	0	172	0	c	III	chinchayote D	191	0	172	0	c	III
cacomiscle H	191	0	172	0	c	III	chipil ad	191	0	172	0	c	II
cajete H	105	6,5	116	0,01	b1	III	chipote ad	105	6,4	139	0	b2	I
camichín E	191	0	172	0	c	VI	chipotle D, H	148	1,3	58	0,11	b3	I
camote C	39	64,2	21	1,12	b1	I	chiquihuite C	82	11,36	172	0	a1	II
canán E	191	0	172	0	c	VI	chocolate C	2	1694,3	6	5,71	b1	I
capulín C	64	22,7	70	0,08	b1	I	chocolatería D, H	191	0	172	0	c	I
capulina D, ad	141	1,9	95	0,03	b3	II	chocolatero D, E	191	0	172	0	c	I
cempasúchil C	162	0,9	107	0,02	b3	I	chomite E	191	0	172	0	c	VI
cenote C	111	5,8	18	1,49	b3	III	chuchuluco H	191	0	139	0,002	a2	III
cenzontle C	91	9	139	0	b2	II	cocol H	148	1,3	118	0,01	b3	I
chachalaca C	148	1,3	172	0	a1	III	cocolazo D, H	166	0,32	172	0	a1	II
chacualear ad	191	0	172	0	c	III	coconete ad	191	0	172	0	c	IV
chahuiscle ad	191	0	172	0	c	IV	colote H	166	0,32	172	0	a1	IV
chalchicuil H	191	0	172	0	c	VI	comal C	22	160	49	0,19	b2	I
chamaco C	6	616,7	36	0,47	b2	I	copal C	47	45,4	39	0,36	b1	III
chamagoso H	141	1,95	172	0	a1	II	coyol H	166	0,3	102	0,03	b3	IV
chapopote C	61	23,3	64	0,1	b1	I	coyotaje D, E	166	0,32	172	0	a1	II

Vocablo	F	Iu (F)	NF	Iu (NF)	G	LP
coyote C	15	277,5	30	0,68	b1	I
coyotera D, H	166	0,32	172	0	a1	III
cuacha E	191	0	172	0	c	V
cuatachismo D, E	191	0	172	0	c	II
cuate C	1	2160,4	17	1,54	b1	I
cuescomate H	191	0	172	0	c	V
cuico ad	97	7,79	172	0	a1	II
cuija H	191	0	172	0	c	III
cuitla H	191	0	172	0	c	VI
ejote C	111	5,8	79	0,06	b3	I
elote C	32	93,5	33	0,5	b1	I
enchapopotar D, E	191	0	172	0	c	I
enchilada C, D	54	33,8	86	0,05	b2	I
enchilarse C, D	191	0	172	0	c	I
enjitomatar D, E	191	0	172	0	c	II
epazote C	36	75,9	28	0,76	b1	I
equipal ad	49	44,1	118	0,01	b2	IV
escuincle C	5	631,01	139	0	b2	I
guacamole H	148	1,3	94	0,04	b3	I
guachinango C	191	0	172	0	c	I
guaje C	45	46,7	77	0,06	b2	I

Vocablo	F	Iu (F)	NF	Iu (NF)	G	LP
jilotillo D, H	191	0	172	0	c	V
jiote C	148	1,3	172	0	a1	II
jiotoso D, H	191	0	172	0	c	II
jitomate C, D	43	55,5	4	8,35	b3	I
jocoque E	59	24,99	172	0	a1	II
jocote E	191	0	139	0,002	a2	V
juil E	191	0	139	0,002	a2	V
macegual ad	84	10,4	61	0,1	b1	V
machincuepa E	166	0,32	172	0	a1	III
machote ad	191	0	139	0,002	a2	III
malacate C	191	0	53	0,14	a2	III
malinchismo H	191	0	109	0,02	a2	III
malinchista C, D	141	1,95	118	0,01	b1	III
mapache E	115	5,19	86	0,05	b3	II
maquech H	146	1,62	172	0	a1	IV
matatena ad	122	3,9	172	0	a1	I
mayate C	97	7,8	102	0,03	b1	II
mecapal E	115	5,19	70	0,08	b3	III
mecapalero D, E	120	4,87	172	0	a1	III
mecatal D, H	191	0	172	0	c	IV
mecate C	31	109,06	40	0,34	b1	I

V	c	0	172	0	191	meclapil H	I	b1	0,85	26	314,5	12	guajolote C
VI	c	0	172	0	191	mecuate H	VI	c	0	172	0	191	guare H
III	c	0	172	0	191	memela H	IV	a2	0,002	135	0	191	guelaguetza H
I	b2	0,07	73	118,1	26	metate C	III	a1	0	172	2,6	139	güila ad
I	b2	0,03	95	545,3	8	mezcal C	I	b3	2,18	13	7,8	97	henequén C
II	a1	0	172	4,87	120	mezcalero D, E	III	c	0	172	0	191	henequenera D, E
V	a2	0,002	139	0	191	mezcalina D, E	II	c	0	172	0	191	henequenero D, H
III	b3	0,12	57	0,32	166	mezquital D, E	I	b2	0,03	95	85,7	34	huacal C
II	b3	0,06	80	86,3	33	mezquite C	III	c	0	172	0	191	huamúchil C
VI	a1	0	172	0,65	164	michi H	I	b3	0,11	59	3,25	128	huapango C
I	b1	4,82	8	154,51	23	milpa C	I	b1	0,99	22	185,6	20	huarache C
III	c	0	172	0	191	milpal D, E	III	c	0	172	0	191	huaracheo D, E
V	c	0	172	0	191	miltomate D, H	III	c	0	172	0	191	huauzoncle H
I	b1	0,1	60	15,6	71	mitote C	IV	b3	0	139	0,3	166	huehuenche E
I	a1	0	172	5,19	115	mitotero D	II	b3	4,73	9	67,1	38	huipil C
III	b1	0,01	117	3,9	122	mixiote E	II	b3	0,03	98	2,9	130	huitlacoche C
I	b1	0,08	69	19,5	68	molcajete C	IV	b1	0,01	118	2,9	130	huizachal D, E
I	b1	0,91	24	576,1	7	mole C	III	b1	0,07	73	40,9	52	huizache E
III	b1	0	139	2,92	130	molote H	V	a1	0	172	0,32	166	huizachera D, E
V	c	0	172	0	191	nacatamal D, ad	I	b1	5,42	7	127,2	25	hule C
III	b1	0,09	67	28,56	56	naco H	I	b2	0,01	118	17,5	70	itacate C
II	b1	0,03	98	7,8	97	nagual C	VI	c	0	172	0	191	ixcamole D, H
I	b1	7,44	5	178,5	21	náhuatl ad	V	c	0	172	0	191	ixtabentín H
V	a2	0,02	109	0	191	nahuatlato D, ad	I	b3	0,5	33	12,9	77	ixtle C

ixtlero D, E	191	0	102	0,03	a2	II	nanche ad	191	0	139	0,002	a2	IV
jacal C	18	227,8	90	0,04	b2	I	nauyaca C	191	0	118	0,01	a2	IV
jícama C	59	24,9	56	0,13	b1	I	neutle E	166	0,32	172	0	a1	V
jicamero D, H	166	0,32	172	0	a1	II	nexcomil H	191	0	172	0	c	V
jícara C	48	44,1	35	0,49	b1	I	nixtamal C	65	20,4	86	0,05	b1	I
jicote ad	148	1,3	135	0	b1	III	nopal C	13	313,6	11	2,57	b1	I
jilote C	191	0	172	0	c	III	nopalera C, D	97	7,8	118	0,01	b1	I
jilotear D, H	148	1,3	172	0	a1	IV	ocelote E	84	10,4	82	0,05	b1	III
jiloteo D, E	191	0	172	0	c	IV	ocochal D, E	191	0	172	0	c	VI

Vocablo	F	Iu (F)	NF	Iu (NF)	G	LP
ocote C	55	32,1	67	0,09	b1	I
ocotero D, E	191	0	172	0	c	II
ocotillo D, E	191	0	139	0	a2	IV
olote C	87	9,74	172	0	a1	II
olotera D, H	191	0	172	0	c	III
otate E	84	10,4	61	0,1	b1	III
otatillo D, ad	191	0	172	0	c	V
oyamel C	130	2,92	45	0,23	b3	III
pagua H	191	0	172	0	c	II
paliacate C	30	109,7	29	0,74	b1	I
papaloquelite D, H	191	0	172	0	c	IV
papalote C	73	14,6	70	0,08	b1	I
papazul H	191	0	172	0	c	V
pascle ad	191	0	172	0	c	VI
paxclal D, H	191	0	172	0	c	VI
pepenador C, D	79	11,7	77	0,06	b1	I
pepenar H	78	12,3	118	0,01	b2	I
petaca C	42	57,1	102	0,03	b2	I
petacón D	191	0	172	0	c	I
petate C	10	413,2	27	0,82	b1	I
petateada D, ad	191	0	172	0	c	IV
petatearse C, D	141	1,95	172	0	a1	I

Vocablo	F	Iu (F)	NF	Iu (NF)	G	LP
tejolote ad	141	1,95	172	0	a1	II
tejuino H	191	0	172	0	c	V
temascal E	191	0	109	0,02	a2	IV
tenamascle C	191	0	172	0	c	V
tenate H	97	7,79	172	0	a1	III
teocali C	148	1,3	139	0	b1	III
tepache C	139	2,6	139	0	b.1	I
tepachería D, E	191	0	172	0	c	I
tepalcate C	97	7,8	118	0,01	b1	II
tepeguaje ad	191	0	172	0	c	V
tepetate C	94	8,11	109	0,02	b1	II
teponaztle H	191	0	139	0,002	a2	III
tequescamoteD, H	191	0	172	0	c	VI
tequesquite H	148	1,3	139	0	b1	II
tequila C	3	1460,7	15	2	b1	I
tequilera D, H	166	0,3	135	0	b3	I
tescal ad	191	0	172	0	c	VI
tezontle C	87	9,7	54	0,13	b3	II
tianguis C	75	13,6	16	1,57	b3	II
tilma ad	79	11,7	43	0,27	b3	III
tinacal E	191	0	172	0	c	III
tiza E	82	11,4	89	0,05	b1	III

Word						
peyote C	87	9,74	44	0,24	b3	IV
pibil H	191	0	139	0,002	a2	III
pilmama ad	148	1,3	139	0	b1	I
pinacate C	105	6,82	98	0,03	b1	III
pinole C	53	37	82	0,05	b2	I
piocha ad	130	2,92	139	0	b1	II
pípilo H	191	0	172	0	c	II
pizote E	164	0,65	118	0,01	b3	VI
popote C	68	19,5	118	0,01	b2	I
popotillo D, E	191	0	172	0	c	III
pozol(e) C	45	46,7	76	0,07	b2	I
pulcazo D E	191	0	172	0	c	II
pulque C	14	309	20	1,17	b1	I
pulquería C, D	49	44,1	50	0,19	b1	I
pulquero C, D	105	6,82	172	0	a1	I
quelite C	111	5,84	64	0,1	b3	II
queltismo D E	191	0	172	0	c	VI
quesquémetl H	191	0	172	0	c	III
quetzal E	56	28,5	32	0,6	b1	III
quilotamal D, H	191	0	172	0	c	VI
quintonil ad	166	0,32	172	0	a1	III
quiote C	166	0,32	172	0	a1	V
salbute H	191	0	139	0,002	a2	VI
socoyote ad	191	0	172	0	c	IV

Word						
tlachique E	191	0	172	0	c	III
tlachiquero D	148	1,3	172	0	a1	III
tlaco ad	87	9,74	172	0	a1	III
tlaconete ad	148	1,3	172	0	a1	II
tlacoyo ad	105	6,82	172	0	a1	III
tlacuache C	41	58,4	51	0,18	b1	II
tlacuil H	191	0	172	0	c	V
tlapalería C	65	20,4	102	0,03	b2	I
tlapalero D, H	166	0,32	172	0	a1	I
tlascal H	191	0	172	0	c	VI
tocayo ad	130	2,9	109	0,02	b1	I
toloache C	122	3,9	139	0	b1	II
tololoche ad	191	0	172	0	c	III
tomate C	28	111,9	3	8,79	b3	I
tompiate ad	191	0	139	0,002	a2	III
topil E	146	1,62	139	0	b1	VI
totol H	148	1,3	139	0	b1	V
totomoxtle H	191	0	172	0	c	VI
totopo H	191	0	172	0	c	II
tucero D, H	191	0	172	0	c	VI
tular D, H	191	0	118	0,01	a2	III
tule C	51	42,2	52	0,14	b1	II
tuza C	65	20,4	109	0,02	b2	II
uchepo H	191	0	172	0	c	V

Word						
talache ad	191	0	172	0	c	II
tamal C	9	429,4	14	2,03	b1	I
tambache E	91	9,09	172	0	a1	I
tatemar H	130	2,9	139	0	b1	II
tayacán E	191	0	172	0	c	VI
tecali ad	166	0,3	139	0	b3	IV
tecolote C	37	68,1	38	0,4	b1	I
tecomate E	191	0	82	0,06	a2	V
tecotehue E	191	0	172	0	c	VI
tejamanil E	91	9,09	90	0,04	b1	III
tejocote C	58	27,3	80	0,06	b1	I
xolosóchil E	191	0	172	0	c	VI
yagual E	191	0	172	0	c	VI
zacahuistle ad	191	0	172	0	c	V
zacamiche H	191	0	172	0	c	VI
zacatal D, E	166	0,32	118	0,01	b3	I
zacate C	29	110,7	31	0,66	b1	I
zacatón D	148	1,3	139	0	b1	III
zacatonal D, E	191	0	172	0	c	IV
zapote H	75	13,6	54	0,13	b1	I
zontle C	191	0	172	0	c	VI
zopilote C	15	277,5	73	0,07	b2	I