Morphosyntactic Variation In Spanish – Global And American Perspectives
Eeva Sippola
University of Helsinki
eeva.sippola@helsinki.fi

Abstract:
This paper models structural variation among Spanish varieties with a special focus on Latin America. Based on a database built on secondary sources, I analyse structural data from 48 Spanish varieties and contact varieties. The results of the analysis show that contact influence is reflected in the grouping of varieties, especially from a global perspective. Among Latin American varieties, regional groupings appear, showing also traces of contact for individual varieties and the influence of national standards. The results reflect previous classifications of Spanish varieties, and show the usefulness of computational methods in comparative studies of global Spanish variation.

1. Introduction

This article studies dialectal classifications of Spanish varieties, with a special focus on varieties spoken outside Europe. Spanish is spoken on four continents, and throughout the world, the Spanish varieties exhibit substantial variation. This variation is due to many factors that have affected the evolution of Spanish, including the natural drift of languages over time, contact with other languages, colonisation and migration patterns, language propagation through missionary activities, urbanisation and the consequent rural–urban sociolinguistic divisions, official language policies, educational systems, literacy, and mass communication media (Lipski 2012: 1). This variation has rarely been approached from a global, comparative perspective, beyond general overviews.
and works focusing on regional variation (e.g. Alvar 1996a, 1996b; Lipski 1996; Moreno Fernández & Otero Roth 2016; Quilis 1992).

In this paper, I focus on modelling structural variation among Spanish dialects and contact languages from a global and regional perspective, especially in the Americas. The aim is to offer new fuel to the debates on the classification of Spanish varieties. Using computational methods, I test how structural comparisons respond to classifications proposed in the previous literature. Finally, the aim is to test if the varieties show outcomes related to prolonged contact between Spanish-speaking settlers and indigenous populations in colonial and postcolonial settings. In general, this article offers new input to Spanish dialectology by examining the potential of phylogenetic analysis in the classification and comparison of Spanish varieties.

Comparative typological studies on varieties of English have served as an inspiration for this study (e.g. Kortmann & Lunkenheimer 2012, Szmrecsanyi & Kortmann 2009, Szmrecsanyi this volume). These studies have successfully shown how general typological and dialectal data can be combined to provide a comprehensive overview of variety types in global languages such as English and Spanish, open new perspectives on the influence of substrate and/or other factors, and help to visualise differences between postcolonial language varieties.

This study is based primarily on data from Lipski’s (1996) cross-dialectal overview of Latin American Spanish and existing databases on Spanish contact varieties (Perez et al. 2017). Lipski (1996) is one of the most comprehensive overviews of the varieties, bringing together several independent studies on dialectal variation in the Americas. Also, its scope of description is greater than that of other overviews of the Spanish spoken in the Americas (e.g. Canfield 1981; Moreno de Alba 1995; Zamora & Guitart 1988; Cotton & Sharp 1988).

This paper is structured as follows. In section two, I give an overview of the current state and history of Spanish dialect classification in the Americas. Section three presents the methodological framework and the feature and variety selection. The analysis follows in four. Sections five and six cover the discussion and the conclusions and set out an agenda for future research.

2. Variation in the Hispanic world
2.1. Classification of Spanish varieties

One of the traditional classifications in order to make sense of the variation in Spanish is the division between the varieties spoken in Spain and in the Americas or elsewhere, such as in Africa and Asia (c.f. Perez this volume). However, linguistically, this grouping is somewhat arbitrary. There is great diversity among the various Latin American varieties of Spanish, and it would be hard to point to one trait shared by all of them which is not also in existence in one or more of the varieties of Spanish used in Spain. Similarly, African and Asian varieties of Spanish show both peninsular traits and other local and contact influences.

In Latin America, there is no consensus on the classification of Spanish varieties for several reasons, ranging from geography and histories of contact to the state of description of the varieties spoken in different regions and the sociolinguistic realities in them (Lipski 2012: 3). The territory where Spanish is spoken is very large and spans over the Americas. From a linguistic point of view, many regions have not been accurately described or documented. In addition, language contact and multilingualism are present in many environments. Sociohistorical developments, such as immigration, internal migrations, and significant rural–urban linguistic polarisation, have also left their imprint on linguistic forms and uses across the Americas. Several dialectal divisions have been proposed, based mostly on phonological features. Factors explaining the variation include geography, migration routes and patterns, and contact (e.g. Canfield 1981; Henriquez Ureña 1921).

Lipski (1996, 2012: 3) provides a dialect classification based on the degree of contact with other languages, models of colonial settlement and administration patterns, and the rural–urban divide. This classification, which results in eleven dialect areas, combines information on sociohistorical and language contact patterns as well as phonetic and morphological traits of the dialects in question. They are listed from north to south as follows:

- Mexico (except for coastal areas) and the U.S. Southwest
- Caribbean region: Cuba, Puerto Rico, Dominican Republic, Panama
- Caribbean coast of Colombia and Venezuela, Caribbean coast of Mexico, and the Mexican Pacific coast
- Guatemala, parts of Yucatán, and Costa Rica
- El Salvador, Honduras, and Nicaragua
- Colombia (interior) and neighbouring highland areas of Venezuela
- Pacific coast of Colombia, Ecuador, and Peru
- Andean regions of Ecuador, Peru, Bolivia, northwestern Argentina, and northeastern Chile
- Chile
- Paraguay, northeastern Argentina, and eastern Bolivia
- Argentina (except for extreme northwest and northeast) and Uruguay

This classification brings together different approaches to dialect division that have been presented in previous studies, and it can be seen as a compromise between them, but no individual grammatical traits are given as defining the groupings. Other classifications have been given throughout the history of studies on Latin American Spanish, and these can be divided into two main groups (following Lipski 2011): classifications based on geographical variation and classifications dealing with sociolinguistic variation. The following overview of the main divisions and approaches will serve as a background to our study, providing insights to the variables and traits generally included in dialectal and variationist studies of Spanish.

2.2. Geographical variation

Geographical variation was central in the early days of studies of Spanish variation. In addition to language-internal developments, it often shows influence from long-term language contact. According to Lipski (2011), classification within this approach to variation is based on the presence or absence of specific traits, with a strong focus on phonology.

Henríquez Ureña (1921) provided one of the first classifications of dialect areas in the Americas based on descriptive studies and his understanding of the indigenous substrate as seen on the lexicon in different regions (Lipski 1996: 18, 22–23). Henríquez Ureña’s dialect areas included Mexico and Central America (Nahuatl substrate), the Caribbean (Caribbean language substrates), Andean Highlands (Quechua substrate), central and southern Chile (Mapudungun/Araucanian substrates), and the Río de la Plata (Guaraní substrate). Other distinguishing features included phonetic traits and voseo, the use of the second person singular pronoun vos and its conjugational verb forms, although he could not match this feature with the dialectal
areas proposed. The phonetic features corresponded to discontinuous dialectal areas, as later observed also by Canfield (1981).

In a similar vein, Rona (1964/1993) proposed a total of 16 geographic dialectal zones and six areas that showed influence from contact with indigenous languages. His analysis focused on phonetic traits, such as *yeismo* and fricative realisations in similar contexts, and *voseo* patterns, which are characteristic of the Río de la Plata area. The study put forward a definition of the Latin American dialects as a dynamic convergence of groups of isoglosses, which reflects the current understanding of Spanish dialectology in the Americas relatively well. However, in practice, Rona’s division is rather abstract and general, and it does not present any overlapping areas or acknowledge any internal variation. Furthermore, Río de la Plata dialect features such as *voseo* are overrepresented, which leads to a detailed description of dialects close to this area, but to oversights and imagined unity in more distant dialect areas (Lipski 1996: 30–31).

Following an approach similar to Rona (1964/1993), Zamora & Guitart (1988) provided yet another classification, resulting in nine dialectal zones. They added a division into phonetically conservative or innovative dialects. Those in the first group, such as the Andean highland areas or the Mexican altiplano, do not typically exhibit different forms of consonant weakening, such as the aspiration of */s/*, loss of intervocalic */d/*, lenition of */x/* to */h/*, loss or lenition of */r/*, liquid neutralisation, etc. Resnick (1975) also used phonetic variables to study dialectal variation, using a method that aimed to overcome the limitations of previous geographic approaches. His binary coding of phonetic variables1 made it possible to identify individual dialects much more precisely than previous methods. However, Lipski (1996: 25) observes that sociolinguistic variation is only very partially included in Resnick’s database.

From this overview it becomes evident that phonetic traits have had a central role in dialect divisions in the geographical space. In addition to phonetic traits, the lexicon is another area that is traditionally used in dialectology. In the case of Latin

---

1 Resnick (1975: 7–8) proposed a binary value coding for traits such as the retention of syllable-final and word-final */s/*, realisation of */r/* and */x/*, distinction between */j/* and */ʎ/*, fricative */b/* in the group */lb/*, word-final */n/* before a following vowel or pause as velar */ŋ/*, */l/* versus */r/* in final position, and full sonority of all vowels in unstressed position, in sentence-final position, and between unvoiced consonants.
American Spanish, the data collected is abundant, but focuses mostly on peculiarities of individual countries or dialects, and is thus of little comparative value (Lipski 1996: 35). As an example, it is worth mentioning Cahuzac’s (1980) lexico-semantic study, which mapped 600 words that have to do with the rural world. Four dialectal zones were distinguished, following Henríquez Ureña’s (1921) substratist division quite closely.

In terms of structural features, most analysis has focused on pronoun forms and related conjugational forms of *voseo*, and no general comparisons based exclusively on structural features have been presented so far. However, more detailed studies have recently examined other structural traits that differentiate Spanish dialects, such as word order patterns, the behaviour of object clitics, and verb tense and mood. These features form part of our database and are presented and discussed in more detail in section 5. More recently, the *Interactive Atlas of Spanish Intonation* (Prieto & Roseano 2009–2013) systematically presents audio and video materials for the study of prosody and intonation in various dialects of Spanish from different age groups. The project aims for a comprehensive study of the great dialectal diversity present in the intonation of Spanish.

These studies and traditional dialect atlases of Spanish (for a list, see Moreno Fernández n.d.) provide a wealth of information to researchers, but take little or no deviation within individual locations into account (Pottier 1992; Lipski 2011: 72). Consequently, studies on geographical variation do not offer a solid base for accurately depicting regional variation that would take sociolinguistic factors into consideration (Lipski 1996: 167–168).

### 2.3. Sociolinguistic variation

Modern sociolinguistic work has been centred on dialect-internal variation that challenges geographical isoglosses and clear boundaries between dialect areas. Social class, age, gender, and other factors, such as identity, ideology, pragmatics, and the functional load of certain traits, can function as explanatory factors for the variation in Latin America (Lipski 2011; Schwenter 2011). Rapid social changes, such as urbanisation, increased mobility, and the spread of new media and communication patterns are also affecting dialect variation today (Lipski 2012: 21–22).

There has been a wealth of individual studies conducted on specific locations or groups (for an overview, see Schwenter 2011; Bentivoglio & Sedano 2011). Overall,
syntactic variation appears in the areas of verbs, clitic pronouns, and subordinate clauses and their matrix clauses. Some examples of particular features are variation in the pluralisation of haber ‘there is/are’, differences between the uses of morphological and periphrastic future constructions (e.g. comeré vs. voy a comer ‘I will eat’), uses of relative que and resumptive elements, and pseudoclefts and constructions with a focaliser ser (e.g. Lo que María compró fue un libro ‘What Maria bought was a book’ vs. María compró fue un libro ‘Maria bought was a book’) (Bentivoglio & Sedano 2011). Whereas most of this variation can be observed throughout the Spanish-speaking world, constructions with focalising ser occur only in some Latin American dialects of Spanish. Many of these features vary not in terms of presence or absence but rather in degree of frequency, which is conditioned by social factors (Bentivoglio & Sedano 2011).

Important advances have also been made with regard to oral corpora created in recent years, with a special focus on the educated speech of urban areas. Some examples include the oral corpus of recordings from 1975–2004 as part of the Corpus de Referencia del Español Actual (Real Academia Española) and the Macrorcorpus de la norma lingüística culta de las principales ciudades del mundo hispánico (Samper et al. 1998). These recent developments are encouraging, but so far, there has been a limited selection of varieties, styles, and registers available for comparative purposes. Currently, the PRESEEA (2014) project aims to create a corpus for the sociolinguistic study of Spanish from Spain and America, with a unified methodology that would provide opportunities for comparative studies on variation. So far, individual studies on selected varieties and features have been published under the umbrella of the project, and the future promises wider approaches.

3. Modelling classifications with networks

Based on the idea that Latin American Spanish variation has a complex mixture of social, geographic, and contact-derived influences, it is necessary to have a method that can model Latin American dialect variation in a representative way. In this study, I use the NeighborNet cluster analysis method in the program SplitsTree v.4.11.3 (Bryant & Moulton 2004; Huson & Bryant 2006). This program is especially suitable for representing the effects of language contact, as it can handle both vertical and lateral
aspects of language development. NeighborNet produces calculations of similarities and differences between individual varieties based on shared features, and ultimately, it creates a graphic representation of a phylogenetic network showing these relationships. The analysis is based on statistical modelling and probabilities and is especially suitable for testing hypotheses.

Phylogenetic network graphs can provide important insights into the degree of similarity or distance between languages and varieties that are known to be related. In our case, all the varieties studied are dialects of Spanish or related to a historical variety of Spanish (e.g. the creoles included in the sample). In these graphs, each set of lines represents a division of varieties into two groups, with the length of the lines representing the average distance between them. Similarity between varieties or languages results in clusters in the network. However, it should be kept in mind that although a common ancestor is assumed on a general level, this clustering does not necessarily mean that the languages share a common origin or path of development. These graphs show implicit networks of similarities and differences, and linguistic and sociolinguistic information must be interpreted from them.

This method has for some time now been used in linguistics, e.g. in studies on language evolution, dialectology (McMahon et al. 2007; Szmrecsanyi & Wolk 2011), typology (Kortmann & Wolk 2012: 191; Wichmann & Good 2014), and creole studies (Bakker et al. 2017), to mention a few (see also Kortmann & Wolk 2012: 191). In the following, two studies that used NeighborNet and are thus relevant for the analysis are presented.

Kortmann & Wolk (2012) present a classification of 74 varieties of English, based on a comparison of 235 morphosyntactictic features in the World Atlas of Varieties of English (WAVE) database. Based on NeighborNet diagrams, their general results show, among others, that the varieties cluster into four different types: traditional L1 varieties, indigenized L2 varieties, pidgins and creoles, and a mixed cluster of pidgin, high-contact L1 varieties, and L2 varieties (Kortmann & Wolk 2012: 920; see also Kortmann & Szmrecsanyi 2004; Szmrecsanyi & Kortmann 2009: 1651). The types can be characterised as follows (Kortmann & Lunkenheimer 2013): Traditional L1 varieties are traditional, regional non-standard varieties that have been spoken as mother-tongue varieties and had a relatively low degree of contact with other varieties since the beginning of the colonial period (e.g. Scottish English, Appalachian English). The high-contact L1 variety group is more diverse, but characterised by a high degree
of contact between different dialects of English and/or between English and other languages. They include transplanted L1 Englishes or colonial standards, language-shift Englishes, and standard L1 varieties (e.g. Irish English, Colloquial American E, Philippine English). The indigenised L2 varieties of English do not have native speakers. They are normally spoken in areas where English was introduced in the colonial era, typically via school, and enjoys prestige and is used in education and other official domains still today (e.g. Pakistani English). Pidgins and creoles are contact languages that developed in situations in which people without a common language developed a new means of communication. Pidgins are often restricted to certain domains and not spoken as mother tongues (e.g. Butler English). Creoles, on the other hand, have become the native language of the majority of the population (e.g. Jamaican Creole). In Kortmann & Wolk (2012), a purely structural morphosyntax-based typological profiling largely matches the socio-historical classifications of the WAVE varieties into these different variety types (c.f. Hackert this volume; Szmrecsanyi this volume).

Regarding Spanish varieties, these methods have been used in exploring the classifications of Afro-Hispanic varieties. Perez et al. (2017) is an exploratory study on how to classify three Afro-Hispanic varieties based on a selection of typological and dialectal data (see also Perez this volume). A total of 72 morphosyntactic and phonetic features were used in the comparison, and the results show Afro-Hispanic varieties clustering as a separate group between Spanish-based creoles and heritage and standard varieties. This supports the classification of the Afro-Hispanic varieties as extremely non-standard varieties of Spanish which reflects the extra-linguistic histories of the varieties.

4. Varieties and the feature sample

Lipski’s (1996) overview of the dialectal traits of Latin American Spanish is the main source of data for the Spanish varieties analysed in this study. Although important research has been conducted in the past two decades, Lipski’s book remains the most comprehensive overview of the Latin American varieties to date. This work relies on an extensive bibliography, which he synthesises into convincing descriptions and explanations. As for the other varieties analysed in this study, the sample is built mainly
on other works by Lipski, as these provide a coherent conceptual and theoretical presentation.

The coding of the data proceeded in the following manner. First, information on the variation and features in Lipski (1996) were listed, and different feature descriptions in the book were compared and divided into different groups according to the area of grammar they belonged to: phonology, nominal categories and syntax, verbal categories, clausal syntax, and pragmatically marked structures (following Dryer & Haspelmath 2013 and Michaelis et al. 2013). Finally, this resulted in a dialectological feature list with 40 grammatical features of which many deal with pronoun forms and behaviour.\(^2\) In addition, features that are shared by the *Atlas of Pidgin and Creole Language Structures* (APiCS) and *World Atlas of Language Structures* (WALS) were selected and converted into binary values, with a total number of 74 features. The complete feature list is presented in Appendix I.

The inclusion of the general typological features was done especially with the contact varieties in mind. Some of the shared APICS and WALS features had already been coded for Spanish in the original WALS database. These were compared for the Spanish varieties with information presented in Lipski (1996), and if no special mention was made of regional or social traits, the general value for “Spanish” was assumed for the varieties in question. Although it might occasionally seem irrelevant to include certain typological traits shared by the great majority of the Spanish varieties in the comparison, doing so can help to obtain a wider perspective, where contact influence or different areal factors, such as distance to urban, more standard-speaking centres or capitals, can have an effect on the traits present in the varieties. These data, which indicate the presence or absence of a range of features in the varieties, were then summarised as numerical values (e.g. absence = 0, presence = 1).

On the basis of the features, data from 34 Latin American Spanish varieties were extracted from Lipski’s dataset. An additional 14 varieties were then added to the database from other sources. These included three Peninsular Spanish varieties (Alvar 1996a), three Afro-Hispanic varieties (based on Perez et al. 2017), four other indigenised and heritage varieties of Spanish (data from Lipski 1985, 1987, 1988; Penny 2004), and three Spanish-based creoles and a mixed language (based on APiCS).

\(^2\) In addition, 28 phonetic features were extracted but due to limitations of space, the focus here is on the morphosyntactic traits.
Afro-Hispanic varieties include Afro-Chinchano (Peru), Afro-Choteño (Ecuador), and Afro-Yungueño (Bolivia). These varieties have been described as extreme non-standard varieties of Spanish that show traces of early creolisation or L2 acquisition of Spanish (Lipski 2007; Perez et al. 2017). Spanish varieties spoken in Equatorial Guinea, Philippines, and in Sabine River (Louisiana), together with Judeo Spanish, are grouped under the label “indigenised and heritage varieties”. They are spoken in multilingual contexts, as in Equatorial Guinea or Philippines, and in the case of the Philippines, Sabine River, and Judeo Spanish in Turkey and the Balkans, Spanish had a more prominent role or used to be spoken more extensively by earlier generations. These varieties show contact influence from other non-Romance languages, or from processes of second-language acquisition or language attrition and shift (Perez et al. 2017: 274). Of the six Hispanic varieties in APiCS, I have chosen to include Media Lengua, Palenquero, Papiamentu, and finally Zamboanga Chabacano as a representative of the Philippine Spanish creole varieties. These languages show very high degrees of restructuring and contact influence from non-Romance languages. This selection provides us with 48 varieties that can be used in the overall comparison, presented in Table 1.

<table>
<thead>
<tr>
<th>Country</th>
<th>Dialect region</th>
<th>Acronym</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Argentina</td>
<td>ARb</td>
</tr>
<tr>
<td></td>
<td>Buenos Aires</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Northwest</td>
<td>ARn</td>
</tr>
<tr>
<td>3</td>
<td>Central region</td>
<td>ARc</td>
</tr>
<tr>
<td>4</td>
<td>Bolivia</td>
<td>BOa</td>
</tr>
<tr>
<td>5</td>
<td>Los Llanos</td>
<td>BOl</td>
</tr>
<tr>
<td>6</td>
<td>Panama</td>
<td>PN</td>
</tr>
<tr>
<td>7</td>
<td>México</td>
<td>MEc</td>
</tr>
<tr>
<td>8</td>
<td>Northwest</td>
<td>MEn</td>
</tr>
<tr>
<td>9</td>
<td>Yucatán</td>
<td>MEy</td>
</tr>
<tr>
<td>10</td>
<td>Cuba</td>
<td>CU</td>
</tr>
<tr>
<td>11</td>
<td>Chile</td>
<td>CHg</td>
</tr>
<tr>
<td>12</td>
<td>General</td>
<td>CHg</td>
</tr>
<tr>
<td>13</td>
<td>Colombia</td>
<td>COa</td>
</tr>
<tr>
<td>14</td>
<td>Highlands of the interior</td>
<td>COc</td>
</tr>
<tr>
<td>15</td>
<td>Caribbean coast</td>
<td>COc</td>
</tr>
<tr>
<td>16</td>
<td>Pacific coast</td>
<td>COp</td>
</tr>
<tr>
<td>17</td>
<td>Amazon region</td>
<td>COM</td>
</tr>
<tr>
<td>18</td>
<td>Costa Rica</td>
<td>CRvc</td>
</tr>
<tr>
<td>19</td>
<td>Central valley</td>
<td>CRgn</td>
</tr>
<tr>
<td>20</td>
<td>Coastal region</td>
<td>ECc</td>
</tr>
</tbody>
</table>
Table 1. Varieties in the sample.

It is necessary to stress that the kind of overview approach adapted here abstracts from detail and from often important characteristics of individual varieties (see also Kortmann & Szmrecsanyi 2004). For example, not all the variation presented in Lipski (1996) made its way into the final database, as occasionally I chose to omit a peculiar regional or social trait from the final comparison. For a comparative purpose, such individual features would probably not make a great difference if all the other features would be akin to the national standard. Similarly, many regional or social varieties were not included, as only insufficient information about them was available.

5. Network analysis of Spanish varieties
In this section, I evaluate the degree of similarity between Spanish varieties and assess the structural similarity based on the selection of different language sets. It should be kept in mind that the networks are the result of an analysis of Spanish varieties based purely on linguistic features, which does not take any historical or other evidence into account.

5.1. Variety type as a basis for classifications

The graph in Figure 1 shows 48 varieties compared on the basis of 114 structural features.

Figure 1. NeighborNet of 48 Spanish varieties with 114 features.

The general picture in Figure 1 reflects the fact that the varieties belong to different types: creoles, Afro-Hispanic varieties, indigenised and heritage varieties, and dialects of Spanish. Creoles, such as Palenquero (PL) and Papiamentu (PP) as well as the mixed language Media Lengua (ML), appear on the right side of the network, while the Afro-Hispanic varieties (AFco, AFci, ACyu) are in the same right-reaching branch, but with shorter independent lines. The creoles and Media Lengua are farthest away from the dialects of Spanish, and they each appear with long independent lines, which means that they are structurally different from each other. Other Spanish varieties from both Latin America and Spain form a star-shaped webbing at the left-hand side of the figure. In-between these, and closer to the star, we find Judeo Spanish (JS), Equatorial Guinea Spanish (EG), and Philippine Spanish (PS). Sabine River Spanish (SR) appears in the middle of the network. On a global level, the network clusters thus first and foremost reflect variety types that include varieties from different continents.
On a general level, geographical location or proximity as an explanatory factor is not reflected in the network. For example, no clear division between European Spanish varieties and Latin American Spanish varieties is represented in the groupings, and European varieties (EScn, ESm, ESca) are in fact close to Mexican varieties (MEn, MEc, MEy). However, varieties from the Andean region and Argentina, Central America, and the Caribbean form areal clusters.

When only varieties that are geographically located in the Americas are compared, the overall result stays the same, although there is a clearer division between regional dialects of Spanish and the Afro-Hispanic varieties and creoles, as seen in the right-reaching branch in Figure 2. Varieties that have often been characterised as bilingual varieties, such as the Spanish spoken in the Andean highlands in Ecuador (ECa), Peru (PEa), or Bolivia (BOa), and Fronterizo (i.e. a mixed lect with Portuguese, cf. Perez, this volume) in Uruguay (URf), do not group closer to these restructured varieties based on this structural comparison. This indicates that restructuring in creoles and Afro-Hispanic varieties leads to different outcomes than in bilingual situations.

Figure 2. NeighborNet of 41 Spanish varieties in the Americas with 114 features.

5.2 Areal signals
When excluding the creoles, Media Lengua, Afro-Hispanic varieties, and indigenised and heritage varieties, and Fronterizos from the comparison, a somewhat different picture emerges. In Figure 3, the same 114 features were applied to 33 varieties of Spanish in Latin America. The split graph has a more star shaped from, producing a webbing in the middle. In this selection, the areal signal seems to be somewhat noticeable, although some exceptions prevail. On the right side of the network, we find varieties from countries in the Central Andes: Spanish varieties from the Ecuadorian
coast (ECc) and highlands (ECa); Peruvian Spanish varieties from the Amazonas (PEm), highlands (PEa), and the central coastal region (PEc); and Bolivian highlands (BOa) and lowlands (BOl) Spanish. However, within this group, the smaller clusters are formed according to national borders, and no regional contact influence is traceable in the clustering. Longer independent lines for the highlands varieties in Ecuador and Bolivia could be seen as representative of contact influence from the indigenous substrates.

Figure 3. NeighborNet of 33 American Spanish dialects with 114 features.

In the middle section of the webbing, Argentinian Spanish varieties form a clear independent cluster, as do the Central American varieties from Guatemala (GU), Nicaragua (NI), Honduras (HO), and El Salvador (EL). In the middle section, we also find Uruguayan (UR), Paraguayan (PA), and Chilean Spanish varieties (CHg, CHn) with independent lines. Three Colombian varieties appear clustered together in the middle, and of these, the Amazonas variety (COm) is somewhat separate from the Colombian Coastal Pacific (COp) and Colombian highlands (COa) varieties. The Costa Rican varieties (CRve, CRcc, CRgn) are also placed together, showing a structural unity.
Caribbean and Mexican Spanish varieties are placed on the left side of the network. A cluster of Caribbean varieties is located at the left of the graph, including varieties from the Dominican Republic (RD), the Caribbean coast of Colombia (COc), Panama (PN), Cuba (CU), Puerto Rico (PR), and Venezuela (VE). Of the Mexican varieties, Yucatán (MEy) appears with a long independent line, indicating some structural differences from the rest of the cluster. Similarly, the Dominican Republic differs somewhat from the rest of the Caribbean cluster, and on the other side of the graph, the Andean highland varieties have a long independent line showing their distinction from neighbouring regional varieties. These patterns support idea of contact influence as a divergent factor.

In summary, when the cases that present the heaviest contact influence or restructuring are excluded from the comparison, regional clusters emerge. However, the results are still highly conditioned by national standards, as the only variety clustering differently from its national sister varieties is the Colombian Caribbean variety of Spanish. This clustering does not completely reproduce the dialect areas by Lipski (1996, 2012) or other researchers, although some overlaps can be seen. Chile appears alone, as it does in most previous classifications. Argentina and Uruguay are not significantly separated from each other, but the Argentinian varieties clearly cluster alone. The Caribbean region clusters together, as do the countries of the Andean region and the Central American varieties, with the Costa Rican exception. One reason for this discrepancy is clear: dialectal classifications take phonetic characteristics into account, but they are completely excluded from this comparison.

6. Discussion

The results presented in section 5.1 support previous findings about the classification of contact varieties of Spanish (Perez et al. 2017), which showed that creoles and Afro-Hispanic varieties clustered differently from standard varieties of Spanish due to their different structural profile. This is not surprising, as part of the database is actually shared between the studies. However, with the extended structural database used in this study, the differences are highlighted even more. These results also reflect similar findings from the structural comparison of English varieties worldwide (Kortmann & Wolk 2012), which suggest that varieties cluster according to their variety type. The
types of traditional L1 English varieties and pidgins and creoles reflect well the dialect and contact variety clusters for Spanish varieties in Figure 1. In contrast to the distinct cluster of indigenous varieties of English, however, the indigenised and heritage varieties of Spanish in this analysis do not cluster together as clearly. For example, Sabine River Spanish never clusters with the other varieties in this group. This could be the case because Sabine River Spanish, being a vestigial variety of Mexican Spanish, does not show much restructuring or contact influence from typologically different languages.

Contact influence has had a prominent role in the geographical dialect classifications of Spanish with focus on phonology and lexicon. In this structural comparison, many dialectal varieties of Spanish spoken in areas with a large numbers of speakers of indigenous languages, such as Yucatán Spanish, the Amazonian variety of Colombian Spanish, and the highlands varieties of Ecuadorian and Bolivian Spanish diverge more from the central webbing. However, this influence is not homogeneous, and even these varieties appear close to their national sister varieties.

The clusters also support the idea that geography matters when the morphosyntax of varieties belonging to the same variety type is compared, as in the case of English varieties (Kortmann & Wolk 2012). Within Latin American varieties of Spanish, areal clusters appear along national and regional borders, with some exceptions. The lower level clustering by national borders is probably due to a shortcoming in the adaptation of the presentation structure in Lipski (1996) where the data are represented according to national boundaries, although he acknowledges that these divisions cannot be seriously maintained except for a handful of small homogeneous nations (Lipski 1996: 17, 2012:3). As the feature and variety selection of this study was built on Lipski (1996), the national standards have also shaped the final database. Although regional dialectal forms and traits by bilingual or bicultural parts of the population in selected regions are mentioned in the individual chapters, the influence of the national standard in the description is strong.

On the basis of the results provided here, selected structural traits from the different clusters can be discussed with special focus on the varieties of Latin American Spanish. A closer look at the presence or absence of certain features shows a very diverse overall picture. This is in line with the current understanding of Latin American dialectology. In general, regional Latin American Spanish varieties behave similarly with regard to most of the shared APiCS-WALS features, but the creoles, Media
Lengua, and the Afro-Hispanic varieties show different patterns of these features. In addition, some regional dialects, such as many of the Caribbean varieties, have a higher frequency of overt subject pronoun usage than others (Lipski 2012, but see Carvalho et al. 2015). Although the WALS-APiCS features only focus on word order in declarative main clauses, some Spanish varieties also show variation in word order patterns in interrogative sentences. For example, the subject generally comes after the verb in Spanish interrogatives, but in some varieties, subject-verb inversion does not occur in absolute interrogatives, in sentences with interrogative words, or when the subject is a pronoun (Lipski 2012: 13).

Pronoun forms play a prominent role in the dialectal feature list. Second person subject pronouns (tú, vos, vosotros, ustedes) and the accompanying verb forms of voseo show variation across the Spanish-speaking world. The pronoun forms are also often applicable to the creoles, but the dialects of Spanish show even more distinctions in the forms and meanings. Voseo, both with regard to the pronoun use and the verbal forms, shows variation according to both national groups and individual varieties.

In Spanish, the object clitics lo, la, and le, replace direct object noun phrases, but there is variation with regard to their use with pronoun objects and nouns according to the dialect. In our database, this variation appears for individual varieties and regional groups. For example, in the Andean region, clitic doubling is common for pronouns and animate nouns, and it can also occur with inanimate definite direct objects (Lipski 1996, 2012: 12–13). Possessive constructions present more variation in Spanish dialects spoken by bilingual populations and the varieties in the creole group than in the Latin American dialects in general.

With regard to the domain of verb tense, the preterite–present perfect distinction (e.g., llegó ‘arrived’ – ha llegado ‘has arrived’) is often mentioned as separating the Spanish and Latin American varieties. In most of Spain, both the preterite and the present perfect can be used when the moment of speaking is not included, while in Latin America, the simple preterite can also include the present moment. However, this feature shows a different picture in the database for this study, as the distinction was not always mentioned for the individual or national varieties in Lipski (1996), but described for Cádiz and Canary Islands. In the area of mood, some individual dialects present a rupture of the consecutio temporum, and permit subjunctive verbs to be used in subordinate clauses in the present tense even when the verb of the main clause is in past tense (e.g. Juan me dijo que lo haga en seguida ‘Juan
told me to do it right away’ (Lipski 1996: 195). Standard Spanish would have *hiciera* ‘do.PST. SBJV.3SG’ instead of *haga* ‘do.PRS.SBJV.3SG’.

Well-known differences are also found in the use of the diminutive suffixes –*ito/-ita* and –*ico/-ica*. In our database, the latter appears in Cuban, Dominican, Colombian, and Costa Rican Spanish, but Lipski (1999) also lists Venezuela and Ecuador as regions that use it.

Although these initial results are encouraging, one should also keep in mind that we are dealing with a preliminary dataset based on Lipski (1996), including partly problematic data and missing data points and features, as seen in this section. For instance, some of the original feature definitions were imprecise, especially from a comparative point of view. For example, direct object omission in the Andean varieties is described with different levels of detail for different varieties. In Bolivia (Lipski 1996: 214), the omission of the clitic can happen when the object has been mentioned previously, but the examples given for Ecuador do not specify if the full noun in object position has to precede the omission (Lipski 1996: 269). With regard to the feature design, the features included in the database are not always applicable to all the varieties. For example, the Afro-Hispanic varieties and creoles differ structurally from Spanish varieties in that they present very limited inflectional patterns. This makes some of the features not applicable to them, such as for example the verbal forms of *voseo*. These are coded differently and excluded from the NeigborNet analysis. Finally, the number of features is relatively low and should be increased in the future to include more features from recent sociolinguistic studies (see 2.3).

In addition, the nature of the reference point varies, as well as the varieties described, are abstractions with different data types (Perez et al. 2017: 285–287). These include L1 and L2 varieties where differing traits are occasionally highlighted in order to make the particularities even more evident. The sociolinguistic settings of the varieties in question also range from monolingual situations to highly multilingual environments, and it is challenging to extract one repertoire or code of internal variation to be included in the comparison. For these reasons, one should read the results with a grain of salt, and they should be complemented by more detailed qualitative studies on individual language varieties and situations.

7. Conclusions
A network analysis of structural features based on a late 20th century overview of varieties of Spanish shows that variety type is a conditioning factor in large-scale comparisons. On the regional level, the results respond to previous classifications relatively well. Contact influences can be interpreted from longer independent lines for certain varieties in the Latin American comparison, but due to the typological differences among the indigenous substrates, the resulting picture is heterogeneous and these varieties do not cluster together. In addition, national borders and standards seem to condition the analysis. Latin American Spanish varieties are thus influenced by language contact but not exclusively shaped by it.

The present study demonstrates that phylogenetic network analysis is a useful tool to complement the current understanding of dialectal variation in Spanish as a global language. These first large-scale results point towards new roads in the dialectology of Spanish in the Americas. The design of features and the selection of varieties should be further defined and motivated from a theoretical perspective, but these results are encouraging as to the possibilities that statistical methods offer for comparative studies. By offering a broad overview of Spanish variation from a regional and global perspective, this study adds to comparative research about the development of postcolonial language varieties in different settings around the world.

**References**


Cahuzac, Philippe. 1980. La división del español de América en zonas dialectales. 
Situación etnolingüística o semántico-dialectal. Lingüística Española Actual 
Canfield, Lincoln D. 1981. Spanish pronunciation in the Americas. Chicago and 
Carvalho, Ana M., Rafael Orozco & Naomi Lapidus Shin (eds.). 2015. Subject 
pronoun expression in Spanish. A cross-dialectal perspective. Washington, 
D.C.: Georgetown University Press.
D.C.: Georgetown University Press.
Structures Online. Leipzig: Max Planck Institute for Evolutionary 
Anthropology.
Henríquez Ureña, Pedro. 1921. Observaciones sobre el español de América. Revista 
de Filología Española 8: 357–390. 
http://www.linguas.net/LinkClick.aspx?fileticket=duFUEpxdPGA%3D&tabid 
=79&mid=421 (10.7.2016)
Huson, Daniel H. & David Bryant. 2006. Application of phylogenetic networks in 
and syntactic variation in English. In Bernd Kortmann, Kate Burridge, Rajend 
Gruyter.
Kortmann, Bernd & Christoph Wolk. 2012. Morphosyntactic variation in the 
Anglophone world: A global perspective. In Bernd Kortmann & Kerstin 
Lunkenheimer (eds.), The Mouton World Atlas of Variation in English, 906– 
Variation in English. Berlin/Boston: Walter de Gruyter.
Kortmann, Bernd & Lunkenheimer, Kerstin (eds.) 2013. The Electronic World Atlas 
of Varieties of English. Leipzig: Max Planck Institute for Evolutionary 
McMahon April, Paul Heggarty, Robert McMahon, & Warren Maguire. 2007. The sound patterns of Englishes: Representing phonetic similarity. English Language and Linguistics 11: 113–142. DOI: 10.1017/S1360674306002139
Perez, Danae, Sandro Sessarego & Eeva Sippola. 2017. Afro-Hispanic varieties in comparison – new light from phylogeny. In Peter Bakker, Finn Borchsenius,


Real Academia Española. (s. f.). Corpus de Referencia del Español Actual (CREA). Madrid: RAE.


Samper, José A., Clara Eugenia Hernández Cabrera & Magnolia Troya (eds.). 1998. *Macrocorpus de la norma lingüística culta de las principales ciudades del mundo hispánico (MC-NLCH)*. Las Palmas de Gran Canaria: Servicio de Publicaciones de la UPGC.


https://dx.doi.org/10.1590/S1984-63982011000200011


**Appendix I**

Feature list
Features 1–74 are modified from the shared APiCS-WALS list (Michaelis et al. 2013), while features 75–114 are a selection from Lipski (1996).

1. Order of subject, object, and verb: SOV
2. Order of subject, object, and verb: VSO
3. Order of subject, object, and verb: SVO
4. Order of possessor and possessum
5. Order of adjective and noun
6. Order of adposition and noun phrase
7. Order of demonstrative and noun
8. Order of cardinal numeral and noun
9. Order of relative clause and noun
10. Order of degree word and adjective
11. Position of interrogative phrases in content questions
12. Gender distinction in personal pronouns
13. Inclusive/exclusive distinction in independent personal pronouns
14. 3-way politeness distinctions in second person pronouns
15. 2-way politeness distinctions in second person pronouns
16. Generic noun based indefinites
17. Occurrence of nominal plural markers
18. Plural word after the noun
19. Plural word precedes the noun
20. Plural suffix
21. No definite article, but indefinite article
22. Definite articles distinct from demonstratives
23. Indefinite articles
24. Pronominal and adnominal demonstratives
25. Distance contrasts in demonstratives
26. Adnominal distributive numerals
27. All ordinal numerals are suppletive
28. Suppletion in low ordinal numerals, but no higher than three
29. Sortal numeral classifiers
30. Marking of possessor noun phrases
31. Secondary surpass marking
32. Locative comparative marking
33. Comparative standard particle marking
34. Suppletion according to tense and aspect
35. The prohibitive: normal imperative, normal negative
36. The prohibitive: normal imperative, special negative
37. The prohibitive: special imperative, normal negative
38. Alignment of case marking of full noun phrases
39. Alignment of case marking of personal pronouns
40. Ditransitive constructions with 'give'
41. Clitics as weak, pronouns as strong
42. Obligatory pronouns
43. Optional pronoun words
44. Expression of pronominal subjects: pronoun affixes
45. Comitatives and instrumentals: identity
46. Noun phrase conjunction and comitative: overlap
47. Noun phrase conjunction and comitative: identity
48. Noun phrase conjunction and comitative: differentiation
49. Nominal and verbal conjunction: identity and differentiation
50. Nominal and verbal conjunction: differentiation
51. Nominal and verbal conjunction: identity
52. Predicative noun phrases: no copula
53. Predicative noun phrases: variable copula
54. Predicative noun phrases: obligatory copula
55. Predicative noun phrases and predicative locative phrases: overlap
56. Predicative noun phrases and predicative locative phrases: identity
57. Predicative noun phrases and predicative locative phrases: differentiation
58. Predicative possession
59. Intensifiers and reflexive pronouns
60. Reciprocal constructions: identity
61. Reciprocal constructions: differentiation
62. Reciprocal constructions: mixed
63. Applicative constructions exist
64. Subject relative clauses: verbal affix
65. Subject relative clauses: relative particle or pronoun
66. ‘Want’ complement subjects left implicit
67. Negative morpheme types: negative particle
68. Negation and indefinite pronouns: predicate negation present
69. Negation and indefinite pronouns: existential
70. Negation and indefinite pronouns: mixed
71. Polar questions: question particle
72. Polar questions: intonation only
73. Polar questions: interrogative verb morphology
74. Polar questions: interrogative word order
75. Infinitive form ending in /–r/
76. Verbal forms of voseo: -ás, -és, -ís
77. Verbal forms of voseo: -áis, -éis, -ís
78. Verbal forms of voseo: -ás, -ís, -ís
79. -ito used for diminutives
80. -ico used for diminutives
81. Omission or alteration of articles
82. Grammatical gender marked in the noun phrase
83. Vos as a 2nd person singular pronoun
84. Usted exists as a 2SG pronoun (formal or familiar)
85. tú as a familiar 2SG pronoun
86. tú exists as a formal 2SG pronoun
87. su merce(d) exists as a pronoun
88. leísmo
89. loísmo
90. Redundant possessives with inalienables: de mi cabeza sufro ‘My head aches’
91. Redundant possessive pronouns in possessive constructions: de la María su casa ‘María’s house’
92. Indefinite article + possessive pronoun: un mi gato ‘my cat’
93. Null direct object / Omission of direct object pronouns: al chófer le di ‘I gave [it] to the driver’
94. Direct object duplication with a clitic when the object is definite or a person: Lo conozco a Juan ‘I know Juan’
95. Repeated clitics pre- and post-verbally: *la he podido conocerla* ‘I was able to meet her’
96. Preverbal *nos* in exhortatives: *nos sentemos* [=sentémonos] ‘Let’s sit down’
97. *yo = a mí* in dative verbs: *yo me parece* ‘It seems to me’
98. No direct object clitic pronouns when direct object in left position: *las elecciones nunca entendí* ‘The elections, I never understood them’
99. Pleonastic *lo: te lo fuiste de mí* ‘you went away from me’
100. Direct object pronouns in nominative: *Cuando él mira nosotros* ‘When he looks at us’
101. Avoidance of double clitic pronouns: *dame = damelo* ‘give it [to me]’
102. Subject pronoun repetition: *yo vengo y yo compro* ‘I come and I buy’
103. Pleonastic/intensive use of *ser: lo conocí fue en la fiesta* ‘I met him at the party’
104. Reflexive *se* generalized: *se [=nos] llevamos bien* ‘we get along well’
105. Reflexive *se* omitted
106. Preterite = present perfect: *no llegó = no ha llegado* ‘He has not arrived [yet]’
107. Periphrastic future preferred: *voy a comer* ‘I will eat’
108. *hasta* = start of an event, e.g. ¿*hasta cuando viene el jefe?* ‘When will the boss come?’
109. *Tener* ‘to have’ as existential verb
110. *ser and estar* have overlapping meanings/functions
111. Changes in the consecutio temporum: *Juan me dijo que lo haga en seguida* ‘Juan told me to do it immediately’
112. Content questions without inversion: ¿*Qué tú compraste?* ‘What did you buy’
113. Double negation
114. *che* as a vocative