Black box approaches to genealogical classification and their shortcomings

Jelena Prokić and Steven Moran

1. Introduction

In the past 20 years, the application of quantitative methods in historical linguistics has received a lot of attention. Traditional historical linguistics relies on the comparative method in order to determine the genealogical relatedness of languages. More recent quantitative approaches attempt to automate this process, either by developing computational tools that complement the comparative method (Steiner et al. 2010) or by applying fully automatized methods that take into account very limited or no linguistic knowledge, e.g. the Levenshtein approach. The Levenshtein method has been extensively used in dialectometry to measure the distances between various dialects (Kessler 1995; Heeringa 2004; Nerbonne 1996). It has also been frequently used to analyze the relatedness between languages, such as Indo-European (Serva and Petroni 2008; Blanchard et al. 2010), Austronesian (Petroni and Serva 2008), and a very large sample of 3002 languages (Holman 2010). In this paper we will examine the performance of the Levenshtein distance against n-gram models and a zipping approach by applying these methods to the same set of language data.

The success of the Levenshtein method is typically evaluated by visually inspecting and comparing the obtained genealogical divisions against already well-established groupings found in the linguistics literature. It has been shown that the Levenshtein method is successful in recovering main languages groups, which for example in the case of Indo-European language family, means that it is able to correctly classify languages into Germanic, Slavic or Romance groups. In a recent analysis of the Austronesian languages by means of Levenshtein distance (Greenhill 2011), the obtained results were evaluated using a more exact method than by visually inspecting the recovered groups. Greenhill (2011) extracted language triplets and compared their subgroupings against those provided by the Ethnologue (Lewis 2009). The possible subgroupings of any three languages included the following: (1) language A is more similar to language B than C, (2) A is more similar to C than B, (3) B is more similar to C than A, or (4) A, B and C are equally

Bereitgestellt von | De Gruyter / TCS
Angemeldet
Heruntergeladen am | 16.10.19 14:00

similar. The comparison of two classifications has shown that the accuracy of the Levenshtein method in languages classification reaches only up to 65%. Furthermore, it has been observed that the accuracy of Levenshtein classification decreases rapidly with phylogenetic distance.

Although the Levenshtein algorithm takes into account very little linguistic knowledge about the segments being compared, those in favor of this approach stress that it gives reasonable results, that it can be computed quickly, and that it can easily be applied to large amounts of data. In this paper, we apply Levenshtein's algorithm to sixty-nine indigenous South American languages, and look into more detail what this algorithm is actually measuring and how meaningful are the groups it obtains. We also analyze the same data set using two very simple techniques: an n-gram model and a gzip file compression method. Both of these methods are very simple and require no linguistic knowledge about the data being analyzed. The n-gram method measures the number of overlapping segments, i.e. in our case unigrams and bigrams of phones in words, without regard to the position of the grams. This approach has been applied by Huffman (2003) to the task of language classification. Gzip is a file compression method based on the Lempel-Ziv algorithm (Ziv and Lempel 1978) that searches for the longest common substring between strings. It has been used by Benedetto (2002) to classify 50 Indo-European languages into genetic groups. We show that there is no significant difference in the performances of these three techniques and that they are only partially successful in finding major language groups. None of these approaches reveals linguistic processes that are responsible for the differences found between the languages. The lack of a language model makes any of these black box approaches unsuitable for the investigation of deep phylogenetic relationships between language varieties. We argue that more linguistically-aware methods, or hybrid methods that use black box approaches coupled with linguistic knowledge, should minimally produce linguistic output that is useful for historical linguists, who remain the front runners in revealing deep genealogical relations between languages.

2. Methods

In this section we give a short introduction to the three methods that we use to measure the distances between languages: the Levenshtein method, the n-gram model and the zipping approach.

Bereitgestellt von | De Gruyter / TCS
Angemeldet
Heruntergeladen am | 16.10.19 14:00

2.1. Levenshtein distance

Levenshtein distance is a metric used to measure the distances between two strings; it was first introduced by Levenshtein (1966). It represents the smallest number of edit operations (insertions, deletions or substitutions) needed to transform one string into the other. At the same time, it aligns the two strings, as illustrated in figure 1, which presents the alignment of pronunciations of the word for 'tree' in two Tucanoan languages, Siriano and Wanano.

j u k ɨ g ɨ
j u k ɨ k ɨ

Figure 1. Alignment of two pronunciations of the word for 'tree'

The aligned strings differ only in position 5, where [g] in Siriano corresponds to [k] in Wanano. The absolute distance between these two strings is 1 since they differ in only one position. There are several variants of the Levenshtein approach, but the most important ones are the normalized approach and phone string comparison. In order to discard the influence of the lengths of the strings being compared, Levenshtein distance can be normalized by dividing it by the length of the longer string (Serva 2007) or by the length of the alignment (Heeringa 2004). In our example in figure 1, both normalization methods would give a distance of 1/6. In the phone string comparison approach, the compared segments are represented as a bundle of features, which allows for a more refined comparison. Since [k] and [g] are both velar plosives, voiceless and voiced respectively, the distance between these two segments can be set to 1/3 instead of 1.[1] If the strings that are being compared are cognate forms that differ in only few segments, then the Levenshtein approach lets us get very accurate alignments. The aligned segments [k] and [g] in our example, thus share the same origin in a hypothetical protoform. However, when comparing languages that are more distantly related, the words become less similar in their surface forms and this makes the applications of the Levenshtein method for their comparison less appropriate.

2.2. N-gram analysis

An n-gram is a subsequence of *n* consecutive items from a given sequence. The size of *n* can range from 1 (unigrams) to the length of the string in question. N-gram models have been applied to language comparison by Huffman

Bereitgestellt von | De Gruyter / TCS
Angemeldet
Heruntergeladen am | 16.10.19 14:00

(2003) and Hoppenbrouwers and Hoppenbrouwers (2001) used frequency of single phones to compare dialect varieties of Dutch. In this paper, we compare the frequency of unigrams and bigrams in order to classify languages in our data set into genetic groups. The method is very simple and the only linguistic knowledge that it requires as input is the information on how to split words into phones. Unlike in the Levenshtein approach, no alignment of the word is involved. The similarity between two words is calculated as the number of shared unigrams or bigrams divided by the length of the longer word. The two words for 'tree' from figure 1 contain the following unigrams, shown in table 1.

Table 1. Two pronunciations of word 'tree' and their phone frequencies

	j	**u**	**k**	**ɨ**	**g**
jukɨgɨ	1	1	1	2	1
jukɨkɨ	1	1	2	2	0

The similarity between these two strings is 5/6 because they share 5 unigrams, i.e. phones: [j], [u], [k] and two times [ɨ]. This method produces a similarity metric between two strings. However, in order to get a distance matrix, similarities are converted into distances by subtracting them from 1. The distance matrix is used to calculate the genetic similarity of the languages under investigation.

2.3. Zipping

File compressors (aka zippers) are algorithms designed to encode a file in such a way that it uses fewer bits than the original and thus compresses its file size. One of the best-known data compression algorithms is the Lempel-Ziv algorithm (Ziv and Lempel 1978), which is used in many public domain compressors, such as *zip* and *gzip*. This algorithm works by searching for the duplicate strings in a file, i.e. longest common substrings, and recoding them into smaller strings. In files with many repeated patterns, there are more recoded strings and the compression rate is greater. Benedetto (2002) presents some of the possibilities of applying this algorithm for language recognition and authorship attribution. The distance between two texts A and B in two different languages is estimated by merging texts from two

Bereitgestellt von | De Gruyter / TCS
Angemeldet
Heruntergeladen am | 16.10.19 14:00

languages and measuring their compression rates. The more similar two texts are, then the higher the compression.[2]

For our approach, we use Normalized Compression Distance (NCD), as presented in Cilibrasi (2004), to measure the distance between two languages:

$$NCD(x,y) = (C(xy) - \min\{C(x),C(y)\})/\max(C(x),C(y))$$

C(xy) is the compressed size of the concatenated texts x and y. C(x) is the compressed size of x. And C(y) is the compressed size of y. To calculate the distance between each of the languages in our data set, we used the publicly available *gzip* compressor.

3. Data set

We tested each black box method on a set of sixty-nine indigenous South American languages extracted from Huber and Reed (1992). The data set consists of 366 word wordlists, based on a list developed by Morris Swadesh and John Rowe.[3] These wordlists were collected for indigenous languages spoken in Columbia. Huber and Reed (1992) classify these languages into 12 language families, most of which are commonly accepted (*ibid* p. V): Chocó, Chibcha, Barbacoa, Kamsá, Quechua, Arawak, Tucano, Carib, Guahibo, Macú-Puinave, Sáliba-Piaroa and Witoto.[4] The number of languages in different groups varies from only one (Kamsá and Quechua) to nineteen (Tucano). We investigate in detail the Tucano language family because it is comprised of the largest number of languages in our data set, and it is well attested in the literature (Campbell 1997; Kaufman 2007; Lewis 2009). According to these three sources, the Tucano language family can be divided into the Western, Eastern and Central Tucanoan branches. Figure 2 illustrates the classification of the Tucanoan languages in Huber and Reed as given in the Ethnologue (Lewis 2009).

According to this classification, the Eastern group can be further divided into Central and Northern, while the Western group is comprised of the Northern, Southern and Tanimuca groups. We use this classification to estimate the performance of the three black box methods.

Bereitgestellt von | De Gruyter / TCS
Angemeldet
Heruntergeladen am | 16.10.19 14:00

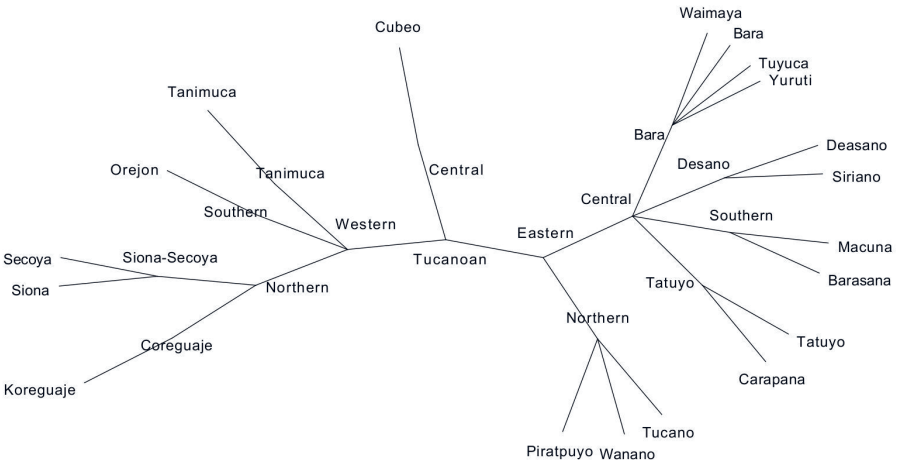

Figure 2. The Tucanoan family tree (Lewis 2009)

4. Results

Distances between the sixty-nine languages in our sample were calculated by means of Levenshtein distance, the amount of shared unigrams and bigrams, and by applying a zipping technique. All distances were analyzed using neighbor-net, as implemented in SplitsTree (Huson and Bryant 2006). Figure 3 shows the neighbor-net of all sixty-nine languages compared, using the Levenshtein algorithm.

The shape of the net in figure 3 is star-like with a very poorly marked hierarchical structure. The only three clearly distinguishable groups are the Chocó, Guahibo and Tucano language families. The rest of the families found in Huber and Reed (1992) can be identified, but the separation between various language groups is not very clear. This may be due to a separate evolution of these languages or it may mark a very weak phylogenetic signal. In figure 4, the same neighbor-net is shown after removing the Chocó, Guahibo and Tucano language families. Witoto and Arawak language families become more distinguishable, but the star-shape is still dominant.

Neighbor-nets of distance matrices obtained using unigram and bigram analyses are shown in figures 5 and 6 respectively. Both networks show high resemblance with the network based on Levenshtein distance. All 12 language families from Huber and Reed (1992) can be identified, with the

Bereitgestellt von | De Gruyter / TCS
Angemeldet
Heruntergeladen am | 16.10.19 14:00

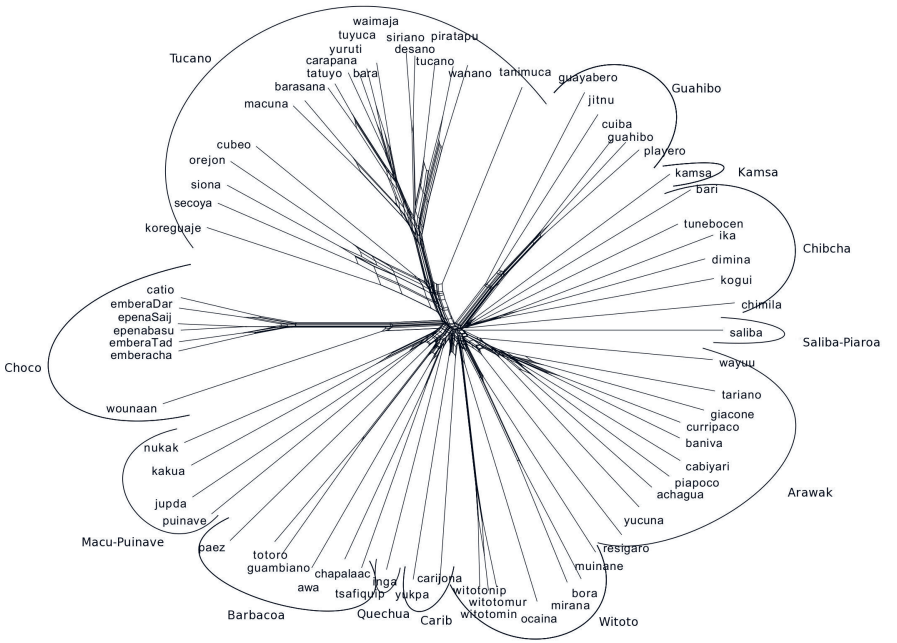

Figure 3. Neighbor-net of the sixty-nine languages compared using the Levenshtein algorithm

Chocó, Guahibo and Tucano families being the only three clearly distinguishable from the others. The rest of the network is star-shaped, which is especially visible in the network based on the bigram analysis. Although more simple than the Levenshtein algorithm, these two techniques give the same results with regard to the language classification on our data set.

In the next step, we analyzed the data by zipping the files as described in section 2.3 and measuring the difference in the compression rates. The results are shown in figure 7. Compared to the classification given in Huber and Reed (1992), some of the languages are misclassified. However, even using this very simple technique, it is possible to identify all language families. The Witoto family is clearly identified, unlike in the other two methods.

Regarding the classification task, the Levenshtein and n-gram models gave very similar results. They show very little hierarchical structure, with the Chocó, Guahibo and Tucano families being the only three exceptions. What makes these three language families different is that their word forms show, on average, much less variation if compared to the other language

Bereitgestellt von | De Gruyter / TCS
Angemeldet
Heruntergeladen am | 16.10.19 14:00

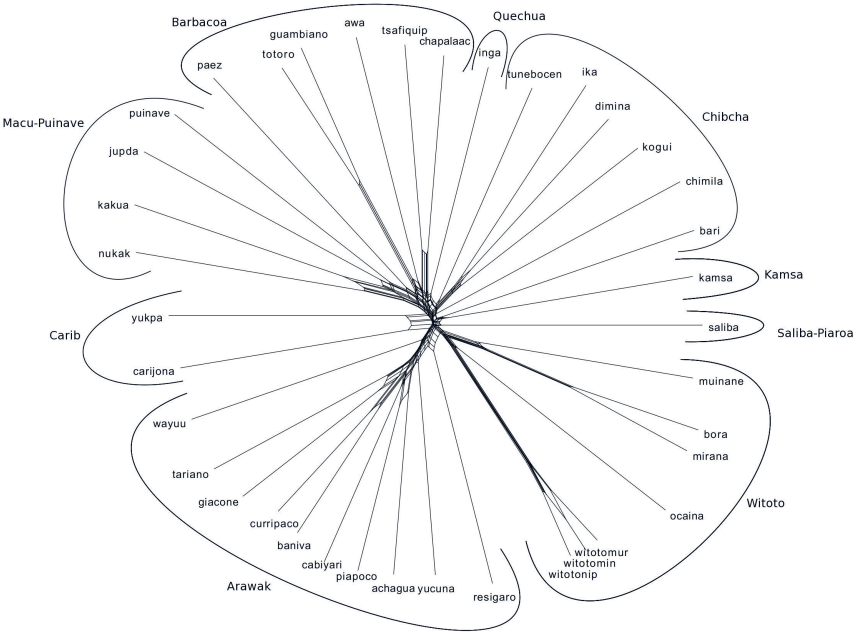

Figure 4. Neighbor-net presented in figure 3 after removing Chocó, Guahibo and Tucano language families

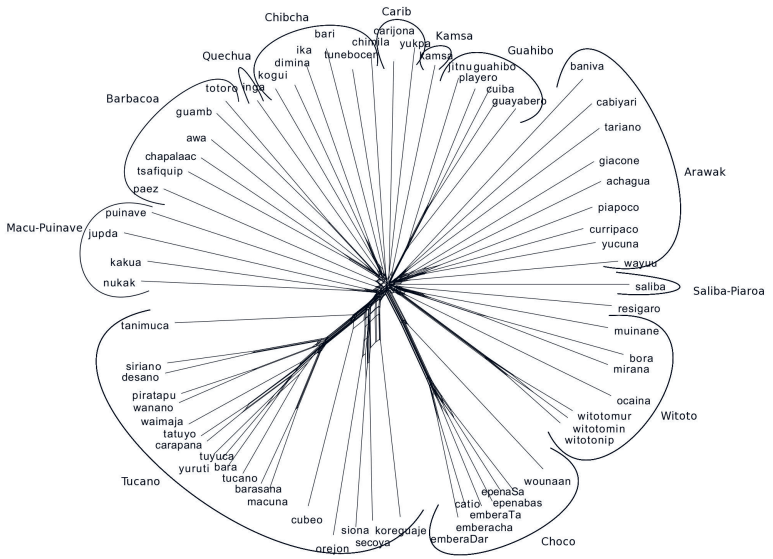

Figure 5. Neighbor-net of the 69 languages compared using the average number of shared unigrams

Bereitgestellt von | De Gruyter / TCS
Angemeldet
Heruntergeladen am | 16.10.19 14:00

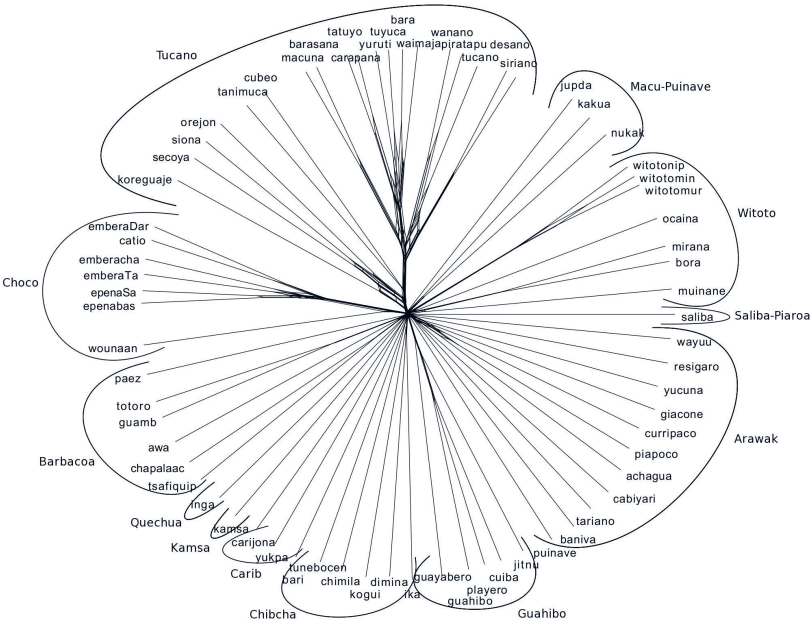

Figure 6. Neighbor-net of the sixty-nine languages compared using the average number of shared bigrams

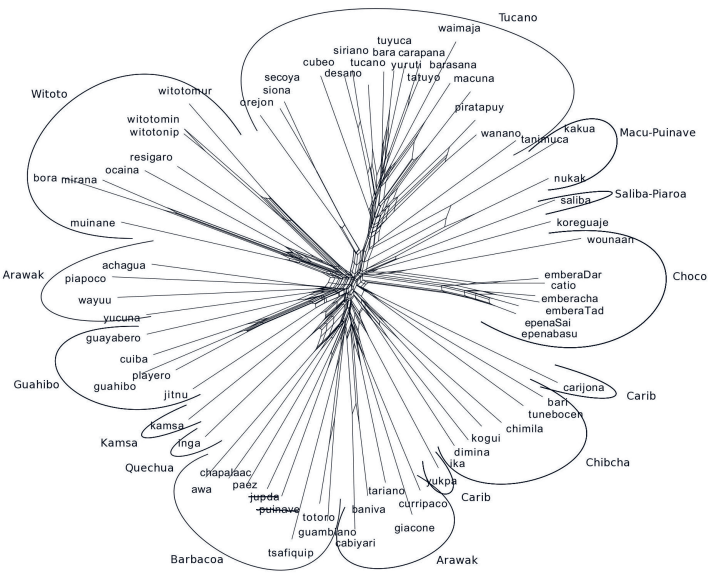

Figure 7. Neighbor-net of the sixty-nine languages compared using the zipping method

Bereitgestellt von | De Gruyter / TCS
Angemeldet
Heruntergeladen am | 16.10.19 14:00

families in the data set. Table 2 shows the pronunciation of the word 'leg' in the Chocó, Tucano and Chibcha language families.

Table 2. Pronunciation of the word 'leg' in Chocó, Tucano and Chibcha

Chocó	**Tucano**	**Chibcha**
hĩɾũ	jẽkãɨ́	ʤúʔkwɨ
hɛ́ɾũ	jĩtʃɨkɨ	kəldə
hĩ́ɾũ	jĩka	kɨna
hĩ́ɾã́	jĩkɨ́	kátto
hĩ́ɾũ	jĩkã̃	bidiinə

Unlike Chocó and Tucano, for languages where the word forms are very different, no internal structure is recovered. We discuss the reasons for the poor performance on remotely related languages in section 5. None of the tested methods gave any information on the relatedness between languages on the macro-family level.

We also checked the performance of the three algorithms on the Tucano languages solely by excluding non-Tucano languages from our analyses. The classification of the Tucano languages given in Ethnologue is given in figure 1. Neighbor-net based on the Levenshtein distances is shown in figure 8.

Although the neighbor-net method correctly reveals a major Eastern-Central-Western split and correctly groups languages at a very low-level in the net, it does not get the precise dividing line correct. For example, the following language pairs are grouped together: Siriano and Desano, Carapana and Tatuyo, and Macuna and Barasana.[5] However, Siriano and Desano are not grouped with the rest of the Central languages of the Eastern group. Tanimuca is classified as Central Tucanoan rather than Western Tucanoan. The net accords with the Ethnologue classification at the very high level (the split of the Tucanoan into Eastern, Central and Western) and at the very low level, but groupings at the intermediate level show differences. In figures 9 and 10, the analyses of the distances obtained by applying unigram and bigram methods to our data set are given. Both networks show the same structure as the network based on the Levenshtein distances. The network based on the zipping technique is presented in figure 11. The Eastern-Western split is less prominent and two languages, Tucano and Waimaja, are misclassified when compared to the Ethnologue's classification. Most of the lower level groupings can still be identified.

Bereitgestellt von | De Gruyter / TCS
Angemeldet
Heruntergeladen am | 16.10.19 14:00

Our analyses show that even in the case where language varieties exhibit relatively small variation, the Levenshtein method is successful only in identifying major splits. However, even more simple and less 'linguistically' informative methods are also able to detect the same major groups in the data. Identification of the subgroups is equally problematic for all three tested methods.

5. Discussion

The comparison of three methods evaluated in this paper shows that there is no significant difference in performance of the Levenshtein and n-gram approaches. Although Levenshtein involves alignments of the strings compared and takes into account the ordering of the segments, the classifications obtained show no improvement over the classifications based on simple phone frequency counts. The zipping method is able to identify main language divisions, but in both analyses (all data and the Tucano subset) it was less accurate than the Levenshtein and n-gram methods. Furthermore, relations between the families at the macro-family level were not retrieved by any of the methods. In order to discover these deep phylogenetic relationships, information about the cognacy of words and their regular sound correspondences is necessary. None of the methods we tested are able to distinguish between cognate and non-cognate words. By applying Levenshtein or n-gram methods on the non-cognate words, we get information on the chance similarity between the words. The fact that two non-cognate words share a certain number of phones does not reflect any genealogical relationship between them. The chance that two languages use non-cognate words to denote the same or similar meaning grows with the phylogenetic distance and black box approaches become unreliable tools for detecting the relationships between the languages.

If compared words are cognates whose surface forms differ in more than only one or two elements, which is often the case with the dialect data, then black box methods are often too simplistic to be able to correctly detect the phylogenetic signal. If we look at the pronunciations of the word 'drink' in Tucano and Siriano, it becomes clear that the black box approaches are overestimating the distance.

Tucano: s ĩ ʔ r ĩ j ã́
Siriano: – i ʔ r í k a

Bereitgestellt von | De Gruyter / TCS
Angemeldet
Heruntergeladen am | 16.10.19 14:00

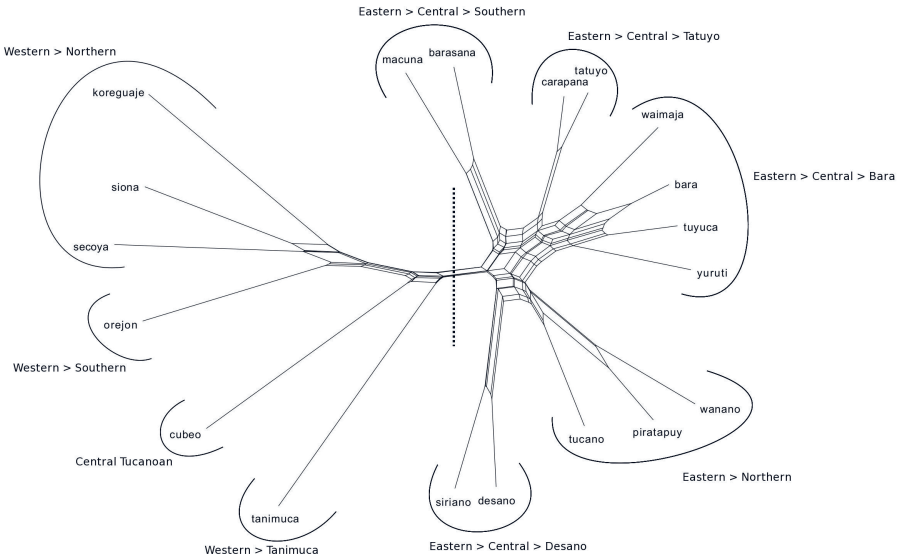

Figure 8. Classification of Tucanoan languages based on Levenshtein distance

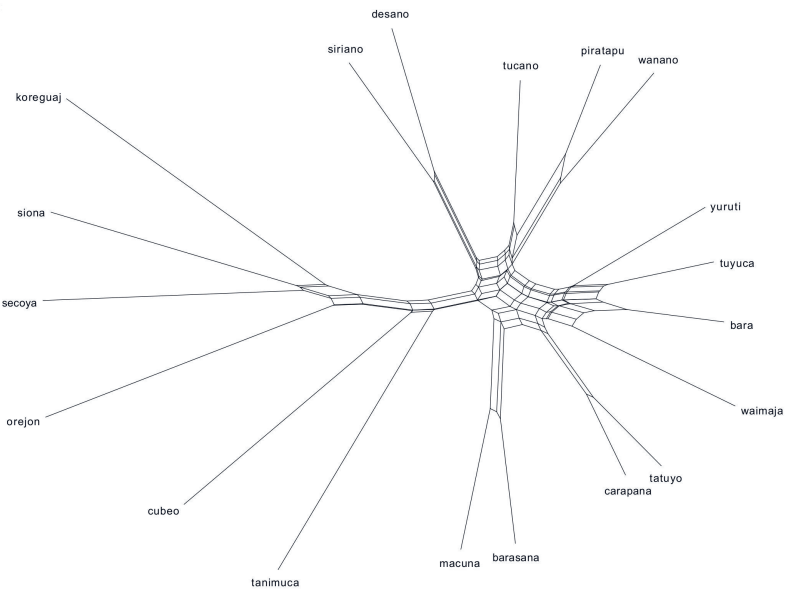

Figure 9. Classification of Tucanoan languages based on the average number of shared unigrams.

Bereitgestellt von | De Gruyter / TCS
Angemeldet
Heruntergeladen am | 16.10.19 14:00

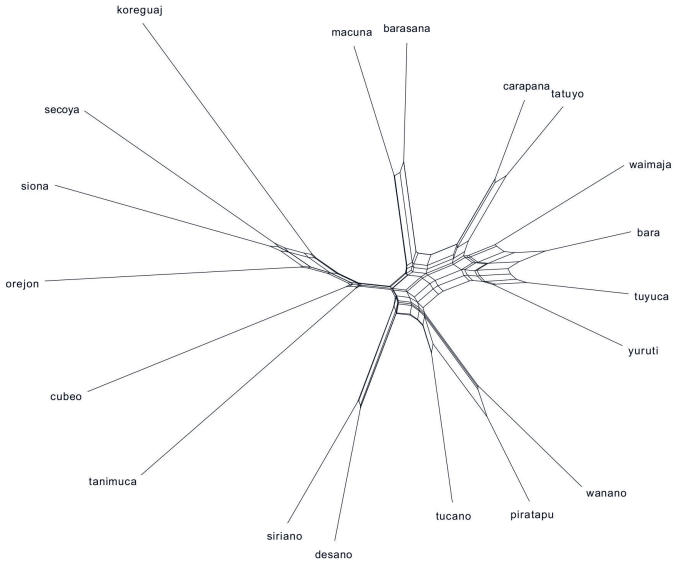

Figure 10. Classification of Tucanoan languages based on the average number of shared bigrams

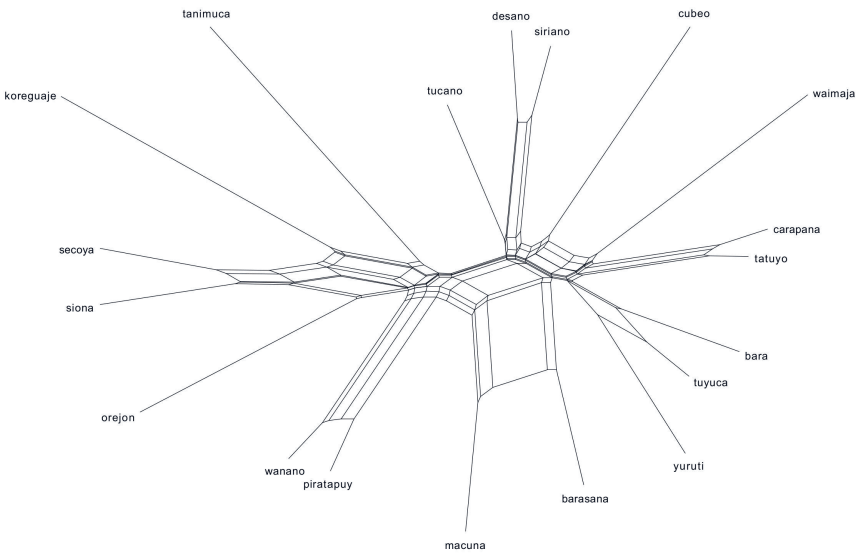

Figure 11. Classification of Tucanoan languages based on the zipping method

Bereitgestellt von | De Gruyter / TCS
Angemeldet
Heruntergeladen am | 16.10.19 14:00

Using the Levenshtein or unigram methods, the distance between these two strings is 5/7, since the only two matching phones are [ʔ] and [r]. If phones are evaluated binarily, i.e. they are either the same or are different, then the differences at the suprasegmental level, as well as secondary phonation types (e.g. nasalization, length, etc.) are weighted too heavily. Therefore, approaches to genetic classification need to go beyond the segment level and need to handle phylogenetic signals at the level of features. A feature system that is language data dependent would thus allow genetic classification approaches to handle issues of divergent but related sound changes, e.g. two related languages that expanded their number of vowels where the first uses nasalization and the second raises vowels.

Phylogenetic approaches need transparent models that explain the evolution of changes in languages and how they relate. These processes should explain how one phone changes into another and should capture changes at the level of phonemic systems. The argument that black box approaches give pretty good results is misleading. Black box approaches, as we have shown on a lesser-studied group of languages, do not give good results. They may trivially capture the higher-level groupings, but beyond those we cannot actually make any claims of genetic relatedness at the deeper genetic levels. Nor do these methods provide any additional information that is useful to historical linguists, e.g. regular sound correspondences, a list of recurrent sound changes, the probability of one change into another, a description of the trigging environments of change, probable cognates, etc. Black box approaches that simply count and align segments do not catch these processes. What is needed is a probabilistic model of language change that describes the relatedness of languages and discharges linguistically-relevant data about these processes.

6. Conclusion

In this paper, we applied three black box approaches, namely Levenshtein distance, two n-gram models and a zipping method, to a data set of sixty-nine South American languages that represent a lesser-studied language family. All three approaches use segment counts and statistics and do not leverage additional linguistic knowledge. We show that these three approaches produce roughly equivalent results, i.e. they capture high-level genetic groups, but fail to discover deep genetic classifications and splits. When these black box methods are used on sets of languages whose genetic classification has

Bereitgestellt von | De Gruyter / TCS
Angemeldet
Heruntergeladen am | 16.10.19 14:00

not been previously described by historical linguists, how can one claim anything beyond possible high-level splits?

Appendix: Languages and their genealogical affiliation according to Huber and Reed (1992)

Achagua	Arawak
Baniva	Arawak
Cabiyarí	Arawak
Curripaco	Arawak
Giacone	Arawak
Piapoco	Arawak
Tariano	Arawak
Wayuu	Arawak
Yucuna	Arawak
Resígaro	Arawak
Awa	Barbacoa
Cha'palaachi	Barbacoa
Guambiano	Barbacoa
Páez	Barbacoa
Totoró	Barbacoa
Tsafiqui pila	Barbacoa
Carijona	Carib
Yukpa	Carib
Barí	Chibcha
Chimila	Chibcha
D̷ḭm̷ina	Chibcha
Ika	Chibcha
Kogui	Chibcha
Tunebo	Chibcha
Tunebo Central	Chibcha
Catío	Chocó
Embera Chamí	Chocó
Embera Darién	Chocó
Embera Tadó	Chocó
Epena Basurudó	Chocó
Epena Saija	Chocó
Wounaan	Chocó
Cuiba	Guahibo
Guahibo	Guahibo
Jitnu	Guahibo
Playero	Guahibo
Guayabero	Guahibo
Kamsá	Kamsá
Jupda	Macú-Puinave
Kakua	Macú-Puinave
Nukak	Macú-Puinave
Puinave	Macú-Puinave
Inga	Quechua
Sáliba	Sáliba-Piaroa
Bará	Tucano
Barasana	Tucano
Carapana	Tucano
Cubeo	Tucano
Desano	Tucano
Koreguaje	Tucano
Macuna	Tucano
Orejón	Tucano
Piratapuyo	Tucano
Secoya	Tucano
Siona	Tucano
Siriano	Tucano
Tanimuca	Tucano
Tatuyo	Tucano
Tucano	Tucano
Tuyuca	Tucano
Waimaja	Tucano
Wanano	Tucano
Yurutí	Tucano
Bora	Witoto
Miraña	Witoto
Muinane	Witoto
Ocaina	Witoto
Witoto M̷ḭn̷ica	Witoto
Witoto Murui	Witoto
Witoto N̷ɨpode	Witoto

Bereitgestellt von | De Gruyter / TCS
Angemeldet
Heruntergeladen am | 16.10.19 14:00

Notes

1. Heeringa (2004) provides a detailed explanation of how to apply the Levenshtein method.
2. For a detailed explanation on how to estimate the distance, see Benedetto (2002).
3. An explanation on the wordlist collection can be found in Huber and Reed (1992).
4. We provide the full list of languages and their classifications in the appendix. Throughout this paper, we use the language names provided by Huber and Reed.
5. According to Campbell (1997), these language pairs are actually dialects.

References

Blanchard, Philippe, Filippo Petroni, Maurizio Serva, and Dimitri Volchenkov
2010 Geometric representations of language taxonomies. *Computer Speech and Language* 25 (3): 679–699.

Benedetto, Dario, Emanuele Caglioti, and Vittorio Loreto
2002 Language trees and zipping. *Physical Review Letters* 88 (4): 048702.

Campbell, Lyle
1997 *American Indian Languages: The Historical Linguistics of Native America*. Oxford: Oxford University Press.

Cilibrasi, Rudi and Paul M.B. Vitanyi
2005 Clustering by compression. *IEEE Transactions on Information Theory* 51 (4): 1523–1545.

Greenhill, Simon
2011 Levenshtein distances fail to identify language relationships accurately. *Computational Linguistics* 37 (4): 689–698.

Heeringa, Wilbert
2004 Measuring dialect pronunciation differences using Levenshtein distance. Ph.D. diss, University of Groningen.

Holman, Eric W.
2010 Do languages originate and become extinct at constant rates? *Diachronica* 27 (2): 214–225.

Huber, Randal Q. and Robert B. Reed
1992 *Vocabulario comparativo: Palabras selectas de lenguas indígenas de Colombia (Comparative vocabulary: Selected words in indigenous languages of Colombia).* Bogota: Instituto Lingüístico de Verano.

Huson, Daniel H. and David Bryant
2006 Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23 (2): 254–267.

Bereitgestellt von | De Gruyter / TCS
Angemeldet
Heruntergeladen am | 16.10.19 14:00

Huffman, Stephen M.
2003 The Genetic classification of languages by n-gram analysis: a computational technique. Ph.D. diss, Georgetown University.
Kaufman, Terrence
2007 *Atlas of the World's Languages*. 2d ed. London, New York: Routledge.
Kessler, Brett
1995 Computational dialectology in Irish Gaelic. In *Proceedings of the EACL*.
Levenshtein, Vladimir
1966 Binary codes capable of correcting deletions, insertions and reversals. *Doklady Akademii Nauk SSSR* 163: 845–848.
Lewis, M. Paul (ed.)
2009 *Ethnologue: Languages of the World*. 16 ed.
Online: http://www.ethnologue.com.
Nerbonne, John, Wilbert Heeringa, Eric van den Hout, Peter van de Kooi, Simone Otten, and Willem van de Vis
1996 Phonetic distances between Dutch dialects. In *CLIN VI: Proceedings of the Sixth CLIN Meeting*, G. Durieux, W. Daelemans, and S. Gillis (eds.), 185–202.
Petroni, Filippo and Maurizio Serva
2008 Language distance and tree reconstruction. *Journal of Statistical Mechanics: Theory and Experiment*. 2008: P08012.
Serva, Maurizio and Filippo Petroni
2008 Indo-european languages tree by Levenshtein distance. *Europhysics Letters* 81 (6): 68005.
Steiner, Lydia, Peter F. Stadler, and Michael Cysouw
2011 A pipeline for computational historical linguistics. *Language Dynamics and Change* 1 (1): 89–127.
Ziv, Jacob and Abraham Lempel
1978 Compression of individual sequences via variable-rate coding. *IEEE Transactions on Information Theory* 24 (5): 530–536.

Bereitgestellt von | De Gruyter / TCS
Angemeldet
Heruntergeladen am | 16.10.19 14:00

Bereitgestellt von | De Gruyter / TCS
Angemeldet
Heruntergeladen am | 16.10.19 14:00