Another Perspective in Historical Linguistics

Dr. Mohammed Alluhaybi

The Arabic Language Institute for Speakers of Other Languages, King Abdulaziz University, Jeddah, Saudi Arabia

malluhaubi@kau.edu.sa

Abstract. In historical linguistics, the comparative method is conventionally used to uncover the genetic relationships between languages. Such a method has shown accurate results in some cases, as in the case of Indo-European languages. However, it seems risky to apply it indiscriminately to other languages. In this paper, we first define the comparative method, and then take Papuan languages as an example to show how this method may not always be effective. Finally, we explain how Typology seems to be a useful way to assess genetic relationships among different languages, or at least support the results found by using the comparative method.

Keywords: Comparative method, Historical linguistics, Language families

1. Introduction

It is fairly common in linguistics to propose a concept and then apply it repeatedly (Dixon, 1997). An example of this can be seen in syntax, where syntacticians attempt to generate one theory that captures every possible way of combining words in all languages. The family tree model is another example. It is a tree diagram that represents the relationships between different languages. In this tree, one can see languages that share essential similarities assigned to another language, called the proto-language. Figure 1 below depicts this tree diagram model.

![Figure 1: Family tree model](image)

In this model, languages A, B and C are considered to be sisters, genetically related, and descend from one mother – the proto-language.

2. Comparative method

The comparative method is the cornerstone of historical linguistics where linguists try to describe and illustrate language change. It is a strategy that has been used to determine relationships among different languages. Upon a successful application of this method, a family tree can be drawn (see Figure 1 above).

Dixon (1997) says that in the family tree, we have an original language that splits into daughters over time, and daughters will also continue to split. Dixon (1997) also suggests two kinds of splitting: gradual and sudden. In both types, however, the splitting starts with one language, and that language becomes two dialects over time. The two dialects in turn will end up being two distinct
languages. If these two dialects did not separate geographically and still have some contact with each other, the split is considered to be sudden. Otherwise, Dixon considers it to be gradual. As a result, the proto-language (the original, the ancestor or the mother language) dies out. The significance of the comparative method arises to solve this problem. The aim of this method is to reconstruct and recover as much as possible of the proto-language by comparing similar items, cognates, in languages that are suspected to be similar. However, it should be borne in mind that cross-linguistically, languages show a great tendency to borrow either for need or prestige. To avoid this problem, linguists would usually start by looking at items that are less likely to be borrowed. These items are sometimes called basic vocabulary, such as kinship terms, low numbers and body parts. Roesch (2012) states that a hierarchy (Figure 2 below) is formulated by Muysken (1981) to rank categories by virtue of their resistance to borrowing. In this hierarchy, nouns are the most common category to be borrowed, while subordinating conjunctions are the most resistant ones.

**Figure 2: Categories based on their resistance to borrowing**

The reconstruction of a proto-language usually starts with phonology, where the ultimate goal is reconstructing the sound system (Campbell, 2004). This can be done using two strategies. The first one is the phonetic plausibility strategy, ‘which requires that any changes posited to account for differences between the proto-forms and later forms must be phonetically plausible’ (O’Grady & Archibald, 2012: 259). The other one is the majority rules strategy. The idea is that when we fail to apply the first strategy, we look for the most common segment in the cognates and consider it to be the proto one. To illustrate how the two strategies work, consider the following cognates from the Romance languages (O’Grady & Archibald, 2012).

<table>
<thead>
<tr>
<th>Language</th>
<th>Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>French</td>
<td>si</td>
</tr>
<tr>
<td>Italian</td>
<td>si</td>
</tr>
<tr>
<td>Romanian</td>
<td>ji</td>
</tr>
<tr>
<td>Spanish</td>
<td>si</td>
</tr>
</tbody>
</table>

Since the vowel [i] is the same in all forms, we consider it to be the original vowel that is preserved in the four languages. However, the initial consonant is different; it is [∫] in Romanian and [s] in the other languages. This variety gives us two possible proto-forms for the word ‘yes’; it is either *∫i or *si. In other words, we can assume that the proto-form has the phoneme [∫], and it has been changed to [s] in the three languages (French, Italian and Spanish), or the proto-form contained [s] and has been changed to [∫] in Romanian. To choose between these two possibilities, let us recall the two strategies used in the comparative

---

1 A cognate set of polysyllabic forms is stronger evidence than monosyllabic ones for suggesting a genetic relationship; however, in this example I used monosyllables because I am trying to explain a methodology rather than prove the genetic relationship among these languages.
Another Perspective in Historical Linguistics

method: the phonetic plausibility and the majority rules strategies. By applying these two strategies here, one can conclude that the proto-form is *si. That is, the process of changing [s] to [∫] is called palatalization, and it is a very common phenomenon in human languages. This is compatible with the first strategy (phonetic plausibility). The second, [s] is the most common consonant found among the four Romance languages. Therefore, with great confidence, one might say *si is the only reasonable option for the proto-form. In the following section, we demonstrate how the comparative method is not effective.

3. Problems with the comparative method

It seems appropriate to explain why the comparative method cannot be applied in some cases by giving specific examples. A good example in this case would be the Papuan languages, but before we go any further, a few words on these languages are in order.

The Papuan languages are spoken in the Pacific Basin, specifically on the island of New Guinea and nearby areas that are not claimed by either Austronesian or Australian family languages. They also extend east to occupy New Britain and some of the adjacent islands. The total number of Papuan languages is approximately 750 (Foley, 1986).

The term Papuan does not propose any genetic relationships among these languages. Therefore, it cannot be treated in the same way as we treat the Indo-European term. While the latter has been used to describe languages that descend from one ancestral language called Proto-Indo-European, the Papuan term does not. It serves only to distinguish certain languages from being either Austronesian or Australian.

An alternative term to describe Papuan languages is non-Austronesian. This term is more accurate as it ‘highlights their purely negative characterizations, i.e., that they are the languages of the Pacific which are not Austronesian’ (Foley, 1986: 3). However, Donohue (2004) suggests that using this term might cause some confusion because Austronesian languages are geographically in contact with another six family languages. In other words, not all non-Austronesian languages are Papuan, some of them might be Australian. Moreover, the Papuan term is widely prevalent in the literature. Consequently, it is the only one that will be used throughout this paper.

Unfortunately, the comparative method cannot be applied to Papuan languages because using such a method requires sufficient data to compare, and this is not available here. In addition, the comparative method has been used confidently with European languages because of the availability of early written documents in Greek, Latin and Sanskrit (Foley, 1986). The significance of these early documents is that once linguists reconstruct the Proto-Indo-European language, they can check their results by comparing the proto-forms to the ones written in the early documents. According to Foley (1986), there are no written documents for the Papuan languages. This means it is not possible to check the reconstructed forms. Another significance of the written documents comes from the information they provide, which can be used to separate borrowings from cognates. Without this separation, the comparative method could be deceptive because borrowing can lead to incorrect results. It is worth mentioning, though, that the
comparative method has been applied successfully to some unwritten languages such as Austronesian. However, the case here is significantly different. Austronesian languages are spread over a large territory; therefore, finding similar words in two languages that are located at the extreme ends of the territory would be good evidence to suggest that these words are cognates rather than being borrowed across this wide area (Foley 1986). Papuan languages, on the other hand, are located close to each other; thus, similar words are probably found because they are borrowed, not cognates. In addition, as Dixon (1997) suggests, languages split when their speakers become separated; for instance, some of them travel to live in a new place, and those travellers have no contact with the original group. In such a situation, speakers will develop two different dialects that will result in two distinct languages. However, this is not the case in Papuan languages as their speakers have some contact with each other. Despite this contact, Papuan languages are different and consequently, the logical explanation for their diversity would be because they descend from different proto-languages. In the following section, we summarise alternative methodologies that have been discussed in the literature to establish the genetic relationships Papuan languages, before we suggest our methodology in such cases.

4. Other methodologies proposed to substitute the comparative method

There are a few attempts in the literature to propose some alternative methods other than the comparative one, to uncover the possible genetic relationships among different languages. In this section, we discuss two of them as representative examples. One of the methodologies is suggested by Foley (1986), who categorises Papuan languages into 60 distinct language families. In his work, Foley notices that most of the borrowings among Papuan languages involve nouns not verbs. The reason is that nouns in Papuan languages are mostly uninflected, and the only place where obligatory morphology can appear is in verbs. Consider the following example from I’saka, one of the Sko languages (Donohue & Roque, 2004): ²

```
pu n-uni-ko Mark
bag 1SG-give-3SG.M.DAT Mark

‘I gave the bag to Mark.’
```

As can be seen from this example, the most complex part is the verb ‘to give’; it must be attached to affixes that encode information about the arguments of the clause, while the nouns ‘bag’ and ‘Mark’ are uninflected forms. Due to this complexity of verbs, they are more resistant to borrowing, which makes them good candidates to look at to establish genetic relationships. However, Foley does not look at verbs in general; instead, he follows a more conservative methodology by looking at verbal morphology. In his methodology, he considers only bound morphemes and the way speakers attach them to verbs. According to him, bound morphemes are too difficult to be isolated and borrowed. ‘If the languages with which speakers of X are in contact all have a system of noun classes (or gender), then language X is likely to develop its own system’ (Dixon,

---

² The abbreviations in this example are as follows: 1 for first person; 3 for third person; SG for singular; M for masculine; and finally, DAT for dative case.
Another methodology is based on the pronoun sets. This one is suggested by Ross (2005). His work involved only 605 languages out of the 750 Papuan languages because no data was available for the remaining ones. The reason for his consideration of the pronoun sets is that 311 contiguous languages were found, and their paradigms of pronouns seem to reflect the same proto-forms (Ross, 2005). Four possibilities can explain this phenomenon: 1) these languages descend from one proto-language, 2) they have borrowed the pronouns from language, 3) they were forced to create similar pronouns, or 4) the similarity occurred by chance. Since the first explanation is the most reasonable one, pronouns seem to be good evidence to take into consideration (Ross, 2005).

Ross is not the first person who used pronouns for grouping languages; Greenberg (1971) used them before. However, there is a significant difference between the two methodologies. Greenberg classifies two languages as sisters if the form of their pronouns contains [n], while Ross considers them sisters if they share two or more forms from the reconstructed paradigm (Ross, 2005). In his methodology, Ross relies on three procedures. First, he identifies both the free and bound pronouns. Then, he reconstructs the proto-form of these pronouns ignoring any suffixes attached to them. Finally, he reconstructs the pronouns for the ancestral language. Consider the following table, which presents the reconstructed pronoun set for the Proto-Trans New Guinea (Ross, 2005).

According to Ross (2005), if languages reflect two or more of the forms in the table, they will be assigned to the Trans-New Guinea family. This methodology categorises 605 Papuan languages into 23 families.

As will be explained in the following section, in this paper, we support a different methodology to establish genetic relationships. That is not to say that this method should be perceived as an alternative to the comparative method; rather, this method could be more practical in cases where the comparative method is not applicable.

### 5. Typology in historical linguistics

In this paper, I argue that the typological method, with some restrictions, could be an effective way to establish genetic relationships among languages when the comparative method cannot be applied. However, before we proceed, we must first define what we mean by typology.

*Linguistic typology* is a feature-based classification, and like several linguistic terms, typology is borrowed from another scientific field. It was probably borrowed from biology around the nineteenth century (Croft, 2003). Initially, *linguistic typology* was concerned with morphology.
only to the point that the term typological morphology was used to address morphological classification as opposed to genealogical classification (Greenberg, 1974). The aim of such a classification was classifying languages into one of the following three groups: fusional, agglutinative or isolating. The language is classified as fusional when it has words that consist of several morphemes and the boundaries between these morphemes are not clear. It is classified as agglutinative when it has words that also consist of more than one morpheme but the boundaries between them are clear. In contrast, the language is classified as isolating when each word represents one morpheme only (Shopen, 2007).

Currently, typology is used more broadly. It is a field of study that investigates similarities as well as differences among languages, and classifies them accordingly, in order to come up with a generalisation that captures what is either possible or impossible in human languages (Song, 2001; Croft, 2003; Velupillai, 2012). In this vein, typologists do not investigate languages as a whole; instead, they concentrate on a specific phenomenon, or perhaps phenomena, across different languages. Then, languages are sorted based on their differences and similarities in regard to the phenomenon in question. However, because languages might be alike if they happen to be in contact with each other, or because they are genetically related, typologists consider languages from different geographical areas and different language families. This is the key in this paper. That is to say, if related languages are excluded in any meaningful typological classifications because their genetic relationships can make them typologically alike, we can infer that languages that are typologically alike might be genetically related. The question then is what kind of typological similarities can be considered indications of potential genetic relationships? For example, world languages are categorised into six groups based on their basic word order. These groups are SOV, SVO, VSO, VOS, OSV and OVS. However, one can never say SOV languages might be genetically related as this type of typological similarity is not unique. Logically speaking, these six word orders are the only possibilities languages can have. A typological similarity between two languages should not be taken into consideration unless that similarity is conspicuous. Consider, for instance, what Aronoff and Fudeman (2011) call root and pattern morphology. This phenomenon is cross-linguistically rare and has only been found among the Semitic languages. In the following example, we demonstrate this Semitic characteristic in Arabic.

Roots in Arabic usually consist of three consonants; however, roots of four and five consonants do exist. Unlike most of the world languages, derivation in Semitic languages involves a change in vowels and syllabic patterns where the consonantal frame stays stable (Aronoff & Fudeman, 2011). The pattern of deriving agents and patients, for example, is different based on the number of consonant roots they have. The different pattern can be seen in the following table. Please note that the roots in (1), (2) and

3 Dryer innovates a new method to control for the two effects; see Dryer (1989) for more details on this.

4 Arguably, a seventh category can be added, namely free word order.
(3) are triconsonantal, but in (4) and (5) they are not. 5

**Triconsonantal roots**

<table>
<thead>
<tr>
<th>Root</th>
<th>Verb</th>
<th>Agent</th>
<th>Patient</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCC</td>
<td>CaCa</td>
<td>CaCaC</td>
<td>CaCCuC</td>
</tr>
<tr>
<td>ktb</td>
<td>kataba</td>
<td>kātib</td>
<td>makūb. ‘write’</td>
</tr>
<tr>
<td>drs</td>
<td>darasa</td>
<td>dāris</td>
<td>madrās ‘study’</td>
</tr>
<tr>
<td>ksr</td>
<td>kasar</td>
<td>kāsir</td>
<td>makṣūr. ‘break’</td>
</tr>
</tbody>
</table>

More than three consonants

<table>
<thead>
<tr>
<th>CCC</th>
<th>CaCCaC</th>
<th>CuCaC</th>
<th>CuCaCCaC</th>
</tr>
</thead>
<tbody>
<tr>
<td>zlzl</td>
<td>zalzala</td>
<td>muzalzil</td>
<td>muzalzal ‘shake’</td>
</tr>
<tr>
<td>bʃ̣øɾ</td>
<td>baʃ̣øara</td>
<td>mubaʃ̣øir</td>
<td>mubaʃ̣øar. ‘disarrange’</td>
</tr>
</tbody>
</table>

The best way to perceive how root and pattern morphology works is to think of the pattern as a template. That is to say, when an Arabic speaker wants, for example, to derive an agent from a triconsonantal root, he or she recalls the appropriate template for it and superimposes the three consonants, in their original order, on that template. The agent *writer*, for example, is derived from a triconsonantal root; thus, the appropriate template in this case is (_ā__i__). After recalling this template, the speaker superimposes the three consonants *kbt* with respect to their original order in the root, and the result is *kātib*. This way of deriving an agent is not cross-linguistically common. Therefore, one might assume that any languages have it could be genetically related.

We have seen that root and pattern morphology could be used as an indicator that Semitic languages are genetically related. However, if this is the case in Semitic languages, what about others? In other words, what are the linguistic features a historical linguist can consider in such cases? There is no straightforward answer to this question as the answer would vary from one linguistic area to another. That is, what is common in one linguistic area might be rare in another. The researcher, then, should look for a rare linguistic feature occurring among the languages he or she wishes to investigate. Take click consonants as an example. They are unique sounds that seem to occur mostly among African languages, and ‘the sole report outside Africa of a language using clicks involves the special case of Damin, a ritual vocabulary of the Lardil of northern Queensland, Australia’ (Traill, 2015). In this case, such a feature seems to be a good start to investigate possible genetic relationships among different languages. Finally, it seems appropriate to emphasise that these typological similarities should not be taken into consideration unless the comparative method cannot be applied for some reason. Otherwise, the comparative method should be the first step to be used, whereas typological similarities might be considered as a second one to support the results already reached.

1. Conclusions

The comparative method has been used traditionally in establishing genetic relationships among different languages. It starts with assembling 100–200 words (basic vocabulary), working out systematic sound correspondence and then reconstructing the proto-forms for these words. This method, however, seems to be unapplicable in some cases.

---

5 Every Arabic example in this paper is based on personal knowledge of the language.
It requires access to a certain amount of data, which is not always available. The Papuan languages represent one of these cases where the available data for study is limited. To deal with these difficulties, Foley (1986) and Ross (2005) suggest different methodologies to be applied. In this paper, however, we suggest another one, and it is worth mentioning that the three methods can be used simultaneously to support each other as there should be no conflict between them. It should be borne in mind, though, that these three methods should not be used unless the comparative method cannot be applied for some other reason; otherwise, they can be used to support previous results.

We argue in this paper that distinct typological similarities among two languages could be a good evidence that may indicate these languages are genetically related. Root and Pattern Morphology could be an example of such cases because it is considered to be a unique feature that observed among Semitic languages only. Future studies, however, might focus on investigating and highlighting other distinguished typological similarities. Any linguistic area is eligible to be studied, i.e., phonology, morphology and syntax. The crucial factor is how rare the considered linguistic phenomenon is. The more cross-linguistically rare the considered phenomenon is, the more accurate the results should be.

Acknowledgement
This project was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, under grant no. G: 360-973-1442. The authors, therefore, acknowledge with thanks DSR for technical and financial support.

Principal Investigator
Dr. Mohammed Alluhaybi

References
وجهة نظر أخرى في علم اللغة التاريخي

د. محمد مقبل اللهيبي

أستاذ مساعد بمعهد اللغة العربية للناطقين بغيرها

جدة - جامعة الملك عبد العزيز - المملكة العربية السعودية

مستخلص، في علم اللغة، يستخدم المنهج المقارن (comparative method) لمعرفة العلاقات التاريخية بين اللغات المختلفة وردها لغة واحدة، تُسمى عادة باللغة الأم. وقد أثبت هذا المنهج دقة ونجاحاً في الكثير من الأحيان لا سيما في رده اللغات الهندية الأوروبية لأصلها وجمعها تحت مظلة عائلة واحدة. لكن يبدو أن هذا المنهج لا يمكن تطبيقه على بعض اللغات، وفي هذا البحث، يعرض الباحث لمنهج اللسانيات المقارنة أولاً ثم يوضح آليته العمل به وكيفية تطبيقه ثم يناقش بعض الأمثلة لشرح كيف أن هذا المنهج غير قابل للتطبيق في بعض الحالات، وأخيراً يختم البحث بالحديث عن علم اللغة التقابلي وشرح منهجه، وكيف يمكن الإفادة من هذا المنهج للتوصل إلى العلاقات التاريخية بين اللغات المختلفة التي لا يمكن استخدام المنهج المقارن معها، وتؤكد العلاقات التاريخية بين اللغات التي سبق أن تم استخدامها معها.

الكلمات المفتاحية: المنهج المقارن، علم اللغة التاريخي، الأسر اللغوية