

Received 8 September 2015.

Accepted 19 December 2015.

## VARIATION, SHIFTING AND IDENTITY IN ARABIC

Hassan R. ABDEL-JAWAD

Sultan Qaboos University (The Sultanate of Oman)\*

hshuqair@squ.edu.om

### Abstract

This paper examines different ideological and identity construction functions of linguistic variation and shifting in Arabic through the analysis of the existing literature on language, politics and national identity, arguing that linguistic variation and shifting may be used for constructing/reconstructing, locating/relocating and/or shifting/abandoning identity, and divided/split identity. A few major cases will be surveyed to exemplify this. It is shown that variation and/or shifting in Arabic may reflect different levels of sociopolitical, ethnic, sectarian, and religious grouping/ divisions. In such contexts, language and/or varieties of language serve as markers of identity and as boundary-setters between groups.

### Keywords

variation, shifting, identity, boundary setting, nationalism

## VARIACIÓN, CAMBIO Y IDENTIDAD EN ÁRABE

### Resumen

Este artículo examina diferentes funciones ideológicas y de construcción de la identidad de la variación lingüística y de cambio en árabe a través del análisis de la bibliografía existente sobre el lenguaje, la política y la identidad nacional, argumentando que la variación lingüística y el cambio pueden ser utilizados para construir/reconstruir, localizar/relocalizar y/o cambiar/abandonar la identidad, y dividir/escindir la identidad. Se encuestarán algunos casos importantes se encuestarán para ejemplificar lo citado. Se demuestra que la variación y/o el cambio en árabe pueden reflejar diferentes niveles de agrupaciones/divisiones sociopolíticas, étnicas, sectarias y religiosas. En tales contextos, la lengua y/o las variedades lingüísticas sirven como marcadores de identidad y como creadores de fronteras entre grupos.

---

\* Department of English, College of Arts and Social Science, Sultan Qaboos University (SQU), Al-Khoudh, PC 132, The Sultanate of Oman.

**Palabras clave**

variación, cambio, identidad, creación de fronteras, nacionalismo

**1. Introduction**

Variation studies of Arabic have frequently focused their attention on studying the correlation between language variation and the traditional socio-economic, cultural and stylistic variables such as age, sex, education, background, etc. In such cases usually a quantitative analysis of collected data is performed to establish these correlations. Then the results show patterns of sociolinguistic tendencies in the distribution of the observed variation. There is no doubt that such a process is the first and fundamental step in discovering these patterns of variation in the linguistic behavior of speakers. Without this it will be subjective and impressionistic to make any subsequent conclusions or generalizations or to give any explanations. Having said this, we can proceed to say that in explaining the findings, some form of qualitative analysis is deemed necessary to identify the factors of, or forces behind these different patterns of observed variation, e.g. ideological, political, social, and cultural, etc. Accordingly, the objective of this paper is to survey previous studies of variation in Arabic to offer a qualitative analysis of the significance, meaning, and function of this variation, with special reference to the function of locating identity.

Furthermore, this geographic area (the Arab world) has its own special characteristics, where the political scene intersects with the religious, cultural, social, ethnic, and national and all of which exhibit themselves in the linguistic situation and behavior as our friend Yasir Suleiman (2006) refers to this in his chapter "Charting the nation: Arabic and the politics of identity". This area is different in that political, social, religious, ethnic, and sectarian issues/classifications/demarcations are important issues as they are deep seated in the consciousness and awareness of the individuals as well as groups. They are special as they cause a lot of polarization and pluralization.

In most, if not all Arab countries, the ruling powers starting from the imperial colonizing ones to the regimes that took over from the colonizers carrying the same

policy of their masters: the main policy being “divide to rule” and thus they have been encouraging/fostering these divisions and splits and all types of conflict which are again linguistically realized in different ways. This has given rise to a set of terms such as *?iqliimiyya* ‘regionalism’, *jihawiyyia* ‘localism’, *Ta?ifiyya* ‘sectarianism’, *3unSuriyya* ‘racism’, *3ilmaaniyya* ‘liberalism/secularism vs salafi/religious/theological’; pan-Arab nationalism (Arab nation or *Al watan Al-Arabi* or *al-Umma al-Arabiyya*) vs State nationalism (Omani, Egyptian, Jordanian, etc.).

This has often given rise from time to time to different types of movements, such as the call for the use of *3aammiyya* (colloquial Arabic) rather than *FusHa* (standard Arabic). Two clear cases can be mentioned here: in the first half of the 20<sup>th</sup> century there was Egyptian National Movement: calling for Egyptianization of language and one in Lebanon calling for Lebanonization of language.

Within the same state: political divisions and conflicts are occasionally provoked causing differences, including linguistic signs, e.g. in Jordan: Jordanian vs. Palestinian. And in Oman Zanzibaris vs. Native Omanis (Arabs). The linguistic boundaries in such cases often coincide with political/power relations. Within the same community one may find divided towns, villages, border towns or different religious groups and/or sects. In some cases, these divided towns or regions may either:

a. belong to two different political systems: Iskenderun (between Turkey and Syria) or the case of many Palestinian villages where the so-called truce line in 1948 cut them in the middle, thus dividing them into an Eastern part belonging to Palestine/Arabs and a western part belonging to Israel; e.g. Jerusalem and *Bart’a*, among many.

or

b. are divided between two or more different Arab countries; e.g. Rafah between Palestine and Egypt, *alburaimi/al3ain* between UAE and Oman... where each one adheres to the varieties in their respective countries.

The overlapping cycles of identity may differ in order: pan-Arabism, nationalism with its two or more dimensions (the local/regional vs the Pan-Arab), religious, localism, which makes it difficult to establish the linguistic boundaries as they often overlap. For instance, there is both an overlap and a conflict between the so-called “Wider

nationalism pan-Arabism" (*alqawmiyya*), and the narrower nationalism referring to the specific countries: Egyptian nationalism, Jordanian nationalism, etc. which require different symbols of identity such as a national anthem, flag, currency, etc. Yet each state considers itself as part of *alʿumma il ʿarabiyya* (Arab Umma), *alwaTan alarbi* (Arab Nation), and/or *alʿumma il ʿislamiyya* (Islamic Umma).

In the cases to be presented in this paper, it is hoped to show the complex roles linguistic/language variation plays in the formation of the overlapping, conflicting, and sometimes divided identities and loyalties among its speakers in the Arab world.

Quite recently, there has been an emphasis on the role of identity with its different types in determining language use, variation and shifting (cf. Suleiman 1994, 1996, 2004, 2006). Language with its different varieties has been used by groups to identify with one political, social, ethnic, religious, national, etc. group or another. It is a common practice that Arab speakers use language features from several language varieties in the same discourse in order to gain authority by tapping into specific language ideologies (Stadlbauer 2010: 8). This paper examines different ideological and identity construction functions of linguistic variation through a detailed analysis of the existing literature on language politics and national identity, arguing that linguistic variation may be used for integration/disintegration, identity reconstruction/reconstruction, locating/relocating and/or shifting/losing identity, and divided/split identity together with the different overlapping cycles of identity (national, ethnic, religious, social, tribal, sectarian, etc.). Beside my own analysis, this study is based on several individual studies done independently around the wide theme of language and identity (see bibliography for a list of these studies). In the following sections, patterns of variation and/or shifting corresponding to different forms of identity formation will be presented.

## **2. Case one: Diglossic variation**

Aside from the contextual domain distribution of the two polar varieties of Diglossic Arabic as outlined by Ferguson 1959, and shown in many subsequent studies (Schmidt 1974; Sallam 1980; Husein 1980; Abdel-Jawad 1981; Ibrahim 1986; Al-Khatib

1988; Haeri 1996; Habib 2008, 2010, 2011 to mention just a sample), the use of these varieties may also be determined by the patterns of identity and ideologies speakers desire to associate with each “Speakers use language features from several language varieties in the same discourse in order to gain authority by tapping into specific language ideologies” (Stadlbauer 2010: 8). In this regard, Holt (1996) states that “the selective use of language features from different varieties signals as much information as the propositional content of the message: choosing features from one variety over another is a significant marker indexing the position of the speaker in society, their knowledge of political and religious values, or their aspiration for social mobility. Arabic is seen as an obvious and inevitable choice [...] as it is the language of the Holy Quran and all the Arab States have a Muslim and Arab Majority” (Holt 1996: 11).

However each of the two polar varieties stands for different sets of values:

a) H variety serves as a neutral, corrective /reference and pan-Arab unifying model, aiming to transcend the boundaries of individual nation-states, while at the same time stands for conflicting drives: Religious, political, national, or ethnic, which determine the direction of change, patterns of variation and/or conflict. This is consistent with “Standard Lang. Ideology” (Lipi-Green 1997) which refers to a cluster of beliefs about the value of linguistic homogeneity. Holt (1996: 11) writes that “it is Islam whose sacred language has retained its original form and still become a national and official language”. Similarly, Haeri (2003: 43) states that this H “socializes people into rituals of Islam, affirms their identity as Muslims and connects them to the realm of purity, morality, and God” and therefore attributes of this variety are equated with the moral virtues of the user.

Furthermore, H is Tribally neutral. In a TRIBAL Arab world, where the tribe is a very strong social and political unit, this H variety has been neutral and universal and has no association with tribes or tribal viewpoints. So membership to it is open to all equally. So Arabic carries “no tribal connotations” (Holt 1996: 13).

b) Diachronically speaking, SA is neutral as it is not based on any of the local dialects thus it stands at equal distance from all Arabs. Holt (1996: 21) argues that “In the Arab world one could argue that all are equally distant from the standard and that on-one therefore benefits from a linguistic advantage”.

Holt (1996), Haeri (2003) Suleiman (2004) and Stadlbauer (2010), among others, have shown that ideological forces gave rise to linguistic conflicts which reflect identity conflicts, such as:

*Religious conservatives*, who are often referred to as “purists”, argue for the purity, supremacy and sacredness of H variety and therefore call for the maintenance and use of this variety as a symbol of Muslim Arab history, morality, and identity as an Muslim Umma ( Nation).

In contrast, *pan-Arab nationalists* call for a reformed modified modern and (religiously neutral) form of the H variety of Arabic as a united Arabic language, widely referred to as Modern Standard Arabic (MSA) to be the neutral unifying force of all Arabic-speaking people in the Arab world of all backgrounds, religious denominations, and affiliations.

Furthermore, *modernists* at this era of IT and globalization, tend to advocate increasing use of English in many social domains in order to connect to the international community. They believe that the H variety is a carrier of tradition and religious morals but has not been able to reflect scientific and economic progress so English has to be used as symbolic capital link to the “prosperity” and modernity of the West (Stadlbauer 2010).

At the other extreme, stand the Nationalists (state nationalism) who call for the use of national forms of Arabic for different Arab countries thus they seem as promoting separatism. Two clear cases were active in the first half of the 20<sup>th</sup> century: one in Egypt (Egyptian National Movement: Egyptinization of language) and one in Lebanon. (Lebanese Christian ideas: Lebanonization of language).

c) Power maintenance: Furthermore, SA has often been used by the older forms of power represented in the ruling families to maintain their power. Holt (1996: 20) explains that such families have seen “it in their interests to appeal to the masses in a language with great symbolic function and which gives them greater legitimacy rather than promoting a vernacular populism”. To them, “the emotional and symbolic appeal of religion would be more successful than appealing to the masses on the basis of nationality” (Holt 1996: 20) which may be served by upgrading local varieties into standard ones. This is evident in recent years and events as major political parties,

especially religious-oriented ones, promote the use of standard not as a national language but a pan-language with wider appeal where religious affiliation is seen as more appropriate than tribal or local/ national affiliation.

d) Prestige: All in all, SA has an overt prestige while the non-standard ones have covert prestige.

e) Religious vs. political: It is interesting to note here that SA can stand for at least two major values/drives...: Religious and political. If seen from a religious point of view, it is acceptable by all Arabs and non-Arabs and even minorities/ethnic groups within the Arab countries accept it as a marker of Muslim identity (as in the case of Kurds, Berbers, etc.). However, when it is seen as a political symbol as marker of national (Arab identity) so it is used as the official language in these countries, then it is looked at like any other language and is resisted so Kurds call for the use of their native language as the official one in their areas and so do the Berbers, etc. (process of inclusion vs. exclusion).

Standard Arabic (H) has served *a national function* during colonization. Holt (1996: 17) writes that “the symbolic function of written Arabic and developing it along nationalist lines during the *nahda*...”. It was a symbol of identity in the face of COLONIZATION... as a reaction and defense strategy.

f) On the other hand, all the L (non-standard) varieties stand for local, regional blocs, sub-national, and state values and identities. At the state level, each country sees itself as a nation state having what may be referred to as Sub-national varieties: Each country has its own form of “ARABIC” (though it includes many different dialects), e.g. Omani, Jordanian, Egyptian or Tunisian etc. Arabic. These are not geographical but territorial, political and therefore territorial Identity markers. Linguistically, there may be an overlap and similarity among them, e.g. northern areas in Jordan are similar to those in the southern parts of Syria, Northern Palestinian is closer to southern Lebanon than it is to Jerusalem one... but they are referred to as Jordanian, Palestinian, Syrian, or Lebanese Arabic. SA is not usually used for local/regional identification because it has no native speakers as it is seen as a pan-Arab variety. It is interesting to add that even names of such countries may indicate the kind of identity they desire to emphasize. E.g. Syrian Arab Republic, Kingdom of Saudi Arabia, or Hashemite Kingdom of Jordan, This

division can be at a regional geographic level (blocs) as in the Gulf (Gulf council countries vs. non-Gulf or *al-itti7aad al magharibi* (Maghribi (western) union).

An examination the names given to countries, papers and other entities in the Arab world shows these tendencies to relocate identity. Egypt was called, 'the United Arab Republic' during Naser's period to refer to a larger Arab entity but it was later changed to 'Egypt Arab Republic' limiting it to a narrower domain, indicating departing the unity dream. Similarly, the terms and names of newspapers, media, etc. are symbolic and indicate a wider scope of identification: *al-sharq al-awsat* Paper (Middle East Paper), *Al-Quds al-Arabi* paper (*Arab Jerusalem paper*), *The voice of Arabs* (radio Egypt during Naser's period), *Al-Jazeera channel*, and so on.

In conclusion, I would like to cite Holt (1996) who maintains that the

linguistic identity in the Arab world is divided between 2 forms which are both independent of identity with the state..... Fusha with all what it stands for ... pan-Arab identity and 3ammiyya, although possessing a vitality and dynamism not echoed in Fusha, leads to an identity which runs deep and is regionally based, and yet politically and historically marginalized (Holt 1996: 23).

Shifting from one variety to the other serves these different identities and values. Shifting to SA is a shift to formality and distance while a shift to non-standard variety is a shift to informality, solidarity, in-group membership: sub-national vs. national (pan)culture. This seems to echo Labov (1966) who maintains that the use of non-standard features is controlled by the norms of vernacular subculture, whilst the use of standard features is controlled by the overt norms of the mainstream culture in society.

### **3. Case two: divided loyalty-political dislocation**

Language is politically significant and is often used as a tool of penetration, participating, legitimacy, identity construction, and integration. Politically divided communities such as the case of many boundary towns: e.g. several Palestinian towns (*Barta3a*), *Buraimi/Ala3ain* between Oman and UAE, *Rafah* (Egypt and Gaza) may exhibit



this political division linguistically. Such political boundaries may set the linguistic boundaries within the same community:

a) Rafah which was originally one town is now divided politically: the Eastern side is with Gaza while the Western side is with Egypt. This political division seems to reflect linguistically where in the Eastern part speakers tend to use a local Palestinian variety while in the Western side they speak an Egyptian variety and often adopt a regional Egyptian variety as their social standard.

b) Alburaimi is an Omani border town sharing a front line with its corresponding UAE town of *Al3ain* which were originally one town with the same families. However, the Omanis in Alburaimi believe (at least psychologically) that they use a Omani variety to mark themselves as Omanis, while on the UAE side, speakers tend to adopt a UAE variety to identify themselves as UAE. However, in case of social identification, the Omanis in Alburaimi tend to adopt linguistic features from *Al3ain* variety as they believe that it is more socially prestigious.

c) The third case (The case of Barta3a) deserves more attention as it sets a more relevant example

In a sociolinguistic study of the situation in a divided Palestinian Village (*Barta'a*), (Amara & Spolsky 1996) presented evidence of a "growing double identity" reflecting socioeconomic, political and religious factors.

*Barta3a* is a Palestinian village which was divided following the 1948 Israeli taking over of Palestine where they set the border line in the middle of the village, thus dividing it into the Western part, annexed to Israel, and Eastern part, annexed to the West Bank of Jordan at the time. Accordingly, the two parts started developing economically, socially, politically and probably linguistically differently: one reflecting the Israeli system while the other the Jordanian Arab system. What deepened the split is the total political blockade where inhabitants of the two parts, in many cases members of divided families, were not allowed any contact of any sort and they were banned from establishing any links under legal punishment, i.e. there had been geographic and familial proximity but total separation and change of directions. In 1967, following the 6-day war, the West Bank fell under Israeli occupation. The official boundary between the

two parts was removed and members were allowed to reunite. Yet, the two parts are still under two different jurisdictions.

d) Questions to be raised here:

-What is the impact of dividing the village for 20 years?

-What is the impact of reunifying the village after the removal of the official boundary?

Does this result in the removal of social and political divisions between the two halves of the village and so the “linguistic frontier” resulting from parallel social, political and cultural differences” continues between them.

What is the impact of rising political, religious, etc. movements such as the Intifada, feeling of abandonment, changing of attitudes with the resurgence of Islamic movements.

e) How does this reflect linguistically? Do the Political and economic dislocation result in Linguistic variation?

According to (Amara & Spolsky 1996) there seems to be a number of Processes OF IDENTITY at work:

*Israelization*: associated with modernization, urbanization and official authority. Are the Arabs in Israel undergoing a process of Israelization? (integrating or not).

*Palestinianization*: associated with the local, national, and cultural values and heritage: do they still identify with Palestinians (Palestinianization?).

*Islamization*: associated with religious values and the resurgence of Islamic movements. Accordingly, a third trend of identity is competing with the first two: i.e. Islamization.

*Alienation / exclusion*: recently, there has been changing in political attitudes and intentions where there have been calls to maintain “a Pure Jewish State” thus excluding all others, including Native Palestine Arabs who start rethinking their political identities.

*Change of attitudes and alternating identities*: Attitudes after 1967, and after Intifadas, the Western part, like all Palestinians west of the green line, were looked at as more modern, richer, and have more movement freedom as they are considered Israeli citizens who have freedom of movement unlike West Bank Palestinians who have always been under severe mobility restrictions.

It was found that linguistic features distinguish each part of the village: lexical variation as many lexical items have been borrowed from Hebrew in the western part of the village. As well as the external linguistic influences may shape the local vernaculars of the two parts: (Arabic, English, and/or Hebrew?). Amara & Spolsky (1996) identify a large amount of linguistic variation between the two halves.

f) For the Western part Arabic is the national native language but is not the major official one since Hebrew is. Both Arabic and Hebrew are formally taught in schools and are informally used in daily interaction. Hebrew being a second language where almost everyone speaks, reads and writes Hebrew.

g) In the Eastern part, Arabic is the only official and formal language with English taught as a foreign language... but Hebrew was rarely taught or spoken pre1967. However after 1967, Hebrew started to be used informally (spoken) by a growing number of laborers, prisoners and those who work with govt. offices who have become fluent in speaking Hebrew.

The authors highlighted several identities by posing the following question to their informants: How do you define yourself, i.e. who are you in terms of:

- Tribal (Hamula)
- Villagers (Rural)
- Palestinian (national)
- Israeli (political)
- Arab (pan-Arabic)
- Muslim (religious)

Accordingly, they established the following trends: older people, esp. born before 1967, show stronger national Arabism (pan-Arab), but the Palestinian identity becomes stronger after 1967 (also following the separation from Jordan and more specifically following the disengagement in 1987 and the establishment of the Palestinian authority)... so in the light of the developing political factors, a lot of locating and dislocating identities have taken place... Disappointment with the Arabs in recent years has led to rethinking of Pan-Arabism. When one group feels abandoned, marginalized, threatened, like Palestinians in Gaza they tend to turn off their alliance with the other

groups. Nowadays the Islamic identity is growing, esp. in light of seeing that Muslim countries like Iran and Turkey support them more than the Arabs... Following their disappointment with their Arab compatriots and the secular PLO leadership they shifted towards Islamic movements...

Conclusion: Political changes lead to changing of identities and loyalties... Sociolinguistic variation (micro) may often reflect this. These identities intersect: which comes first and why? Changing identities depending on political and may be religious associations/considerations.

#### **4. Case three: In-group vs. out-group**

Local/internal Dialects may be in conflict: the case of Palestinian Vs. Jordanian groups in Jordan which shows that “language serves as a marker of identity and as a boundary-setter between the in-group (ourselves) and the out-group (others)” (Suleiman 2004: 7).

There is always “Interaction between language and national/ethnic identity in situations of intra- and intergroup conflict” (Suleiman 2004: 8). In Jordan, like in many other communities, there are patterns of variation other than diglossic: i.e. Jordanian vs. Palestinian linguistic markers.

Three principles govern this pattern according to Suleiman (2004):

First, “the power-language relationship is an important aspect of ideological contestation, and of identity assertion and negotiation in inter-and-intra-group interaction” (p. 13).

Second, language is a marker of identity” (p. 13).

Third, both one and two operate at the communicative and symbolic levels” (p. 14).

What is the Interaction between linguistic choices and political events/situations?

*Language situation* in Jordan is closely related to the political situation and political division: a case of ethno-political situation.

*Linguistic division/boundaries* represent group divisions: to go from one group to another you have to cross over these linguistic boundaries so variation or shifting here

indicates divided political loyalty... (consider some Palestinian families residing in Jordan early in the 20<sup>th</sup> century and how they see themselves as Jordanians or Palestinians and how this is marked linguistically ).

This shows how conflict in the Middle East (whether internal, regional, or...) can alter the linguistic map of a country and the dynamics of the sociopolitical evaluation...

Power and conflict are important factors in shaping the linguistic situation (same pattern applies to Bahrain (Cf. Holes 1980, 1983a 1983b, 1986a and 1986b).

The use of (Q) variants has been particularly motivated by / or motivate the use of ethno linguistic labels (*beljiki*) “ (used to refer to a Palestinian in Jordan) as a boundary setting label” which is similar to ethnic group labels such as Bellushis, Zadjalis, Druze, Alawaites, etc. regardless of their background origin to distinguish them from the rest. Such terms have become “identity laden labels” (Suleiman 2004) which mark Insiders vs. outsiders, power imbalance, not equal share of power and employment, dominant group vs. dominated group or superordinate group vs. subordinate one

This is often accompanied by other symbols/artifacts of identity: Red-checked vs. black-checked head dress (Kufiyya vs. Hatta).

In many cases, this variation may distinguish the population in the same country as “nationals vs. naturalized” which has also given rise to a number of dividing terms such as “asli (original) vs. “naqli” (not original... borrowing terms used to describe spare parts), “mulHaq (annexed)” or even using electrical terms such 220 vs. 110 (220 referring to originals while 110 referring to incoming groups (naturalized)”. This sociopolitical pattern of VARIATION corresponds to social stratification: first-class citizen, second-class with whatever linguistic and cultural markers associated with it.

In such communities (multi-ethnic) linguistic/cultural accommodation must be at play: it relates variation to ethnic/national identity in intergroup relations. At least three trends can be identified in this regard, which reflect patterns of identity as well:

a. Convergence: approval of the interlocutor and may be for gaining social and/or political acceptance (where outsiders converge to the insiders’ norms, thus hiding their original identity and adopting the so-called insiders).

b. Divergence: exaggeration of the difference (where the so-called outsiders maintain their marking linguistic as well cultural norms and even exaggerate using them to emphasize their group identity and challenge the others).

c. Maintenance: (no-convergence) each side maintains its own features (without any challenge or exaggeration involved) where tolerance governs the situation (as in Oman where groups are free to use their varieties or language tolerating each other but all must identify with Oman as a covering umbrella).

In Jordan, the [g] variant (of the voiceless uvula stop (q)) in Jordan has been established as the norm of the in-group, ruling and politically powerful group which has the full control, and so it is a symbol of POWER and STATUS. Switching to this variant may be integral or instrumental. Switching to [g] is considered by Suleiman (2004: 131) as related “to the formation of nation-state in Jordan”. Jordanizing Jordan and de-Palestinization of Jordan This has been officially marked as Late king Hussein used a dividing term (with historical religious significance) to describe the Palestinian-Jordanian duality: Muhajiriin (immigrants... outsiders) vs. Ansaar (hosting group).

With these political and group divisions, one has to consider Vitality (ethno linguistic validity) in this context where variables can be

- a. Status variables
- b. Demographic variables
- c. Institutionally-supported variables

In such communities, in any political tension, each group resorts to its identity markers (linguistic and otherwise): East Jordanian sees himself as the majority, host, employer, ruling, while the Palestinian is a guest, annexed and not original. This has given rise to a number of political linguistic slogans such: Jordanians of all origins (*min shatta il-manaabiti wal-?uSool*), Jordan First (*Al-?urdu ?awwalan*), Jordan for Jordanians (unity slogan) (*al-?urdun lil-?uroniyyiin*) and so on. Thus one can refer to covert/overt policies of inclusion vs. exclusion as shown in the following points:

a. Political demographic classification which is linguistically realized (though all were under same jurisdiction up to 1987).

- b. Pure Jordanians who have full citizenships with all it entails (national #).

c. Jordanian of Palestinian origin which make up two groups: those who hold Israeli/Palestinian ID (believed to have two Identities) who have full rights except a very few technical ones. This group is marked by being given special yellow Cards (referred to as Bridge statistical card).

d. The third group is pure Palestinian who are seen now as foreigners but with some Jordanian links and so they are given a special passport (with no national number) and their bridge cards are green not yellow, while a blue card is given to Palestinians from Jerusalem.

e. In this context, identity controversial questions such as: Who is the Jordanian? How to distinguish oneself accordingly? Are often raised.

f. Furthermore, such complicated context also causes what may be termed as "identity conflict". For instance, how do the early settlers of Jordan of Palestinian origin (moved to East Jordan early in the 20<sup>th</sup> century) they see or identify themselves? How do they use lg. to identify with this side or that side and what accommodation processes they follow?

g. Jordanian women who live in big cities like Amman and Irbid normally use the urban variety (which is believed to have Palestinian origins) but switch to pure Jordanian variety in case of tension... (divided loyalty: covert prestige vs. overt prestige).

Conclusion: This is a clear case of SOCIO POLITICAL PATTERN OF VARIATION where linguistic variation serves as a marker of identity and as a boundary-setter between the "in-group (ourselves) and the out-group (others)" (Suleiman 2004: 7).

## **5. Religious identification and division: the case of Bethlehem**

Different religious groups may use language variation or shifting as a way to change identity and affiliation. In a study done by Amara (2005), he found that among Christians in the city of Bethlehem who used to be the dominant group:

Younger women and some Christian men are tending to adopt an urban pronunciation like that of nearby East Jerusalem, at the same time as the speech of younger educated Muslims is showing the growing influence of the standard

variety, of Arabic. By relating the use of linguistic variants to changes in identity, this study shows that Bethlehem is a town in transition, being transformed from its previous status as a Christian Arab town into an important Palestinian and dominantly Muslim city (Amara 2005: Abstract).

“By relating the use of linguistic variants to changes in identity, this study shows that Bethlehem is a town in transition, being transformed from its previous status as a Christian town into a Muslim one” (Amara 2005: Abstract). Bethlehem had been widely seen as a Christian town but now there is a transition into a Muslim town. Here, linguistic variation is shown to have produced new distinctions. A case of a town in transition, being transformed from its previous status as a Christian Arab town into an important Palestinian and dominantly Muslim city. This also provides a unique case for testing shifting of social identities: non-urban to urban. It illustrates how urbanization and migration are reflected in the sociolinguistic changes of the Arabic spoken in the town. One of the most salient linguistic features of this is the use of the variants of (q). Whereas most residents formerly used the common (k) Palestinian variety for the standard /q/, similar to that spoken in Palestinian villages, younger women and some Christian men tend to adopt an urban pronunciation like that of nearby East Jerusalem, while, at the same time, the speech of younger educated Muslims is showing the growing influence of the standard variety of Arabic. This linguistic behavior may be seen as a symbol of religious identification which may be traced in other Arab communities as well (cf. Blanc 1964; Holes 1983a).

## **6. The case of returning-immigrants variation: relocating identity; e.g. Zanzibaris in Oman**

Many Arab Omanis and Omani families immigrated to Zanzibar over a long period of time in the last few centuries where they established themselves as the ruling and upper class. Then in the 60s of last Century they were forced to leave the country following a local revolution by the native Africans. Few of these (returnees/re-immigrants) identify with what they consider as native country; yet many still face what



may be seen as “identity conflict” or “divided loyalty”: whether they see themselves as Omanis returning home or Zanzibari Omanis immigrating to hosting Oman? This is also linguistically marked through the variant use of English-Arabic and Swahili.

Kharusi (2012: 13) explains that “language choice among Swahili-speaking Zanzibaris is very much influenced by the individual’s perception and interpretation of the label Zanzibaris, which in turn determines their acceptance or rejection of it”. Accordingly, Kharusi identified three different subgroups of Zanzibaris: those who continue using Swahili in both public and private domains; those who use it only in private domains; and those who do not use or refuse to use Swahili in either domain and even they pretend they do not know the language.

This explains the different patterns of language shifting and use among them: the first group continue to use Swahili to the exclusion of Arabic, the second group is still using Swahili predominantly with a weaker version of Arabic though they have been in the country for most of their lives, while the third group has shifted to Arabic with the exclusion of Swahili. This pattern of linguistic variation is closely tied with the issue of identity for them. Many see themselves as pure Omanis descending from Omani tribes which maintain their existence in Oman and so they have integrated with these tribes; others feel that they have lost their tribal roots as a result of population mix and intermarriages so they have not fully integrated within the rest.

Psychologically, they have developed different attitudes. When they started returning to the country back in the last century, they were not well received by the people and they were even looked down on. Then at later stages when they started taking higher jobs because they were more educated and have more command of English, they started to feel superior to the others. It can be assumed in this case that their language behavior is one way of distinguishing themselves from the rest: i.e. a way of distinction and probably superiority (changing of attitudes). Psychologically, Kharusi (2012) explains this saying that the first group, which belongs to the upper class, uses Swahili predominantly as they do not feel that their identity is threatened. However,

[m]embers of the second subgroup, whose use of the language is domain dependent, associate themselves with the label only when they perceive the

context in which it is used as being unthreatening to their 'Arabness'. Among other Zinjibaris, individuals in the second subgroup readily identify themselves with the group, for it is within this context that they interpret the term to represent shared positive cultural values and a means of fostering solidarity. However, in the presence of non-Zinjibaris, they are concerned that the label might be used as a challenge to their Arab identity and, consequently, they might be disinclined to use Swahili in public (Kharusi 2012: 13-14).

As for the third group, members of this group reject to be identified as Zanzibaris and accordingly they totally shift to Arabic and abandon Swahili which they see as a marker of "inferiority". This group division according to Kharusi (2012: 16) "this collocates with social boundaries too: higher class, middle class and lower class respectively".

At all levels, the language issue is still at the center of this identity conflict for these returning or immigrating groups.

## **7. Conclusion**

It is evident from these case studies that variation and shifting in Arabic in its different forms and manifestations fulfills different ideological and identity construction functions. It may be used for constructing/reconstructing, locating/ relocating and/or shifting/losing identity, and divided/split identity. At the diglossic level, each of the two polar varieties stands for different sets of values: SA can stand for at least two major values/drives and identities political and religious. It stands at an equal distance from all non-standard varieties, acting as a corrective and — reference model — umbrella. Non-standard varieties on the other hand stand for Sub-national, territorial, political identity markers. It may serve, on the one hand, to express unity, pan-Arabism, nationhood (umma), and Islam, and on the other hand, it may mark localism, regionalism, or sub-nationalism.

Within each state, variation may reflect different levels of sociopolitical, ethnic, sectarian, and religious grouping/ divisions. In such contexts, Language serves as a

marker of identity and as a “boundary-setter between groups” (the case of Jordanians and Palestinians) where linguistic boundaries correspond to political divisions. Accordingly, with any political tension or conflict, each group resorts to its identity markers, including language, where in some cases language variation can save or cost life. Ultimately, political changes lead to changing of identities and loyalties. Religiously, Different religious groups may use Language variation as a way to relocate identity (the case of Bethlehem). Linguistic variation may also reflect political and economic dislocation and divided loyalties. Politically divided communities may present evidence of “growing double identities” which correspond to linguistic divisions. Yet, further complications of the picture will ultimately lead to start reconstructing of new political identities (the case of *Barta3a*). Finally the case of “The Returning-Immigrants variation” shows “identity conflict” or “divided loyalty”.

## References

- ABDEL-JAWAD, H. (1981) *Lexical and phonological variation in spoken Arabic in Amman*, Unpublished PhD dissertation, University of Pennsylvania.
- AL-KHATIB, M. (1988) *Change in an Expanding Context: A Case Study of Irbid City, Jordan*, unpublished PhD Thesis, University of Durham.
- AMARA, M. & B. SPOLSKY (1996) “The Construction of identity in a divided Palestinian village: Sociolinguistic Evidence”, in Yasir Suleiman (ed.), *Language and Identity in the Middle East and North Africa*, London: Curzon, 81-99.
- AMARA, M. (2005) “Language, migration, and urbanization: the Case of Bethlehem”, *Linguistics*, 43, 883-902.
- BLANC, H. (1964) *Communal Dialects in Baghdad*, Cambridge-Massachusetts: Harvard University Press.
- FERGUSON, C.A. (1959) “Diglossia”, *Word*, 15, 325-340.
- HABIB, R. (2008) “Humor and disagreement: identity construction and cross-cultural enrichment”, *Journal of Pragmatics*, 40 (6), 1117-1145.
- HABIB, R. (2010) “Rural Migration and Language Variation in Hims, Syria”, *SKY Journal of Linguistics*, 23, 61-99.

- HABIB, R. (2011) "Meaningful Variation and Bidirectional Change in Rural Child and Adolescent Language", *University of Pennsylvania Working Papers in Linguistics*, vol. 17: 2, Selected Papers from New Ways of Analyzing Variation 39 conferences, 81-90.
- HAERI, N. (1996) *The Sociolinguistic Market of Cairo: Gender, Class and Education*, London and New York: Kegan Paul International.
- HAERI, N. (2003) *Sacred Language, Ordinary People: Dilemmas of Culture and Politics in Egypt*, New York: Palgrave Macmillan.
- HOLES, C. (1980) "Phonological variation in Bahraini Arabic: the [j] and [y] allophones of /j/", *Zeitschrift fur arabische Linguistik*, 4, 72-89.
- HOLES, C. (1983a) "Patterns of communal language variation in Bahrain", *Language in Society*, 12, 433-457.
- HOLES, C. (1983b) "Bahraini dialects: sectarian dialects and the sedentary/nomadic split", *Zeitschrift fur arabische Linguistik*, 10, 7-37.
- HOLES, C. (1986a) "The Social motivation for phonological convergence in three Arabic dialects", *International Journal of the Sociology of Language*, 61, 33-51.
- HOLES, C. (1986b) "Communicative function of pronominal variation in Bahraini Arabic", *Anthropological Linguistics*, 28/1, 10-30.
- HOLT, M. (1996) "Divided Loyalty: Language and ethnic identity in the Arab World", in Y. Suleiman (ed.), *Language and Identity in the Middle East and North America*, London: Curzon: 11-25.
- HUSEIN, R. (1980) *The Case Of Triglossia in Arabic with Special Emphasis on Jordan*, Unpublished PhD Thesis, University of Buffalo.
- IBRAHIM, M. (1986) "Standard and Prestige Language: A Problem in Arabic Sociolinguistics", *Anthropological Linguistics*, 28, 115-126.
- KHARUSI, Nafla S. (2012) "The ethnic label *Zinjibari*: Politics and language choice implications among Swahili speakers in Oman", *Ethnicities*, 12 (3), 335-353 <The online version of this article can be found at: <http://etn.sagepub.com/>> (accessed 1 September 2015)
- LABOV, W. (1966) *The Social stratification of English in New York City*, Washington, DC: Center for Applied Linguistics.
- LIPPI-GREEN, R. (1997) *English with an accent: Language, ideology, and discrimination in the United States*, London and New York: Routledge
- SALLAM, A. (1980) "Phonological variation in educated spoken Arabic: a study of the uvular and related plosive types", *Bulletin of the school of Oriental and African studies*, XLIII, 1, 77-100.

- SCHMIDT, R. W. (1974) *Sociolinguistic variation in Spoken Egyptian Arabic. A re-examination of the concept of diglossia*, PhD Dissertation, Georgetown University, Washington, DC.
- STADLBAUER, S. (2010) "Language Ideologies in the Arabic Diglossia of Egypt", *Colorado Research in Linguistics*, 22, 1-19.
- SULEIMAN, Y. (1994) "Nationalism and the Arabic Language: an Historical Overview", in Y. Suleiman (ed.), *Arabic Sociolinguistics: Issues and Perspectives*. Richmond: Cruzon Press, 3-24.
- SULEIMAN, Y. (1996) *Language and Identity in the Middle East and North America*, London: Curzon.
- SULEIMAN, Y. (2004) *A war of Words: Language and Conflict in the Middle East*, Cambridge: Cambridge University Press.
- SULEIMAN, Y. (2006) "Charting the nations: Arabic and Politics of identity", *Annual Review of Applied Linguistics*, 26, 125-148.

Received 8 July 2015.

Accepted 12 October 2015.

## NEW YORUBA IDIOMS AND IDIOMATIC EXPRESSIONS: A NEW MODE OF EXPRESSION IN POLITICAL ARENA

Dayo AKANMU

Adeniran Ogunsanya College of Education\*

dayoakanmu2002@yahoo.com

### Abstract

New idioms and idiomatic expressions, which are modern stock expressions, constitute communicative clogs in Yorùbá routine discourses because of their semantic complexity and deviant nature. Existing studies have established their scope of usage in Yorùbá music but have hardly addressed their communicative adaptability in politics. This paper investigated issues expressed with these idioms, context and strategies for using them in political arena. This is with a view to establishing their communicative and stylistic relevance in Yorùbá discourses. The paper adopted Mukarovsky's theory of Standard Language because of its capacity to explain the "differentia specifica" between the language of everyday interaction and literary language. Data were collected from different routine communicative discourses of politicians and party members on radio, during political rallies or campaigns and were subjected to pragmatic and stylistic analysis. Political issues were expressed using new idioms and idiomatic expressions in the following sociopolitical contexts: *Eté* for corruption/stealing; *Ojúyọbọ* for under achiever; *Yóó wọlé èẹken si* for deceit; *Ó bọjẹ tì* for collective support; *Oyin ni o* for a successful tenure; *Ajímọbi* for collective responsibility; *Nàìjá* for affection/disappointment and *Kárí go/Kárí kọmù* for continuity and discontinuity. Nominalisation, Pidginisation and dialect expressions were the strategies employed for the formation of new idioms in this paper. These strategies were used to express different political happenings contextualized in underachievement, sloganeering for support, deceit, and acceptance/rejection for second term in office respectively. New idioms and idiomatic expressions, used

---

\* School of Languages. Adeniran Ogunsanya College of Education. Otto/Ijanikin P.M.B 007, Festac Town – Lagos, Nigeria.

to express sociopolitical issues in Yorùbá routine communication, occurred in mediated and non-mediated contexts and were conveyed through nominalization, pidginisation, and dialect expressions. These idioms reflect dynamism and modernity-constrained stylistic choices in Yorùbá.

### Keywords

new idioms, Yorùbá, strategies, politics, stylistics

## NUEVOS MODISMOS Y EXPRESIONES IDIOMÁTICAS EN YORUBA: UN NUEVO MODO DE EXPRESIÓN EN POLÍTICA

### Resumen

Los nuevos modismos y expresiones idiomáticas, que son expresiones comunes modernas, constituyen obstrucciones comunicativas en los discursos rutinarios en yoruba debido a su complejidad semántica y naturaleza desviada. Los estudios existentes han establecido su uso en la música en yoruba, pero apenas han abordado su adaptabilidad comunicativa en la política. Este trabajo investiga las expresiones usadas con estos modismos, y el contexto y las estrategias para utilizarlos en el ámbito político. El fin es establecer su relevancia comunicativa y estilística en los discursos en yoruba. El artículo adopta la teoría de la lengua estándar de Mukarovsky debido a su capacidad para explicar la “especificidad diferencial” entre el lenguaje de la interacción cotidiana y el lenguaje literario. Los datos fueron extraídos de diferentes discursos comunicativos rutinarios de políticos y miembros del partido en la radio, durante campañas políticas, y fueron sometidos a análisis pragmático y estilístico. Los temas políticos se expresaron utilizando nuevos modismos y expresiones idiomáticas en los siguientes contextos sociopolíticos: *Eté* para corrupción/robo; *Ojúyọbọ* para inexperto; *Yóó wọlé ẹẹken sí* para el engaño; *Ó bọje tí* para el apoyo colectivo; *Oyin ni o* para un mandato exitoso; *Ajímọbi* para la responsabilidad colectiva; *Nàìjá* para afecto/decepción y *Kárí go/Kárí kọmù* para continuidad y discontinuidad. Expresiones nominalizadas, dialectales y pidgin fueron las estrategias empleadas para la formación de nuevos modismos. Estas estrategias se utilizaron para expresar diferentes acontecimientos políticos contextualizados en el fracaso, consignas para el apoyo, el engaño y la aceptación/rechazo en el segundo mandato en el cargo, respectivamente. Los nuevos modismos y expresiones idiomáticas, utilizados para expresar temas sociopolíticos en la comunicación rutinaria en yoruba, se daban en contextos mediados y no mediados y se transmitían a través de la nominalización, la pidginización y las expresiones dialectales. Estos modismos reflejan el dinamismo y las opciones estilísticas constreñidas por la modernidad en yoruba.

### Palabras clave

nuevos modismos, yoruba, estrategias, política, estilística

## **1. Introduction**

This paper was motivated by the saying ‘the world is a global village’ which imposes new global communicative challenges on the Yorùbá people especially the politicians and their followers. The evolution of new idioms and idiomatic expressions thus becomes a coping strategy to meet the new communicative challenges. Globalisation is a phenomenon that filters into all domain of human experience, especially politics. As a result of this, the Yorùbá, politicians like any other politicians in the Third World countries, are faced with the challenges of how to express their ideas and experiences emanating from globalization. These politicians have a duty to inform, enlighten and educate their people about the latest development brought about by the wind of change blowing across the globe. At a time that Nigeria is trying to develop and sustain a democratic system of government, the Yoruba politicians cannot afford to be aloof to this global development. The politicians must be in the forefront of the urgent need to educate Nigerians about new developments in politics. In doing this, they are incapacitated with certain linguistic constraints imposed on them by their language such as the unavailability of words in their language and difficulties in getting appropriate lexical expressions for their ideas. To overcome this, the choice left for politicians according to Babalọlá (1972), Awóbùlúyì (1992: 26) and Olateju (2005) is to coin new words or idioms and also to attach new meanings to the existing ones. In this paper, such new coinages, words and expressions are hereby referred to as ‘new idioms or expressions’.

By and large, in this paper, attempt was made to address political issues expressed with new idioms and idiomatic expressions in Yorùbá and strategies for using them with a view to establishing their communicative and stylistic potential in Yorùbá discourses.

## **2. New idioms and idiomatic expressions**

New idioms and idiomatic expressions are quintessentially modern stock expressions, formed in response to the needs of the moment arising as a result of the



linguistic constraints of the users which consequently necessitates creation of new words and expressions. They are so called not only because of their complexity which requires the overlapping tendencies of their semantic realizations, but also because of their distinctive difference from old, traditional idioms. As a matter of fact, they straddle metaphor and euphemism and also serve as replication of the established Yorùbá idioms both in form and meaning. They are in spoken and written forms and mainly operate at the lexical and phrasal levels. They are also multi-functional because of their semantic unity. They may equally function as nouns, verbs and adjectives. New idioms and idiomatic expressions are created either through coining, (creating and recreating of new words and expressions) or by investing old words and expressions with new meanings. New idioms are creative, new, strange and fresh by nature. They are not vague expressions and their origins are also known to the users who are always dazzled whenever they come across them in any context.

### **3. Existing studies on new idioms**

Few scholars have turned their attention on the study of new idioms and idiomatic expressions in Yoruba. These scholars include Àkànmú (2003) and Ọlátéjú (2005). Some scholars have described idioms as terms referring to words whose meanings cannot be predicted from the individual elements in them (Babalọlá 1972; Bámgbósé 1975; Owólabí 1976; Awóbùlúyì 1992; and Yusuf 2002). These studies have not directly dealt with some issues raised in this paper but they serve as the springboard for this paper.

While discussing idioms and coinages, Babalọlá (1972) states that neologism is used for expressing new items and ideas brought into Yorùbá culture by foreigners. His view bears direct relevance to this study in that, neologism is one of the linguistic strategies employed in the formation of new idioms.

Bámgbósé (1975) identifies vocabulary-coining strategies to include nominalization, explication, semantic extension, and loan words, among others. Through this, the Yorùbá lexicon is being expanded. Àkànmú's (2003) is not specifically on new idioms but it has some semblance of new idioms in the discussion and examples

cited. He employed a sociolinguistic analytical method for the interpretation of the work which he referred to as coded language used by certain groups of the Yorùbá society. He cites numerous examples of such expressions that are used in different contexts and discusses their functions in daily activities of the users, even though many of the examples cited as slang are not slang but new idioms used by a considerable number of people.

Olátéjú's (2005) work is a remarkable contribution on new idioms, their formation and interpretation. He examines new idioms and idiomatic expressions from both the linguistic and literary points of view. He adopted Chomskyan Transformational Generative Grammar (TGG) while stylistic and some sociolinguistic variables were used to bring out aesthetic and communicative potential of the new idioms and idiomatic expressions.

#### **4. Methodological approach**

Data were collected from different routine communicative discourses of some prominent Yorùbá politicians and parties' members on radio, during political rallies, campaigns, sloganeering in Lagos, Ìbàdàn and Òşogbo. These three cities were chosen because of their urbanized nature and the constant usage of new idioms and idiomatic expressions by the politicians. Data were subjected to pragmatic and stylistic analysis.

#### **5. Standard language vs literary language**

The theory of Standard Language was formulated in 1970 by Jan Mukarousky to solve the problem of relationship between Standard Language (SL) and Literary Language (LL). The suitability of this model lies in the fact that it can be used to explain the differentia specifica between the language of ordinary usage and literary expressions as well as using it to explain and interpret the deviant and inventive nature of new idioms and idiomatic expressions.

According to Mukarovsky (1970), Standard Language (SL) is the language of everyday conversation. Its purpose is the understanding between the speaker and the audience with the aim of enhancing effective communication (Crystal 1997: 68). Language of ordinary discourse is casual and devoid of ornamentation. It is unexamined and uncritical. It does not draw attention to itself or open up provocative questions to the nature of its coding. It is often used in schools, on radio, during political campaigns and preaching. Its priority is effective communication which calls for employment of words and phrases that can be fully comprehended. Therefore, to enhance undistorted communication in Standard Language (SL), language must conform to the entire linguistic norms. Concepts crucial to the Standard Language are 'backgrounding' and automatization illustrated by the example below:

*Làkàsórò* Literal meaning: corrupt coinage from *La casera* drink  
 Idiomatic: No story/end of discussion

The above expression is a phonological creation from the name of a particular non-alcoholic drink in circulation all over the country. The expression '*làkàsérà* (*La casera*) is the standard form that has been foregrounded or de-automatized as '*làkàsórò*'. In other words, *làkàsórò* is the artistic form of *La casera*. There is no expression like '*làkàsórò*' in the standard language. '*Làkàsérà*' is automatized and serves as the background from which '*làkàsórò*' is created and used as an idiom. '*Làkàsórò*' which can be interpreted as 'no story/end of discussion' is made new, thereby becoming element of surprise because it has violated the Yorùbá lexical formation rules.

Although, foregrounding is logically prohibited in Standard Language, this is not to say that it is not visible in Standard Language as Mukarovsky indicates its possibility in journalistic language where it can be used as subordinate to communication in order to draw the readers' attention more closely to the issue at hand or the item of information articulated. This must have justified the use of new idioms and idiomatic expressions in political routine communicative situation discussed in this paper.

## 6. Linguistic strategies for the formation of new idioms

According to Bámbósé (1975), word formation is a universal linguistic concept that is concerned with the study of the patterns on which a language forms new lexical item. As far as formation of new idioms and idiomatic expressions are concerned in this paper, the linguistic strategies involved in their formation include the following:

### i) Nominalisation

Nominalisation is universally known as one of the veritable tools for generating new words in virtually every language. Ruwet (1973: 172) views it as the derivation of a noun phrase from an underlying clause or sentence; or the process of forming a noun from other word classes. Below are some examples of new idioms and idiomatic expressions created from the linguistic process of nominalization.

Idioms	Literal meaning	Idiomatic meaning
oní + àsà + kí + àşà (owner-of-bad-culture)	aláşàkaşà (owner of bad culture)	Oníşokúşò (vulgar person /user of nonsensical language)
oní + ebòlò (owner-of-ebòlò)	Elébòlò (seller of ebòlò vegetable)	A prostitute
má + fo + gótà (do-not-jump-over-the gutter)	mini skirt	indecent dressing

### ii) Pidginisation

According to Nichols (1980), pidgin is a contact language which is an amalgam of linguistic elements of two or more languages and which arises in social and economic transactions between, at least two groups speaking different languages by a process of restriction and simplification of one of the languages of these groups, usually that is in a socially superior position. This process of restriction and simplification is termed Pidginisation. By definition, a pidgin is no one's native language. In the process of becoming notarized, the pidgin undergoes extension and elaboration and thus becomes creolized.

Pidginisation, as used in this paper, is a strategy for loaning words or expressions that are not only pidginized but also foregrounded with idiomatic undertones for communicative purposes, especially by the youths and non-fluent speakers of the language. Examples of such idioms and idiomatic expressions include:

Idioms	Literal meaning	Idiomatic meaning
Vb + Vb or V <sub>1</sub> + V <sub>1</sub> Kárí kọ̀mù - kárí + kọ̀mù = carry come	Carry and come	Loan form of the pidgin English word 'carry come' for: come in/pack.
Kari go - kari + go = carry go	Carry go	Loaned form of the pidgin English word 'carry go' for: move on/ride on/go away
Vb + Noun		
Kari moni - kari + moni = carry money	Put money on the head	Loaned form of the English word 'carry money' for: carry passengers.

### iii) Dialect Expression

Dialect, according to Salami (2006), is a distinct form or variety of a language that is associated with a recognizable regional, social or ethnic groups different from other forms of the language by specific linguistic features such as pronunciation, grammar or vocabulary, especially a variety of speech differing from the standard language or speech pattern of the culture in which it exists. As far as the new idiom is concerned, dialect expression is seen as a dialect variant of word present in the standard language and sometimes used consciously for entertainment. Examples of where dialect are used as idioms and idiomatic expressions include:

Idioms	Literal meaning	Idiomatic meaning
E e ti ri <u>nnken ken</u> (you have not seen anything).	You have only seen little.	Greater show awaits you/that is just a tip of the iceberg.

The underlined word in the above excerpt is borrowed from the Oke-Ogun dialect of Yoruba. In standard Yoruba, the word should have appeared thus:

*E e ti i ri nnkan kan*

(You have not seen anything)

The above expression rendered in Ibarapa dialect is often found in the adverts, addresses and speeches of politicians or even, in music of some artistes and it often has a stylistic significance of comic effect or entertainment.

## **7. Analysis of new Yoruba idioms and idiomatic expressions used in sociopolitical context**

This section is devoted to analyzing and interpreting idioms and idiomatic expressions found in the data collected for this paper. This analysis is to provide insight into communicative and stylistic relevance of the new idioms. Because of the possibility of new idioms and idiomatic expressions possessing many interpretations and the fact that meaning cannot be determined by their individual lexical constituents, this phenomenon is better studied and analyzed from pragmatic and stylistic perspectives with a view to determined their communicative potential in political discourses.

- (i) - *Eté* literal meaning: Yorubanized version of the name *Etteh*  
Idiomatic: thief/looter

The context in which the name emerged is a context of corruption involving a former Speaker of the Nigerian House of Representatives. The expression '*Eté*' as a new idiom is commonly used today by people, especially in the present political dispensation in Nigeria. It is a corrupt coinage of the name *Etteh*. *Eté* is metaphorically used to refer to 'a thief or a looter of public treasury' like the former speaker of the Nigerian House of Representatives who was allegedly accused of corruption and forced to resign on October 31, 2008 over the crisis that emanated from the N628 million building renovation scandal in 2007 before she was discharged and acquitted in a court of law.

- (ii) - 'Ojúyọbọ' Literal meaning: bulging eyes  
 Idiomatic: physical description of an under-achiever

The above is one of the new idioms used to express 'abuse'. The expression which literally means 'bulging eyes' is contextualized in one of the political statements credited to a prominent politician and a one time gubernatorial candidate in Lagos State. In attempt to run down the ruling party in one of his political campaigns in Yaba, he remarks:

*...Ará Èkó kílí ẹ̀ Babiháalà  
 Ẹ̀ sọ́ fún àwọn ojúyọ́bọ́ kí wọ́n kó jíjí wọ́n lọ sí òkè lẹ́hùn ún.  
 (Ray Power, March 2007)*

...Lagosians are not beggars  
 Tell the Ojúyọbọs to carry their eye glasses to the North.

Apparently, The politician's good performance as governor during his first tenure was not recognised by his political opponents. Therefore the opponents employed derogatory terms (idioms) to run him down during electioneering campaigns so that they could have an upper hand. The use of this idiom as shown in the above political context makes interpretation easier especially for people who are already familiar with the political issues around which the context revolved. The name 'ojúyọbọ' is derived from his physical appearance. This politician is endowed with big eyeballs. But suddenly because of some challenges he had during his tenure as governor which many of his supporters blamed on non-release of the local governments allocation funds by the Federal Government, the endowment later turned to the object of ridicule. Since then, the name *ojúyọbọ* has been used as an expression to abuse one's enemy or a political opponent

- (iii) - 'Yó wọlé` ẹ̀ẹ̀kẹ̀n sí' Literal meaning: He would win once again  
 Idiomatic: Deceitful comment (for: never again)

In the above, the expression *yó wọlé è̀èkẹ̀n si* which literally means he would emerge the winner once again; is from a specially waxed record for political advert few years ago for a prominent politician who was bidding for a second term as governor of a State in the south-western Nigeria by Ìyá Aládùkẹ̀, a Dadakúàdà exponent. The song pervaded all the radio and television stations in the south-western part of the country. Since it was generally believed that this politician did not do well enough to get a ticket for the second term by his opponents, the song was scornfully imitated by many people. It was very easy for people to imitate and comprehend not only because of its usage in such a familiar political context, but also because of the dialectal variation ‘è̀èkẹ̀n’ that was skilfully employed to create a subtle humour in the song. The expression è̀èkẹ̀n is from Òkè Ògùn dialect which comprises Şakí, Ìsẹ̀yìn, Baba ọ̀dẹ̀, Şẹ̀pẹ̀tẹ̀rí and some parts of Kwara State, especially Ìlọ̀rìn, where Ìyá Aládùkẹ̀ hails from. The standard form of è̀èkẹ̀n is è̀èkan (once again/more). It is used contemptuously by the opposition party to mean ‘he will never win again’. Apart from its idiomatic connotation, it also has humorous and entertainment effects.

(iv) - ‘Ó bàjẹ̀ tì’	Literal meaning: It would never retrogress or go backward
	Idiomatc: Positive assertion/sloganeering for support, collective action for good governance/ project

The expression ‘ó bàjẹ̀ tì’, literally interpreted as ‘it would not retrogress or go backward’ is a refrain to the political sloganeering of a political party in the south-western part of Nigeria. The governor of a particular State in Nigeria is fond of saying ‘È̀kó ò ní bàjẹ̀ o’ (Lagos will not be backward/destroyed) the response or refrain of which is ‘ó bàjẹ̀ tì’<sup>2</sup>(it would not retrogress or go backward). The governor uses the expression which has become a household idiom in the State, to solicit the support of the people, for the various developmental programmes of his government. In some other States, there are similar political expressions or slogans which have become idiomatic expressions that are meant to sensitise people to the activities of government. In one of these States, during the tenure of the former governor, the sloganeering idiom was ‘Oyin ni o’:



(v) - *Oyin ni o'* Literal meaning: It is honey/shortened of name of Oyinlólá  
Idiomatic: Successful tenure

The expression *Oyin ni o*, is used not only as an appellation, but also as a slogan during the political campaign of a politician whose name and tenure are expected to be as sweet as honey. This politician's tenure as governor was metaphorically predicated as *oyin* (honey) during an election in his State compared to the hardship experienced by the people of the State under the immediate past governor, whose tenure was characterized by mass retrenchment of teachers. As a metaphor, the expression *oyin ni o* (it is honey) means someone whose name and tenure are expected to be as sweet as honey. The expression is used to hail the former Governor of a particular State in Nigeria, who always got excited and entertained on hearing *oyin ni o*. This attribute is interpreted based on the socio-political context and in the socio-political views of some people in that particular State as sweet tenure, soothing tenure or a tenure that brings succour. In other words, the context of its usage always provides information which enables people to give the expression an interpretation of a peaceful and prosperous tenure. In another State today, the sloganeering expression is:

(vi) - *Àjùmòṣe* Literal meaning: collective implementation  
Idiomatic meaning: governance is a collective responsibility

The above is a slogan-turned idiomatic expression used virtually in every home and public place thus:

*'k'Ọyọ́ lè dára àjùmòṣe gbogbo wa ni o!*

(for Oyo to be peaceful and prosperous it is our collective responsibilities)

The word '*Àjùmòṣe*' in the sloganeering idiom is derived from the governor's name '*Ajímòbi* meaning 'together we woke up to know the manner of birth' The idiom, *Àjùmòṣe* (that-which-is-executed-collectively) shares the same derivational strategy with the first three syllable (A/jí/mò) in the name *Ajímòbi* (a person we woke up together to experience his birth) with another connotative meaning of 'collective effort' derived

from the first syllable and vowel (A) in the two names, a first person plural pronoun which eventually expresses 'collective responsibility'. Today, whether on the television or radio, every advert placed by the Oyo State government is ended with *k'Ọyọ́ lè dádá àjùmọ̀sẹ̀ gbogbo wa ni o!* (for Oyo to be peaceful and prosperous, it is our collective responsibilities). The sloganeering has equally become a household expression with the meaning of 'collective responsibility'; even governance is said to be collective responsibility.

(vii) - *Nàìjá* Literal meaning: Yorubanized and shortened of the name Nigeria

Idiomatic: Affectionate way of addressing the country Nigeria exclamatory words of disappointment when something strange happens in the polity.

The expression '*Nàìjá*' is also a new idiom that is frequently used by many Nigerians, especially for entertainment purposes. It has become a way of affectionately referring to the country, Nigeria and Nigerians, especially those abroad. Expression like *ìgbàwo lè ní lọ Nàìjá?* (when are you going to *Nàìjá*?) *'sẹ́ o ti gbọ́ láti Nàìjá?* (have you heard from *Nàìjá*?), '*Nàìjá lodù*' (*Nàìjá* is the real thing) are common expressions among Nigerians, especially during a football match between Nigeria and any other country. Whenever it is used in contexts such as the foregoing, it facilitates information or background that connotes 'an expression of unity' among Nigerians as far as the meaning is concerned. There are also Nigerian musicians, actors, actresses, comedians and youths who have chosen *Nàìjá* as their affectionate way to address their country and sometimes to express exclamatory words of disappointment. When something crazy happens in the polity, people always exclaim *Háà, Nàìjá! Nàìjá* is a created word pronounced in Yorùbá or Pidgin English. It is a name given to Nigeria by Nigerians. It has resonated very well with the educated, semi-educated and the illiterates.

(viii) - *Kárí go/Kárí kọ̀mù* Literal meaning: Go away/move away

Idiomatic: Continue/you are not wanted

These are new idioms loaned from English. The original English words were correctly used in their Yorùbá form '*kárí go*' and '*kárí kòòmù*' under different political contexts to connote different meanings. For instance, during the 2003 gubernatorial election in Nigeria the association of market women in a particular State told the former governor of the State in one of his political rallies to *kárí go*. That is, his candidature is accepted by the people, therefore, he should continue with his ambition to contest as they (the electorate) are solidly behind him. The same year, in another political campaigns/rallies in one of the strongholds of his opponent, he was asked to *kárí go* by the supporters of his opponent. Here, the expression was aided by the body language that connotes 'go away', 'you have overstayed', 'enough is enough', 'you are not wanted any more'. It should be stressed here also that, interpretation of these new idioms are made possible with their various contexts of usage including the extra contextual use of the body language which also aids the understanding of the idiom. In another context, the same candidate was told to '*kárí kòòmù*' by his supporters during one of his political campaigns in 2003. '*Kárí kòòmù*' means come and govern, people want you.

(ix) - *Jẹun sápò*

Literal meaning: Eat into the pocket

Idiomatic: Enrich your pocket

(x) - *Jẹun sókè*

Literal meaning: Eat into the upper stomach

Idiomatic: Move on/go on

Expressions such as the ones above are foregrounded because they are idiomatic. Their meanings cannot be derived from each of their lexical components. Expressions such as these are not allowed in standard Yorùbá because they are semantically deviant. These items are frequently used to describe fraudulent tendencies of some politicians. For instance, in '*Ọmọ Gómìnà*' (Governor's Child), a film produced and directed by Áńtá Láníyan, a character called *Oşùfálé* uses the expression in a discussion with his friend *Akínbò* to portray politicians as fraudulent and corrupt individuals when complaining about shortage of money in circulation thus:

Akínbò: *Oşùfálé, òjẹ o rántí san owó egbé tírẹ fún akòwé egbé kí o tó kúrò nínú ìpàdẹ.*

Oṣùṣàlálé: Kí lò n sọ nígboro ẹnu yí? Níbo ni o ti fẹ kí n rówó? Àfi bí ẹni pé àwọn olóṣèlú ti kó gbogbo owó ilú nílẹ tán. Wọ́n kàn n jẹun sápò ni.

Akínbò: Oṣùṣàlálé, did you remember to pay your own association money to the secretary before you left the meeting?

Oṣùṣàlálé: What exactly are you saying? Where do you expect me to get money? It is like the politicians have packed all the money in circulation. They are just enriching their pockets.

The underlined expression, which literally means ‘eat into the pocket,’ can be interpreted as “our politicians are just enriching themselves with public funds”. This interpretation is arrived at because of the political information expressed in the context which gives clue to the comprehension of the excerpt. In a similar vein, the expression ‘jẹun sókè’ for “move/go on” can be used in different contexts. Let’s consider the example below:

*Òkè ni A.C.N wà nínú ìwé ìdìbò yẹn,*

*Òkè ni kí ẹ jẹun sí*

The A.C.N is located at the uppermost position  
in the ballot paper, eat into the upper stomach.

The excerpt is credited to the incumbent governor of Lagos State, Babátúndé Rájí Fáṣọlá (BRF) during one of the political rallies of his party (Action Congress of Nigeria) held on the 5th of March 2011 at the Tàfáwà Bàléwà Square (TBS) Lagos State. The rally was covered by the Lagos Television (L.T.V) and TV Continental (TVC). The underlined expression is idiomatic and requires socio-political information, as reflected in the context such as the above for a meaningful interpretation. It is abnormal to eat into the upper stomach (jẹun sókè). Eat into the stomach (jẹun síkùn) is the normal way of saying it. However, in this context, the expression is embedded with the socio-political

meaning of 'cast your vote for the A.C.N', whose logo and name appear in the upper part of the ballot paper. The Governor is trying to educate his supporters on where to cast their votes, that is, the position of the party logo on the ballot paper. This idiom elicits laughter. Consequently, the mammoth crowd who found joy, happiness and entertainment in the use of the expression 'òkè ni kí ẹ jẹun sí' (eat into the upper stomach) could not hide their feelings as Fàṣọlá himself was hailed and adored.

## 8. Findings

In this paper, we discovered that new Yoruba idioms and idiomatic expressions were found to be used to express certain socio-political issues in Nigeria, especially in some part of western Nigeria. These political issues were expressed in a special and unique ways in the following socio-political contexts: 'Eté' (name of a politician) for corruption/stealing; *Ojúyọbọ* (physical description of a politician) for under achiever; *Yóó wọlé ẹ̀ẹken si* (He would win once again) for deceit; *Ó bọ̀jẹ̀ tì* (It would never retrogress or go backward) for collective support; *Oyin ni o* (It is honey) for a successful tenure; *Ajímòbi/Àjùmòṣe* (a politician's name) for collective responsibility; *Nàìjá* (pidgin name for Nigeria) for affection/ disappointment and *Kárí go/Kárí kọ̀mù* (carry go/carry come) for continuity and discontinuity.

Nominalisation, pidginisation and dialect expressions were the strategies employed for the formation of new idioms in this paper. These strategies were used to express different political happenings contextualized in underachievement, sloganeering for support, deceit and acceptance/rejection for second term in office respectively.

New idioms and idiomatic expressions, used to express socio-political issues in Yorùbá routine communication, occurred in mediated and non-mediated contexts and were conveyed through nominalization, pidginisation, and dialect expressions. These idioms reflect dynamism and modernity-constrained stylistic choices in Yorùbá.

## 9. Conclusion

New idioms and idiomatic expressions are quintessentially unique expressions, mostly used in a special and peculiar way when viewed from communicative and stylistic perspective. We therefore, conclude from the foregoing, especially the linguistic strategies involved in the formation of the new idioms and idiomatic expressions, that in spite of the globalisation, which imposes new global communicative challenges on the people, especially the politicians and their followers, the Yorùbá language has the mechanism to cope with such communicative exigencies. The evolution of the new idioms and idiomatic expressions thus becomes a coping strategy to meet the new challenges.

## References

- AKÀNMÚ, Dayo (2003) “Àṣàdídá Ìgbàlódé àti ìlò wọ́n fún Ìdàgbàsókè Èdè Yorùbá”, *Làngbàsà*, 10, 4, March 2003, 42-53.
- AWÓBÙLÚYÌ, Oladele (1992) “Lexical Expansion in Yorùbá: Techniques and Principles”, *Burbark*, 2, 2, June 1992, 14-32.
- BABALQĹÁ, Adeboye (1972) “The Growth of the Yorùbá Language: Problems encountered and Solutions Essayed”, *Paper presented to the Tenth West African Languages Congress*, University of Ghana.
- BÁMGBÓSÉ, A. (1975) “Relative Clauses and Nominalized Sentences in Yorùbá”, in R. K. Herbert (ed.), *Proceedings of the Sixth Conference on African Languages*, Columbus, Ohio: Department of Linguistics, Ohio State University, 202-209.
- CRYSTAL, David (1997)<sup>2</sup> *The Cambridge Encyclopedia of Language*, Cambridge: Cambridge University Press
- MUKAROUSKY, Jan (1970) “Standard Language and Poetic Language”, in D. C. Freeman (ed.), *Linguistic and Literary Style*, Massachusetts: Holt, Rinehart and Winston Inc., 40-59.
- NICHOLS, Johanna (1980) “Pidginisation and Foreigner Talk: Chinese pidgin Russian”, in E. C. Traugott, R. Labrum & S. Shepherd (eds.), *Fourth International conference on historical linguistics*, Amsterdam: Benjamins, 397-407.

- QLÁTÉJÚ, Adeshola (2005) "The Phenomenon of New idioms in Yoruba Literary and routine Communication", in Toyin Falola & Ann Jenova (eds.), *Yoruba Creativity, Fiction, Language, Life and Songs*, Africa World Press, Inc., 273-287.
- OWÓLABÍ, Kola (1976) "Noun-noun Constructions in Yorùbá Language", Ph.D. Thesis, Arts, University of Ibadan, Ìbàdàn.
- RUWET, Nicolas (1973) *An Introduction to Generative Grammar*, Amsterdam: North-Holland.
- SÀLÁMÌ, Yinusa Kehinde (2006) "Yorùbá Proverbs and Democratic Ethos", *Proverbium*, 6, 3, 303-325.
- YUSUF, Omolayo (2002) *Understanding English Usage*, Ibadan: Joyful Printing Press

Received 11 March 2015.

Accepted 10 June 2015.

## SIKH AND HINDU INDIAN THAI NAMING BY SEMANTIC DOMAINS

Warintorn BENJASRI

Mahidol University (Thailand)\*

variinn@gmail.com

### Abstract

This article will explore the traits of Sikh and Hindu Indian Thai naming by semantic domains. According to Nida (1975: 174), groups of meanings, such as names, which share a similar semantic component are known as semantic domains. Ottenheimer (2006: 19) saw a semantic domain as “a specific area of cultural emphasis”. The results of this study will shed light on the different worldviews of Hindus and Sikhs. The results show the semantic domain of Hindu Indian Thais’ naming is divided into 3 types; 1) entities, 2) abstract and 3) events in sequence. The semantic domains of Sikh Indian Thais naming is 1) abstract, 2) entities and 3) events in sequence.

### Keywords

naming, semantic domains, ethnolinguistics, Indian Thai

## SIJ Y TAILANDÉS HINDÚ Y LA DESIGNACIÓN DE CAMPOS SEMÁNTICOS

### Resumen

Este artículo explorará los rasgos de Sij y tailandés hindú en la designación de campos semánticos. Según Nida (1975: 174), los grupos de significados, como los nombres, que comparten un componente semántico similar, se denominan campos semánticos. Ottenheimer (2006: 19) consideró un campo semántico como “una área específica de énfasis cultural”. Los resultados de este estudio arrojarán luz sobre las diferentes visiones de mundo de hindúes y sijes y demostrarán que el campo semántico de la

---

\* 999 Phuttamonthon 4 Road, Salaya, Nakhon Pathom 73170, Thailand.



denominación de los tailandeses hindúes se divide en 3 tipos; 1) entidades, 2) abstracciones y 3) eventos en secuencia. Los campos semánticos de la denominación tailandesa india sij son: 1) abstracciones, 2) entidades y 3) eventos en secuencia.

**Palabras clave**

denominación, campos semánticos, etnolingüística, tailandés hindú

## **1. Introduction**

One of the most prominent and unique features of any and all societies is the names which its people bear. For it is in the names of the people that all of the roots and influences on that society and its culture are reflected. In Mphande's (2006) opinion, these special linguistic phenomena are not only self-significant but are also indicative of cultural and societal influences such as social status, legendary histories, and phonological processes. For all of these reasons, names are an important subject for linguistic study. According to James (1975), the study of these influences takes place within the field of Ethnosemantics, which by definition focuses on the behavior of ethnic groups and attempts to understand the semantic domains (categories) used by each. Ethnosemantics especially focuses on studying the vocabulary of a particular language given that the vocabulary will generally somewhat reflect the entire range of semantic ideas within that language's speakers.

The nation of India has been an important influence on Thailand, particularly with regard to the court system. Dhiravegin (2008) stated that, generally speaking, the Brahminism of the Indian subcontinent and its peoples has been one of the most powerful influencers on Thai beliefs and values. The Office of the Defence Attache, Royal Thai Embassy (2011) reports that currently there are approximately 150,000 Indian people living in Thailand; in addition to these, many have migrated to the nearby countries of Brunei, Indonesia, Singapore, and Malaysia. With regard to belief systems, the two main groups of Indians that have settled in Thailand can be differentiated into Hindus and Sikhs. The majority of these are Hindus.

Archaeologists have discovered that it was during the era of Ashoka the Great that much of Indian culture was brought into Thailand by migrants; evidence for this includes the erection of several statues at the site of Phra Pathom in Nakhon Pathom and the site of Phing Tuk. The statues there have stood since the 3rd or 4th century A.D. (Coedès 1968 17-18). Great conflict within India after the reign of King Rama V resulted in mass migration to Thailand.

Just over a century ago, the Sikh migration began. In 1885, two groups of Sikhs—the Khalsa and the Namdhari—migrated to Sukhumvit Road and Ban Khaek, Lampang, and Phitsanulok. Pongsapit (1991) stated that 91 Sikhs moved to Thailand at this time. Now, there are between 30,000 and 40,000 Sikh Indian Thais in the city of Bangkok, most of them having settled in Pahurat, Thaphra, Ban Khaek, Sukhumvit and Klongton.

From the perspective of the Thai people, most Hindus and Sikhs are seen as foreigners even though they have been granted legal citizenship as Thais. This is due to the fact that these peoples still adhere to their own native customs and traditions and still speak their native languages. Furthermore Thai Indians do not intermarry because of the great differences between the two cultures. One way for them to maintain their ethnic identity is by adhering to their traditional naming conventions, usually abstaining from the use of Thai names, except in certain circumstances.

Traditionally in India, names were based on location, religion, caste, and profession and were composed of a phrase, single unit-word, or word-division. According to Kaushik (2000), the probability of occurrence of the last name does not impact the reliable distribution in all terminology places. In many cases, the first part of a name will refer to the place of one's home, and the last part of a name will refer to caste or some other reference.

Previously there have been few studies about Indian Thai groups, and the research that has been done has tended to be about descriptive and historical features of these groups. This research focuses on Hindu and Sikh Indian Thais due to the fact that they are the largest and oldest groups in Thailand. This is the first study to inquire into the conventions for assigning names and surnames to Indian Thais in terms of ethnolinguistics and the cultural perspectives reflected in the names

## **2. Objective**

The main intent of this research is to examine the linguistic structure of Indian Thai naming conventions. Various linguistic components of the names will be analyzed semantic domains.

## **3. Materials and Methods**

This research will follow a standardized procedure involving the following four categories: (1) Data Limitation, (2) Data Preparation, (3) Data collection and (4) Data analyses.

### *3.1 Data Limitation*

The data limitation for this research is Data for this research approximately 1000 Indian Thais in Bangkok, gathered for analysis of conventions and meanings.

### *3.2 Data Preparation*

Important to this study will be the use of journals, articles, textbooks and theses that are pertinent to Indian Thai naming conventions. The theoretical frameworks therein will be important for this study. Foundational to this study will be the collection of data (Indian Thai names) from 1000 informants around Thailand (Bangkok, Pathumthani, Nakhon Pathom, Chonburi, Chiang Mai, Lampun, Lampang, Nakhon Ratchasrima, Phuket, Yala and Trang).

### *3.3 Data collection*

For data about naming conventions, sources such as textbooks, dissertations, and journals will be consulted. This will allow the researcher to investigate the structures of

the naming conventions and the word components themselves, leading to a coherent semantic analysis of social and cultural factors influencing naming conventions.

A sample base of approximately 1000 Indian Thais (500 being Sikh and 500 being Hindu) and their relatives will be interviewed for this research. The sample will consist of Indian Thais whose ancestors migrated to Thailand from India. This data will include focus groups within three to four generations per sample.

### *3.4 Data analyses*

Because most Thai names are Sanskritized, the roots of the words/names must be determined. To do this, linguistic features will be analyzed from the data from an ethnolinguistic and semantic perspective. This will be followed by semantic analysis which will involve the categorization of the words based on meaning and a variety of socio-cultural factors. Finally, there will be a discussion of the findings.

Foundational to this study will be the collection of data (Indian Thai names) from 1000 informants around Thailand (Bangkok, Pathumthani, Nakhon Pathom, Chonburi, Chiang Mai, Lampun, Lampang, Nakhon Ratchasrima, Phuket, Yala and Trang).

Because most Thai names are Sanskritized, the roots of the words/names must be determined. To do this, linguistic features will be analyzed from the data from an ethnolinguistic and semantic perspective. This will be followed by semantic analysis which will involve the categorization of the words based on meaning and. Finally, there will be a discussion of the findings.

## **4. Hindu Indian Thai naming**

Hindu Indian Thai names are influenced by a variety of social constructs, including the four castes: the brahmanas, or priests, representing purity; the kshatriyas, or warriors, representing courage; the Vaishyas, or businessmen, representing prosperity; and the Shudras, or laborers, representing hard work. The Shudras are often seen as the

lowest caste in Indian society. Shudras include servants (Kaushik 2000: 13-42). Nowadays, more weight is put in personal preference than in caste.

#### 4.1 First names

##### 4.1.1 Entities semantic domain

Entities are used for 62% of Hindu Thai names, as follows:

##### 4.1.1.1 Inanimate entities

a) *Natural inanimate entities*: Natural phenomena account for 26.8% of Hindu Indian Thai names (15% of male names; 11.8% of female names), as follows:

Celestial-atmospheric entities include the following examples (15.2% of names are celestial-atmospheric entities; 8.68% of male names; 6.51% of female names):

- |               |                  |               |                     |
|---------------|------------------|---------------|---------------------|
| (1) Raveena   | (Female, Urdu)   | /ravina/      | 'beauty of the Sun' |
| (2) Nabh      | (Male, Sanskrit) | /nab/         | 'the sky'           |
| (3) Arundhati | (Female, Hindi)  | /ʔarunthatiʔ/ | 'the star'          |
| (4) Sabrang   | (Male, Hindi)    | /sapraŋ/      | 'rainbow'           |

Entities related to earth, geography, and names of cities include the following examples (3% of names are earth, geography (1.5% of male names; 1.5% of female names):

- |             |               |           |                     |
|-------------|---------------|-----------|---------------------|
| (5) Pavan   | (Male, Hindi) | /pawan/   | 'wind, air, breeze' |
| (6) Hemadri | (Male, Hindi) | /himatri/ | 'Himalaya mountain' |
| (7) Swarg   | (Male, Hindi) | /sawak/   | 'name of city'      |

Gems make up for 3.2% of names, include the following examples (1.92% of male names; 1.28% of female names):

- |            |                  |         |           |
|------------|------------------|---------|-----------|
| (8) Himank | (Male, Hindi)    | /himaŋ/ | 'diamond' |
| (9) Neelar | (Male, Sanskrit) | /nila/  | 'emerald' |
| (10) Panna | (Female, Hindi)  | /panna/ | 'gem'     |

Names reflecting fire and flame include the following examples [1.2% of names are fire and flame (0.8% of male names; 0.4% of female names)]:

- |               |                    |               |        |
|---------------|--------------------|---------------|--------|
| (11) Patag    | (Male, Hindi)      | /patak/       | 'fire' |
| (12) Prajuala | (Female, Sanskrit) | /prattchuala/ | 'fire' |

Names reflecting water include the following [1.4% of names are various forms of water (0.7% of male names; 0.7% of female names)]:

- |             |                    |          |        |
|-------------|--------------------|----------|--------|
| (13) Karana | (Male, Hindi)      | /karana/ | 'tide' |
| (14) Kavana | (Female, Sanskrit) | /kawana/ | 'tide' |

Flora and plant products include the following [2.8 % of names are flora and plants (1.4 % of male names; 1.4% of female names as in the following examples)]:

- |                |                          |             |                        |
|----------------|--------------------------|-------------|------------------------|
| (15) Kalakar   | (Male, Hindi)            | /kalakan/   | 'tree'                 |
| (16) Kusumakar | (Male, Sanskrit)         | /kusumakan/ | 'the bundle of flower' |
| (17) Laksha    | (Female, Sanskrit)       | /laktcha/   | 'flower, white rose'   |
| (18) Kamal     | (Female, Hindi, Punjabi) | /kaman/     | 'lotus'                |

b) Manufactured or constructed inanimate entities: Only 2.8% (2.45% of male names; 0.35% of female names) of Hindu Indian Thais' naming are based on these entities as in the following examples:

Decoration:

- |              |                    |              |            |
|--------------|--------------------|--------------|------------|
| (19) Kirit   | (Male, Sanskrit)   | /ki?rit/     | 'crown'    |
| (20) Mudrika | (Female, Sanskrit) | /mattharika/ | 'ring'     |
| (21) Trishar | (Male, Hindi)      | /tritcha/    | 'necklace' |

Lighting and lamps:

- |              |                         |           |                |
|--------------|-------------------------|-----------|----------------|
| (22) Pradeep | (Male, Sanskrit, Hindi) | /prathip/ | 'lamp'         |
| (23) Deepak  | (Male, Hindi)           | /dipak/   | 'lamp, kindle' |

Musical instruments:

- |             |                        |         |         |
|-------------|------------------------|---------|---------|
| (24) Banshi | (Male, Hindi, Punjabi) | /bansi/ | 'flute' |
|-------------|------------------------|---------|---------|

Processed substances:

- |             |                  |           |             |
|-------------|------------------|-----------|-------------|
| (25) Ikshu  | (Male, Sanskrit) | /?iktchu/ | 'sugarcane' |
| (26) Madhur | (Male, Hindi)    | /mathura/ | 'sweet'     |

## 4.1.1.2 Animate entities

These names make up 32.4 % (27.7% of male names; 4.7% of female names) of Hindu Indian Thai names as in the following examples:

a) Animals: 3.14% of names are related to animals:

1.57% of names refer to poultry:

(27) Vihang (Male, Sanskrit, Hindi) /wihan/ 'bird'

1.05% of names refer to quadrupeds:

(28) Kamboj (Male, Sanskrit) /kambot/ 'elephant'

(29) Neshu (Male, Sanskrit) /net/ 'lion'

0.52% of names refer to body parts of animal:

(30) Muktā (Female, Sanskrit) /mukta/ 'pearl'

b) Human beings

13.6% of names are human beings (male 12.54%, female 1.05%) consist of the following sub-domains:

Generic and distinguished by age and sex:

(31) Bhupen (Male, Bangala) /phupen/ 'king'

(32) Kanvar (Male, Punjabi) /khanwara/ 'young Prince'

(33) Riya (Female, Arabic) /riya/ 'singer'

Males:

(34) Manish (Male, Sanskrit, Hindi) /manit/ 'man who can control thought'

(35) Hadrik (Male, Urdu) /hattharika/ 'boy'

Offspring:

(36) Hareh Kumar (Male, Hindi) /haretkuman/ 'The son of Lord Siva'

(37) Jitendra Kumar (Male, Hindi) /tçithendrakuman/ 'Son of virtuous father'

(38) Mahesh Kumar (Male, Hindi) /mahetkuman/ 'the son of God'

(40) Anusha (Male, Sanskrit) /?anutcha/ 'younger brother'

(41) Falak Kumar (Male, Urdu) /falakkuman/ 'son of the sky'

Affinals:

(42) Ira	(Female, Hindi)	/ʔira/	'the wife of Manu'
(43) Harini	(Female, Urdu)	/harini/	'deer, darling'

Brothers and sisters:

(44) Pehlaj	(Male, Persian)	/pelat/	'elder brother'
(45) Anusha	(Male, Sanskrit)	/ʔanut̪cha/	'younger sister'

Group consists of socio-politics:

(46) Pandey	(Male, Hindi)	/pande/	'a Brahmin sub-caste'
-------------	---------------	---------	-----------------------

Body, body parts, and body expressions:

(47) Citrākṣa	(Male, Hindi)	/t̪cit̪rak̪sa/	'expressive eyes'
(48) Nayan	(Male, Sanskrit, Hindi)	/najan/	'eyes'

c) Supernatural powers or beings

15.26% of supernatural powers or beings are divided into powers or personifications and personal beings:

4.73% of names refer to powers or personifications:

(49) Kripa	(Female, Sanskrit)	/kripha/	'merciful'
(50) Rujul	(Male, Sanskrit)	/rut̪t̪hun/	'honest, faithful'
(51) Riddhi	(Female, Sanskrit)	/rit̪thiʔ/	'power'

10.53% of names refer to personal beings:

(52) Adhi	(Male, Sanskrit)	/ʔathiʔ/	'Name of Lord Ganesha'
(53) Biharikunj	(Male, Sanskrit)	/phihanrikun /	'Name of Lord Krishna'
(54) Gajadhar	(Male, Hindi)	/kat̪t̪hathan/	'Name of Lord Vishnu'
(55) Jogishwar	(Male, Sanskrit)	/jokhit̪hawa/	'The God of yoga'
(56) Kumar Nandana	(Male, Sanskrit)	/kumannandana/	'son of God'
(57) Abhayā	(Female, Hindi)	/ʔapheja/	'Goddess'
(58) Mahesh Kumar	(Male, Hindi)	/mahetkuman/	'Name of Lord Gan@esa'



## 4.1.2 Abstract semantic domain

22.8% of abstract domain is used for naming which refers to the following name groups:

8.2% of names refer to capacity:

(59) Ashvath	(Male, Hindi)	/ʔattɕawat/	'Strong like banyan tree'
(60) Dakṣi	(Female, Sanskrit)	/thaksi/	'expert'
(61) Aiswarya	(Female, Hindi)	/ʔaytɕhawa:raya:/	'prosperous'

5.6% of names refer to number:

(62) Anit	(Male, Hindi)	/ʔanit/	'small, a little'
(63) Balraj	(Male, Hindi)	/banrat/	'great, big'

4.8% of names refer to time:

(64) Arun	(Male, Punjabi)	/ʔarun/	'dawn'
(65) Rajni	(Female, Hindi) /	ratɕhani/	'night'

3.4% of names refer to good-bad:

(66) Uttam	(Male, Hindi)	/ʔattam/	'good'
(67) Suchir	(Male, Hindi)	/sutɕhə/	'Good for long time'

2.6% of names refer to attractiveness:

(68) Sanjula	(Female, Hindi)	/sanjula/	'beautiful'
(69) Jahi	(Male, Persian, Hindi)	/jahi/	'chic, stylish, handsome'
(70) Kashish	(Female, Persian, Hindi)	/katchit/	'charming'

2.2% of names refer to status:

(71) Aiswarya	(Female, Hindi)	/ʔaytɕhawaraya/	'prosperous'
(72) Udbal	(Male, Hindi)	/ʔatbaw/	'famous'

2% of names refer to color:

(73) Shweta	(Female, Hindi)	/tɕhaweta/	'white'
(74) Neela	(Female, Hindi, Urdu)	/nila/	'blue'

1.2% of names refer to volume which are males:

(75) Suran	(Male, Hindi)	/suran/	'melodious'
------------	---------------	---------	-------------

0.2% of names refer to velocity:

(76) Ranam	(Male, Urdu)	/ranam/	'fast/quick'
------------	--------------	---------	--------------

0.2% of names refer to truth-falsehood:

(77) Tatbir	(Male, Persian)	/thatbə/	'truth'
-------------	-----------------	----------	---------

#### 4.1.3 Events semantic domain

This semantic domain relates to emotions, intellect, and control. Hindu Indian Thai names are based on 15.2% of this sub-domain.

16.4% of names refer to emotions:

- |               |                         |             |             |
|---------------|-------------------------|-------------|-------------|
| (78) Khushi   | (Female, Punjabi, Urdu) | /khut̪chi/  | 'happiness' |
| (79) Akanksha | (Female, Hindi)         | /ʔakant̪ʃa/ | 'desire'    |
| (80) Himal    | (Urdu, Punjabi)         | /himan/     | 'calm'      |

1.64% of names refer to intellect:

- |              |                           |              |             |
|--------------|---------------------------|--------------|-------------|
| (81) Jignasa | (Female, Sanskrit, Hindi) | /t̪ʃhiknasa/ | 'curiosity' |
|--------------|---------------------------|--------------|-------------|

3.28% of names refer to control:

- |               |                           |               |           |
|---------------|---------------------------|---------------|-----------|
| (82) Amarjeet | (Male, Punjabi)           | /ʔamarat̪hit/ | 'victory' |
| (83) Shamak   | (Male, Persian, Sanskrit) | /t̪ʃhamak/    | 'peace'   |

#### 4.2 Surnames

The Hindu Indian Thai surnames can be analyzed into 3 semantic domains and several sub domains as follows:

##### 4.2.1 Entities semantic domain

This semantic domain accounts for 48.6% of Hindu Indian Thai Surnames, with the following examples:

###### 4.2.1.1 Inanimate entities

The inanimate (sub-)domain refers to nature, the supernatural, substances, flora/plant entities and manufactured or constructed entities, as follows:

a) Natural inanimate entities: Natural phenomena account for 22.4% of Hindu Indian Thai surnames, as follows:

12.92% of surnames refer to celestial-atmospheric entities:

- |                  |            |                   |                     |
|------------------|------------|-------------------|---------------------|
| (84) Lalitaditya | (Sanskrit) | /lalita:thittaya/ | 'the beautiful sun' |
|------------------|------------|-------------------|---------------------|

(85) Kshitij (Hindi) /kasidit/ 'horizon'

3.3% of surnames refer to geography:

(86) Jaldhar (Hindi) /tçhanthara/ 'stream, waterway'

(87) Sagardutt (Hindi) /sakaradut/ 'gift of ocean'

3.2% of surnames refer to gems:

(88) Kanchan (Hindi) /kantçhan/ 'gold'

(89) Neelam (Hindi) /nilam/ 'blue sapphire'

0.9 % of surnames refer to various forms of water:

(90) Mehul (Persian, Hindi) /mehun/ 'rain'

(91) Jaldhar (Hindi) /jathan/ 'stream, waterway'

2.08 % of surnames refer to flowers:

(92) Kusum (Hindi) /kusum/ 'flower'

b) Manufactured or constructed inanimate entities: Only 2.2% of Hindu Indian surnames are based on these entities as in the following examples:

Decoration:

(93) Suvarnamālā (Sanskrit) /suwanmala/ 'gold necklace'

#### 4.2.1.2 Animate subdomain

24% of the animate sub-domain is used for Hindu Indian Thai surnames according to the following name groups:

a) Human beings: 7% of names related to human beings. These consist of 0.4% of kinship terms:

(94) Nandini (Hindi) /nandini/ 'Daughter'

(95) Hridyanshu (Hindi) /haritjantçhu/ 'Light of heart'

b) Supernatural powers or beings: Supernatural powers or beings (17%) consist of the following name groups:

2.2% of surnames refer to powers or personifications:

(96) Dhaval (Sanskrit, Hindi) /thawan/ 'fair'

(97) Palin (Hindi) /palin/ 'safeguard'

14.8% of surnames refer to personal beings:

(98) Khagesh	(Hindi)	/khaket/	'God of birds'
(99) Manju	(Hindi)	/mantchu/	'Lord Siva'

#### 4.2.2 Abstract semantic domain

42.4% of abstract domain is used for surnames which refer to the following name groups:

10.8% of surnames refer to time:

(100) Nishita	(Hindi)	/nitçhita/	'night '
(101) Prathysha	(Hindi)	/pratya/	'morning'

0.2% of surnames refer to velocity:

(102) Chapal	(Hindi)	/tçha:pan/	'fast'
--------------	---------	------------	--------

2.8% of surnames refer to number:

(103) Yakta	(Hindi)	/yakta/	'unique'
(104) Nikhil	(Sanskrit)	/nikhiw/	'infinite, limitless'

3.45% of surnames refer to status:

(105) Darlat	(Persian, Hindi)	/dalat/	'stability'
(106) Dhureen	(Punjabi, Hindi)	/thuri:n/	'achievement'

8.2% of surnames refer to religious character:

(107) Ashish	(Hindi)	/ʔatçhit/	'blessing'
(108) Oni	(Hindi)	/ʔoni/	'born in a holy place'

8.6% of surnames refer to attractiveness:

(109) Manjulas	(Hindi)	/mantçhulat/	'loveable'
(110) Mohit	(Hindi)	/mohit/	'charming'
(111) Nikhil	(Sanskrit)	/nikhiw/	'perfection'

0.2% of surnames refer to truth-falsehood:

(112) Tatbir	(Persian)	/tatbæ/	'truth'
--------------	-----------	---------	---------

0.2% of surnames refer to Good-bad:

(113) Sheil	(Hindi)	/tçhiw/	'good conduct'
-------------	---------	---------	----------------

7.4% of surnames refer to capacity:

(114) Sonya	(Punjabi, Hindi)	/sonya/	'wisdom, clever'
-------------	------------------	---------	------------------

(115) Yukta (Hindi) /yukta/ 'attentive, skillful'

#### 4.2.3 Events semantic domain

8.8% of events are used as surnames relate to physiological entities; emotions; intellect, control and movement as in the following examples:

0.4% of surnames refer to physiological entities:

(116) Janam (Hindi) /tʃanam/ 'birth'

5.75% of surnames refer to emotions:

(117) Kanth (Hindi) /kan/ 'beloved'

(118) Parmanand (Hindi) /paramanan/ 'superlative joy'

2.46% of surnames refer to intellectual:

(119) Tatva (Hindi) /tatwa/ 'Fundamental idea'

2.46% of surnames refer to control:

(120) Baljit (Punjabi) /bantʃit/ 'winning power'

(121) Jishnu (Hindi) /tʃhitsanu/ 'victorious'

0.4% of surnames refer to movement:

(122) Pretvan (Hindi) /pretwan/ 'moving along'

## 5. Sikh Indian Thai naming

Cole & Sambhi (1978: 43-97) explained that in the Naming Ceremony of Sikhism, a hymn from the Granth (Sikhism's central text) is chosen at random. The first letter of the hymn, any letter of the alphabet, is chosen as the initial of the child's name. After this, the congregation is asked to suggest a suitable name beginning with that letter. The Sikh Indian Thais' naming can be divided into first names and surnames:

### 5.1 First names

Sikh Indian Thai first names can be analyzed into three semantic domains (Abstract, Entities, and Events).

### 5.1.1 Abstract semantic domain

52.68% of abstract domain is used for the Sikh names which refers to religious character, time, attractiveness, capacity, number, status, good-bad, velocity, and color as in the following detail. Sikhism worships the Guru Granth Sahib and believes in peace.

39.86% of names refer to religious character (30.97% of male; 8.89% of female):

(123) Darvesh	(Male, Hindi)	/darawet/	'religious, gentleman'
(124) Diksha	(Female, Punjabi)	/thit̥cha/	'initiation, consecration'

3.21% of names refer to time (2.5 % of male; 0.71% of female):

(125) Basant	(Female, Punjabi)	/phasan/	'spring season'
(126) Chiranjiv	(Male, Sanskrit, Hindi)	/t̥chirant̥chip/	'immortal'

2.81% of names refer to attractiveness (1.405% of male; 1.405% of female):

(127) Komal	(Female, Punjabi)	/koman/	'tender, soft, delicate'
(128) Mohanbir	(Male, Punjabi)	/mo:hanb̥ə:/	'beautiful, brave'

2.2% of names refer to capacity (1.65% of male; 0.99% of female):

(129) Chathuv	(Female, Punjabi)	/t̥cat̥thuwa/	'clever'
---------------	-------------------	---------------	----------

1.4% of names refer to number (0.33% of male; 0.33% of female):

(130) Asankh	(Male, Punjabi)	/ʔasan/	'countless'
(131) Ekam	(Female, Punjabi)	/ʔekam/	'one, united'

1% of names refer to status: (0.75% of male; 0.25% of female)

(132) Anokh	(Male, Punjabi)	/ʔanok/	'extraordinary'
(133) Darvesh	(Male, Hindi)	/darawet/	'humble'

1% of names refer to good-bad (0.75% of male; 0.25% of female):

(134) Adarsh	(Male, Hindi)	/ʔadat/	'ideal'
(135) Nirmai	(Female, Hindi)	/niramay/	'without flaw'

0.4% of names refer to velocity (0.4% of male):

(136) Teja	(Male, Punjabi)	/t̥et̥cha/	'fast, speedy, sharp'
------------	-----------------	------------	-----------------------

0.2% of names refer to color (0.2% of female):

(137) Basant	(Female, Punjabi)	/phasan/	'yellow'
--------------	-------------------	----------	----------

0.6% of names refer to age: (0.2% of only female):

(138) Yuvleen	(Male, Punjabi)	/yuwalin/	'absorbed in youthfulness'
---------------	-----------------	-----------	----------------------------

### 5.1.2 Entities semantic domain

42.32% of entities domain are used for Sikh names, including animate entities (animal, human beings and supernatural powers or beings) and inanimate entities (natural and manufactured or constructed entities).

#### 5.1.2.1 Animate entities

Animate entities sub-domain consists of animals, human beings and supernatural as follows:

a) Animals: 1% of names are Animals, as follows:

(139) Arjun	(Male, Hindi)	/ʔarajun/	'peacock'
(140) Variam	(Male, Punjabi)	/wariʔam/	'lion'

b) Human beings: 2.6% of names are human beings (2.2% of male; 0.4% of female), as follows:

Generic and distinctions by age and sex:

(141) Hamin	(Female, Urdu)	/hamin/	'woman'
(142) Nina	(Female, European)	/nina/	'little girl'

Group consists of sociopolitical:

(143) Fauja	(Male, Punjabi)	/faoja/	'army, general'
(144) Meeta	(Male, Punjabi)	/mita/	'friend'

Body, body parts, and body products:

(145) Mukhlapa	(Female, Punjabi)	/mukhalapha/	'face'
(146) Netter	(Male, Punjabi)	/net/	'eyes'

c) Supernatural powers or beings: 20.3% of names are related to supernatural powers or beings, as follows:

Powers or personifications:

(147) Taan	(Male, Punjabi)	/taʔan/	'power'
(148) Wafa	(Male, Urdu)	/wafa/	'faithful'

Personal beings:

- |                 |                         |               |                        |
|-----------------|-------------------------|---------------|------------------------|
| (149) Aditya    | (Male, Sanskrit, Hindi) | /ʔathittaya/  | 'sun, lord Surya'      |
| (150) Arvinderā | (Male, Punjabi)         | /ʔarawindəra/ | 'lord of Horses, lord' |

5.1.1.2 Inanimate entities

a) Natural inanimate entities: Natural phenomena account for 11.42% of Sikh Indian Thais names, as follows:

6% of names refer to celestial-atmospheric entities:

- |              |                   |           |                           |
|--------------|-------------------|-----------|---------------------------|
| (151) Dhoop  | (Male, Punjabi)   | /thup/    | 'at the edge of sunshine' |
| (152) Divjot | (Female, Punjabi) | /thipjot/ | 'divine light'            |

2.1% of names refer to entities related to earth, geography:

- |                 |                 |           |          |
|-----------------|-----------------|-----------|----------|
| (153) Deenaagal | (Male, Punjabi) | /dinakan/ | 'valley' |
| (154) Gulzar    | (Male, Urdu)    | /kunsa/   | 'garden' |

0.52% of names refer to gems:

- |               |                 |            |        |
|---------------|-----------------|------------|--------|
| (155) Kanchan | (Female, Hindi) | /kantchan/ | 'gold' |
|---------------|-----------------|------------|--------|

2% of names refer to fire and flame:

- |              |                         |             |         |
|--------------|-------------------------|-------------|---------|
| (156) Jyoti  | (Male, Sanskrit, Hindi) | /tchayoti?/ | 'flame' |
| (157) Jujhar | (Male, Punjabi)         | /tchuthan/  | 'flame' |

0.8% of names refer to flora and plant products:

- |                |                 |            |          |
|----------------|-----------------|------------|----------|
| (158) Harzadan | (Male, Punjabi) | /hasadan/  | 'flower' |
| (159) Kinshuk  | (Male, Punjabi) | /kintchuk/ | 'flower' |

b) Manufactured or constructed inanimate entities: Only 2.2% of Sikh Indian Thais' naming are based on these entities as in the following examples:

1.8% of names refer to offense weapons:

- |              |              |          |                   |
|--------------|--------------|----------|-------------------|
| (160) Talwar | (Male, Urdu) | /tanwan/ | 'sword or dagger' |
|--------------|--------------|----------|-------------------|

0.4% of names refer to decoration:

- |                 |                 |           |              |
|-----------------|-----------------|-----------|--------------|
| (161) Deenaagal | (Male, Punjabi) | /dinakan/ | 'decoration' |
|-----------------|-----------------|-----------|--------------|

1.96% of names refer to lighting and lamp stands:

- |              |                     |         |        |
|--------------|---------------------|---------|--------|
| (162) Deepak | (Male, Male, Hindi) | /dipak/ | 'lamp' |
|--------------|---------------------|---------|--------|



### 5.1.2 Events semantic domain

Events (5%) are related to physiology, emotion, intellect, control, and transfer. The meaning of events is indicated by the Sikh Indian Thai people's ideal character.

1% of names refer to physiology:

(163) Ankur	(Male, Hindi)	/ʔanjun/	'sprout, glow'
(164) Deepak	(Male, Sanskrit, Hindi)	/dipak/	'kindle'

3.2% of names refer to emotion (2% of males; 0.2% of females):

(165) Akalsukh	(Male, Punjabi)	/ʔakansuk/	'delight'
(166) Premdeep	(Female, Hindi)	/premthip/	'The lamp of love'

1.8% of names refer to intellect:

(167) Gianparkash	(Male, Punjabi)	/kiʔanpakat/	'Light of divine knowledge'
(168) Gianpreet	(Male/Female, Punjabi)	/kiʔanprit/	'Lover of divine knowledge'

### 5.2 Surnames

The Sikh Indian Thai surnames can be analyzed into 3 semantic domains and several sub domains as follows:

#### 5.2.1 Abstract semantic domain

Abstract domain is used for the Sikhs' surnames (52.57%) as in the following examples:

35.11% of surnames refer to religious character:

(169) Adish	(Punjabi)	/ʔadit/	'exalted'
(170) Sharan	(Punjabi)	/tcharan/	'taking shelter of the Guru'

10.04% of surnames refer to status:

(171) Yash	(Sanskrit, Hindi)	/yat/	'glory'
(172) Yashas	(Sanskrit, Hindi)	/yatchat/	'famous'

4.41% of surnames refer to attractiveness:

(173) Mehar	(Punjabi, Hindi)	/mehan/	'kindness, grace'
(174) Mohit	(Punjabi)	/mohit/	'beauty, charming'

2.81% of surnames refer to number:

(175) Pehila	(Punjabi)	/pehila/	'first'
--------------	-----------	----------	---------

0.2% of surnames refer to time:

(176) Prabhat	(Hindi)	/praphat/	'dawn'
---------------	---------	-----------	--------

## 5.2.2 Entities semantic domain

41.34% of Sikh surnames are from the entities domain, which refers to animate entities (animal, human beings and supernatural powers or beings) and inanimate entities (natural and manufactured or constructed entities).

### 5.2.2.1 Animate entities

The animate (sub-)domain refers to animals, human beings, and the supernatural, as follows:

a) Supernatural powers or beings: Supernatural powers or beings include the following examples:

0.2% of surnames relate to powers or personifications:

(177) Kalyan	(Hindi)	/kala:yan/	'auspicious'
--------------	---------	------------	--------------

19.87% of surnames relate to supernatural beings

(178) Akaal	(Punjabi)	/ʔakan/	'God/Goddess Saraswati'
-------------	-----------	---------	-------------------------

(179) Mahadev	(Hindi)	/mahathep/	'Lord of Siva'
---------------	---------	------------	----------------

(180) Surina	(Punjabi)	/surina/	'Goddess'
--------------	-----------	----------	-----------

b) Human Beings: Human beings (18.47%) are represented by social caste, as in the following examples:

(181) Singh	(Punjabi)	/siŋ/	'derive from Kshatriya caste'
-------------	-----------	-------	-------------------------------

(182) Sirinetkul	(Sanskrit)	/sirinetkun/	'derived from Kshatriya caste of Hindu Varna system from Rajputs of Rajasthan'
------------------	------------	--------------	--

### 5.2.2.2 Inanimate entities

The inanimate (sub-)domain refers to nature, the supernatural, substances, flora/plant entities and manufactured or constructed entities, as follows:

a) Natural inanimate entities: Natural phenomena account for 2.4% of Sikh Indian Thai surnames:

1.6% of surnames relate to celestial-atmospheric entities:

(183) Mehal	(Urdu)	/mehan/	'cloud'
(184) Sudipti	(Punjabi)	/suthipti?/	'brightness'

0.2% of surnames relate to gems:

(185) Hira	(Punjabi)	/hira/	'diamond'
------------	-----------	--------	-----------

0.4% of surnames relate to flora and plant products:

(186) Vigsai	(Punjabi)	/vigsay/	'blossom'
--------------	-----------	----------	-----------

b) Manufactured or constructed inanimate entities: Manufactured or constructed entities refer to artifacts which can be divided into generic items and offensive weapons. Only 0.6% of Sikh Indian Thais names are based on these entities as in the following examples:

0.2% of surnames are generic:

(187) Nidhi	(Hindi)	/ni?thi?/	'treasure, wealth'
-------------	---------	-----------	--------------------

0.4% of surnames relate to offensive weapons:

(188) Tulwa	(Punjabi)	/tunwa/	'sword'
-------------	-----------	---------	---------

### 5.2.3 Events semantic domain

Events are used for the Sikhs' surnames at 6.09%. The meaning of events, in term of semantic domains, conveys the Sikh Indian Thai people's ideal character.

3.02% of surnames relate to emotion:

(189) Kaan	(Punjabi)	/ka/an/	'wish, desire'
(190) Preet	(Hindi)	/prit/	'love'

1.78% of surnames relate to intellect:

(191) Sochai	(Punjabi)	/sotçhay/	'by thinking'
--------------	-----------	-----------	---------------

1.29% of surnames relate to movement:

(192) Rijak	(Sanskrit)	/ritçhak/	'sustenance'
(193) Vigsai	(Punjabi)	/vigsay/	'to come into bloom'

## 6. Summary

Hindu and Sikh names are most common in regards to those names which reflect 'emotions.' Regarding female names, those Hindu and Sikh female names which refer to 'emotions' are most similar, and those referring to 'religious characters' are second-most similar.

Hindu males and females are named similarly as regards entities; of these, 'natural' and 'animals' sub-domains are highest. The 'emotion' sub-domain the second-highest.

Sikh males and females are named mostly as 'religious characters'.

From this study, the principal semantic domain is 'religious characters.' For Hindu Indian Thais, values about holy God, spirit and cosmos are important. Most Hindus believe in a few foundational ideas, such as inherent sacredness in nature. Hindus believe that an educated expert, or satguru, is required for an average person to know transcendence. They believe that all life forms are holy, to be loved; therefore they exercise ahimsa, or non-injury.

Sikhism is a monotheistic faith. Sikhism identifies God as being only one. To Sikhs, God is the Designer of the Galaxy, omniscient, timeless and omnipotent. Sikhs do not recognize any other gods.

The results of the data analysis are consistent with the concept of Nida (1975: 174). The result exposed that a semantic sector is comprised of a number of meanings which share semantic elements.

## 7. Popular Indian Thai Naming

Ferguson (1864) said: "...the etymology of proper names is the only branch then of the subject which can in any sense be called popular, for what men, even of those who care not to enquire the origin of the language they speak, feel some interest or curiosity in knowing the meaning of the names they bear..." (Ferguson *apud* Rosenthal 2005: 3).

### 7.1 Male Hindu Indian Thai names

Name	Percentage
Arsok 'Without sorrow, happiness, name of emperor'	18%
Suchir 'Good for a long time'	12%
Aruna 'Dawn, a god of the Hindus'	9%
Others	61%

Table 1. Popular Male Hindu Indian Thai names

Three such names were identified. Arsok is a form of the Indian Ashoka, denoting the importance of happiness. Ashoka, an ancient Indian ruler, urged his kin to practice Buddhist values, he encouraged an air of appreciation for all religions. Ashoka served as an illustration of tolerance for his subjects (Thapar 1973: 51). Suchir (the second name) is the representative of Hinduism which is the "Everlasting way". Everlasting in Hinduism implies those living creatures who will stay alive on Earth until the end of this modern age. The third name, Aruna, is an exemplification of the ruddy gleam of the climbing Sun, which is considered to have otherworldly powers. In Hindu Tradition, Aruna is the charioteer of Surya (the Sun God) and he is additionally the red sky of first light or throughout dawn. The story of his heavenly conception is said in the Astika Parva of Mahabharata (Klostermaier 2007).

### 7.2 Female Hindu Indian Thai names

Name	Percentage
Ridhhi 'Power, Hindu God'	20%
Neela 'A character in Hindu mythology, beautiful'	11%

Mukta 'Liberated, pearl'	7%
Others	62%

Table 2. Popular Female Hindu Indian Thai names

The name Ridhhi ranks the highest in popularity. Siddhi and Riddhi are the wives of the god Ganesha. Thus, anyone who honors Ganesha are likewise honored by Riddhi and Siddhi and can accomplish everything in their life (Brown 1991: 115-140). The second name, Neela/Nala, is the goddess of Nishadha Kingdom, child of Veerasena and possessed by the God Shani (Doniger 1999). This name is associated with the Hindu female ideal. The third name, Mukta, represents freedom. This is reflected in the fact that many informants migrated from India to start a life in Thailand.

### 7.3 Male Sikh Indian Thai names

Name	Percentage
Harsharan 'One who takes shelter in the Lord'	40%
Deepak 'Lamp, kindle	15%
Hukam 'Gods who Command, authority, direction'	9%
Others	36%

Table 3. Popular Male Sikh Indian Thai names

Harsharan is the most popular of Male Sikh Indian names. Sikhism recognizes that the spiritual and physical world are entwined (Cole & Sambhi 1995) and thus seeks to secure the privileges of all individuals. The second name, Deepak, means an oil light. The oil in the Deepak represents the dark part of a human spirit that people have a tendency to sustain, and burning the wick is typical of the atman (self). To accomplish illumination and unite with the preeminent force, one must dispose of materialism (Cole & Sambhi 1995).

The third name, Hukam, is also associated with Sikhism. In the Sikh Scripture, Guru Nanak, the author of the religion, says: "O Nanak, by the Hukam of God's Command, we travel every which way in resurrection. The entirety of the Universe is liable to the Hukam of the Creator God. Nothing ever happens without the Will of Him." This is acknowledged as one of the essential ideas of Sikhism.

#### 7.4 Female Sikh Indian Thai names

Name	Percentage
Livsharan 'Absorbed in the Lotus Feet of God'	35%
Ishwari 'Supreme Goddess'	18%
Jasdeep 'The lamp radiating God's Glories'	8%
Others	39%

Table 4. Popular Female Sikh Indian Thai names

Livsharan is the most popular female Sikh Indian Thai name, followed by Ishwari and Jasdeep. These three names refer to the Sikh Oneness of God.

According to the data, the most popular of male and female Sikh Indian names are associated with Sikhism. The names of Sikhs are consistent in both sexes.

#### 7.5 The language used in naming

Language	Hindu's Naming	Sikh's Naming
Hindi	73.13%	24.39%
Sanskrit	13.53%	2.38%
Punjabi	1.38%	65.36%
Others	11.95%	7.87%
Total	100%	100%

Table 5. Languages used in naming

For Hindu Indian Thai names, Hindi is the most popular for names, followed by Punjabi and then Sanskrit. For Sikhs, Punjabi is the most popular for names, followed by Hindi and then Sanskrit. A different language is sometimes used because of Sikh migration from Uttar Pradesh.

The semantic domains in Indian Thais Naming are composed of 3 domains: the semantic domain of Hindu's names is up to entities, abstract and events. The semantic domain of Hindu's surnames is up to entities, abstract and events. The semantic domain of Sikh's names is up to abstract, entities and event. And the semantic domain of Sikh's surnames is up to abstract, entities and event. There is not relational because the

relational is the grammatical words. This word is not identifying an individual as content words.

## **8. Conclusion**

The results of the data analysis are consistent with the concept of Nida (1975: 174). The result exposed that a semantic sector comprises basically of a number of meanings which share certain semantic elements of each group. The salient of the semantic domains are religious characters. Hindu Indian Thais people named entities (natural and animals) that related to the main concepts of Hinduism. Values about holy God, spirit and cosmos are important to a person's lifestyle. Sikhism respects a monotheistic faith. Sikhism identifies God as the only One. God is the Designer of the Galaxy, courageous, not topic to time or area, and does not take beginning or die. It reflects that both of groups still believe firmly own philosophy of region.

## **Acknowledgements**

The research presented here was financially supported by Thailand Research Fund (TRF) through the Royal Golden Jubilee Ph.D. Program (Grant No. PHD/0341/2550). I would like to express my gratitude to Assoc. Prof. Sophana Srichampa, my dissertation advisor, for her kind support in my research. In addition, it gives me great pleasure in acknowledging the support and help of Assoc. Prof. Amarjiva Lochan, an expertise in Indian studies, Prof. Suwilai Premrirat, whose enthusiasm for linguistics conventions had lasting effect. My deep gratitude is also due to the generosity of Assoc. Professor Amon Sangmanee, and Assoc. Prof. Samniang Leurmsai, whose guidance is the key to my success. A very special gratitude needs to be given to my Indian Thai informants who dedicated their valuable time and patience during the data collection. Many thanks must be given to Thailand Research Fund who granted me the Royal Golden Jubilee Scholarship throughout the research process and provided me several noteworthy opportunities.



## References

- BROWN, Robert (1991) *Ganesh: Studies of an Asian God*. Albany: State University of New York.
- COLE, O. & SAMBHI, A. K. (1995) *The Sikhs: Their Religious Beliefs and Practices*. Sussex (England): Sussex Academic Press.
- CÆDÈS, G. (1968) *The Indianized State of Southeast Asia*, Kuala Lumpur: University of Malaya Press.
- DHIRAVEGIN, Likhit (2008) *Coalition Politics, Political Reform and Democratic Development in Thailand*. Bangkok: Thammasat University.
- DONIGER, W. (1999) *Splitting the Difference: Gender and Myth in Ancient Greece and India*. The 1996 Jordan Lectures. Chicago and London: University of London Press and University of Chicago Press.
- FERGUSON, R. (1864) *Teutonic Name-System: The Teutonic name system applied to the family names of France, England & Germany*. Carlisle: R. & J. Steel.
- JAMES, L. (1975) *Notes on ethnosemantics*, Honolulu: University of Hawaii.
- KAUSHIK, D. K. (2000) *Cataloguing of Indic names In AACR-2*. New Delhi: Low Price Publications.
- KLOSTERMAIER, K. (2007) *A survey of Hinduism*. New York, New York: State University of New York Press.
- MPHANDE, L. (2006) *Naming and linguistic Africanisms in African American culture. Selected Proceedings of the 35th Annual Conference on African Linguistics: African Languages and Linguistics in Broad Perspectives*. Somerville, Mass: Cascadilla Proceedings Project.
- NIDA, E. A. (1975) *Componential analysis of meaning: an introduction to semantic structure*. New York: Mouton Publishers.
- OFFICE OF THE DEFENCE ATTACHE, ROYAL THAI EMBASSY (2011) *Indian Migrants*. New Delhi: Government Printing Office.
- ROSENTHAL, D. (2005) *Conscious and Mind*, England: Clarendon Press.
- THAPAR, R. (1973) *Asoka and the Decline of the Mauryas*, Oxford (England): Oxford University Press.

Received 7 December 2015.

Accepted 27 April 2016.

## ON DOCUMENTING LOW RESOURCED INDIAN LANGUAGES INSIGHTS FROM KANAUJI SPEECH CORPUS

Pankaj DWIVEDI & Somdev KAR

Indian Institute of Technology Ropar\*

pankajd@iitrpr.ac.in / somdev.kar@iitrpr.ac.in

### Abstract

Well-designed and well-developed corpora can considerably be helpful in bridging the gap between theory and practice in language documentation and revitalization process, in building language technology applications, in testing language hypothesis and in numerous other important areas. Developing a corpus for an under-resourced or endangered language encounters several problems and issues. The present study starts with an overview of the role that corpora (speech corpora in particular) can play in language documentation and revitalization process. It then provides a brief account of the situation of endangered languages and corpora development efforts in India. Thereafter, it discusses the various issues involved in the construction of a speech corpus for low resourced languages. Insights are followed from speech database of Kanauji of Kanpur, an endangered variety of Western Hindi, spoken in Uttar Pradesh. Kanauji speech database is being developed at Indian Institute of Technology Ropar, Punjab.

### Keywords

speech corpus, Kanauji, language documentation, endangered language, Western Hindi

## DOCUMENTACIÓN DE OBSERVACIONES SOBRE LENGUAS HINDIS DE POCOS RECURSOS A PARTIR DE UN CORPUS ORAL DE KANAUJI

### Resumen

Los corpus bien diseñados y bien desarrollados pueden ser considerablemente útiles para salvar la brecha entre la teoría y la práctica en la documentación de la lengua y los procesos de revitalización, en la

---

\* Indian Institute Of Technology Ropar (IIT Ropar), Rupnagar 140001, Punjab, India.

construcción de aplicaciones de tecnología lingüística, en la prueba de hipótesis lingüísticas y en muchas otras áreas importantes. El desarrollo de un corpus para una lengua con pocos recursos o amenazada encuentra varios problemas. El presente estudio se inicia con una perspectiva general sobre el papel que los corpus (los corpus orales en particular) pueden desempeñar en la documentación del lenguaje y en los procesos de revitalización. A continuación, ofrece una breve reseña de la situación de las lenguas amenazadas y de los esfuerzos de desarrollo de corpus en la India. Posteriormente, se analizan las diversas cuestiones relacionadas con la construcción de un corpus oral para lenguas con pocos recursos. Las observaciones se extraen de la base de datos de Kanauji de Kanpur, una variedad amenazada del Hindi occidental, hablada en Uttar Pradesh. La base de datos del discurso de Kanauji se está desarrollando en el Indian Institute of Technology Ropar, de Punjab.

### Palabras clave

corpus oral, lengua kanauji, documentación de la lengua, lenguas amenazadas, hindi occidental

## 1. Introduction

The old quip attributed to Uriel Weinreich, that a language is a dialect with an army and a navy, is being replaced in these progressive days: a language is a dialect with a dictionary, grammar, parser and a multi-million-word corpus of texts –and they would better all be computer tractable. When you have got all of those, get yourself a speech database,<sup>1</sup> and your language will be poised to complete on terms of equality in the new Information Society (Nicholas Ostler *apud* Borin 2006).

What Nicholas Ostler meant here is that a language will not be able to survive in the today's world, the world which looks as if it has turned into a technological park, if it is not used in language technology applications. Mentioning the importance of speech database, Ostler says: "...parser and a multi-million-world corpus of texts [...] get yourself a speech database" (p. 317). Trosterud (2006: 293) mentions "Languages may

---

<sup>1</sup> In this paper the term 'speech corpus' refers to an organized collection of spoken data of a language in a way so that it is used as a base for the various purposes ranging from phonetic analysis to multipurpose speech applications. Hence, the corpus will further need to be customized depending upon the focus of inquiry. It should be taken as a step, which, if appropriately dealt with, can act as platform for product development. Therefore, the terms speech or spoken corpora, speech database and digital spoken archives are more or less equivalent in nature.

live on without orthography. But no language will be able to function as administrative language in a modern society without a developed language technology applications”.

First half of the 19<sup>th</sup> century saw a gradual shift from rule-based approach to corpus-based approach. Before it, a linguist’s typical job involved to sit at a table and think of a language. Until 1950s it was realized that it is easier to formulate a hypothesis, compare results and perform a recheck if you have larger data at your disposal, which grew into concept what we today know as corpus linguistics. However, use of electronic corpora was not very clear in the picture till then. The first motivation in the field of electronic corpora came from the work of Jesuit priest Roberto Busa, who created an electronic lemmatised index of the complete works of St Thomas Aquinas, *Index Thomisticus*, beginning in the 1950s and completing it in the late 1970s (Tognini-Bonelli 2010). Since then both electronic text and speech corpora came to be used widely in all areas of language research including language documentation.

Language Corpora, including archives, can play a significant role in the documentation of endangered languages in following ways: a) help to preserve the dying languages for the future generations; b) facilitate the use of primary materials, such as filed notes, audio and video recordings; and thereby helping language maintenance and revitalization; anthropological, typological, historical and comparative studies; c) provide a platform for the products and deliverables not only to the researchers but also to the communities speaking endangered languages.

## **2. Language documentation and corpus linguistics**

Documentation of a language often ends with a lot of material as it aims to collect the examples of the full spectrum of language forms and uses that the language community employs (Johnson 2004). Till the first quarter of the 20<sup>th</sup> century, language documentation meant collection of data on paper and therefore each documentation project resulted into large bodies of texts, that is, text corpora. Task such as dictation, transcription, translations, elicitation, analysis, etc. were all done only on papers. Linguists and other language researchers such as anthropologists worked painstakingly

to collect and preserve these collections of texts. However, lately, technology entered into the scene and revamped the entire process all through. It provided the documentary linguists with the useful tools and resources leading to better organization and long-term preservation of the language data apart from significantly reducing the manual efforts. It became possible to store the text data in electronic form using floppy disks, CD ROMs, cassettes etc. According to Bird & Simons (2003) the process of documenting and describing the world's language is undergoing radical transformation with the rapid uptake of new digital technologies for capture, storage, annotation and dissemination. With the further advancement in the technology, linguists could also make good quality recordings of the spoken from itself and store them very longer time and thereby making spoken corpus.

The best thing about a carefully designed corpus is that one can perform almost any kind of linguistics without the help of language consultants/informants or any linguistic fieldwork. Large collection of language data lets you formulate a potential hypothesis and later to cross verify it. Corpus linguistics as well as language documentation share many common points of interaction. For example, one of the major points of interaction is data collection. McEnery & Ostler (2000: 410) define corpus linguistics as a methodological toolkit dealing with construction and analysis of consistent collection of data, whereas Johnson's (2004) take on language documentation is 'effort to produce permanent and reusable *collections of diverse linguistic data*'. Cox (2009) opines that corpus linguistics and language documentation complements each other in the following ways:

For corpus linguistics, language documentation offers a diverse and well-catalogued data, and 'raw material' for corpus construction for under-represented languages-which is a standing challenge to corpus linguistics, whereas for language documentation, corpus linguistics offers a methodological perspective on the documentary record and a set of new tools for analysis. To language documentation corpus linguistics also offers another means of rendering the documentary record available, both to academic and non-academic communities (Cox 2009: 4).

### 3. Language endangerment

Language endangerment is a serious concern to which linguists and language planners have turned their attention in the last several decades. The famous linguist and professor David Crystal (see Jansen & Sørensen 2005)<sup>2</sup> says “When a language is lost, a vision to the world is lost”. For a variety of reasons, speakers of many smaller, less dominant languages stop using their heritage language and begin using another. For example, parents begin to use only the second language with their children and gradually the intergenerational transmission of the heritage language is reduced and may even cease. As a consequence there may be no speakers who use the language as their first or primary language and eventually the language may no longer be used at all (Paul, Simons & Fennig 2013). Crystal (2000) claims that the rate of language disappearance is as high as two languages each month (cf. Allwood 2006). Larger languages are held responsible on the endangered situation for the other languages. Writing about as to how English is devouring other smaller languages, McWhorter (2009) writes:

There are about 20 languages that are slowly eating up the other 6,000. That’s essentially because of how England developed a global presence starting in the 1600s, and the language they happened to carry with them was English. What we’re seeing is an increasingly Anglophone world and an increasingly oral, rather than written world. So many of the other languages are falling by the wayside that we may lose 90 percent of the languages we have now by the year 2100 (John McWhorter *apud* Stephenson 2013).

### 4. Language endangerment in India: an overview

*Times of India* (TOI, August 9, 2013), a reputed English newspaper, reads (via People Linguistic Survey of India (PLSI) conducted by Bhasha Research & Publication

---

<sup>2</sup> <http://www.dfi.dk/Service/English/News-and-publications/FILM-Magazine/Artikler-fra-tidsskriftet-FILM/60/Voices-of-the-World-Language-is-the-House-of-Being.aspx>

Centre, Gujrat, India) that India currently speaks more than 880 languages. Up till 1961, about 1,100 languages would be spoken in India. Of these, 220 languages have already disappeared in the last 50 years, that is, we have lost about 20% of total languages spoken in India during the period. TOI further reports that more than 150 out of 880 languages currently spoken are going to become extinct in the next 50 years, that is, about 17.05% are going to die. The situation seems to be awfully alarming.

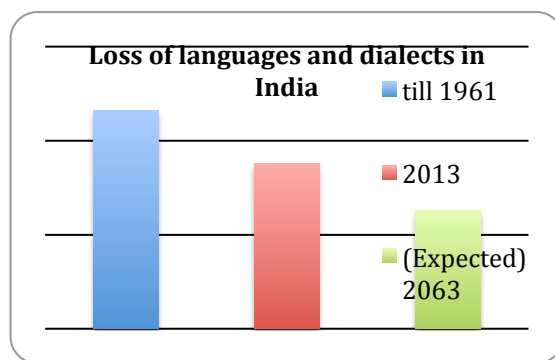


Figure 1. Loss of languages and dialects in India

According to 2001 census and language report,<sup>3</sup> 122 languages have been recognized as by the government. Of these 122 languages, 22 languages<sup>4</sup> have been classified as scheduled languages and 100 languages<sup>5</sup> have been categorized as non-scheduled languages. Statement 9 of the report reads that 99.82% of the population, i.e. 1,028,610,328, recognize their mother tongue among of these 122 languages. It further adds that only 0.17%, i.e. 1,762, 388, of the total population speaks other languages or dialects. These languages were not identifiable for they were returned by less than 10, 000 people at all India level as their mother tongue. If we assume that no language loss has occurred between 2001-2013, 768 languages were spoken by only 0.17 of the population of the India in 2001. And if we take PLSI statistics in account, about 32 languages are dying per decade in India. About 790 languages were spoken by merely

<sup>3</sup> Latest census was done in the year 2011 by the govt. of India. However, the govt. released no data pertaining to the no. of language spoken in India. According to news and reports, govt. plans to conduct a separate linguistic survey of the country.

<sup>4</sup> In the eight schedule, following languages have been recognized as scheduled languages:- Assamese, Bengali, Bodo, Dogri, Gujarati, Hindi, Kannada, Kashmiri, Konkani, Maithili, Malayalam, Manipuri, Marathi, Nepali, Oriya, Punjabi, Sanskrit, Santhali, Sindhi, Tamil, Telugu and Urdu.

<sup>5</sup> List of non-scheduled languages can be retrieved from <<http://censusindia.gov.in/>> (accessed 9th August 2015).

0.17% population of the country in 2001, which is extremely alarming for linguistic diversity of a country such as India.

There are various interrelated factors which together lead to endangerment of a language. Ethnologue uses Expanded Graded Intergenerational Disruption Scale (hence EGDIS) to estimate the vitality of the languages in the countries where they are primarily spoken. For India, Ethnologue lists following languages in the category of ‘Threatened’ (group-1), shifting (group-2), Nearly Extinct (group-3), and Dormant (group-4) on EDGIS scale:

Group I: Aiton, Allar, Andh, Apatani, Aranadan, Bantawa, Byangsi, Chamling, Chaudangsi, Chin-Bawm, Darmiya, Dubli, Gadaba, Bodo, Gahri, Gata’, Godwari, Gurung-Western, Hruso, Indo-Portuguese, Kachari, Kadar, *Kanauji*, Khamti, Koraga, Korra, Koraga-Mudu, Koro, Kui, Kulung, Kumbaran, Kupia, Kurichiya, Magar-Eastern, Mahali, Mal Paharia, Malavedan, Mannan, Nihali, Öñge, Pardhan, Powari, Sentinel, Thachanadan, Thangmi and Vishavan. Group II: A’tong, Bazigar, Bellari, Majhi, Majhwar, Manna-Dora, Ralte, Rawat, Sansi, Yakkha, and Zakhring. Group III: Turi, Great Andamanese, Khamyang, Nefamese, Parenga, and Ruga. Group IV: Malaryan, Rangkas, Ullatan, Urali, Ahom, Pali.

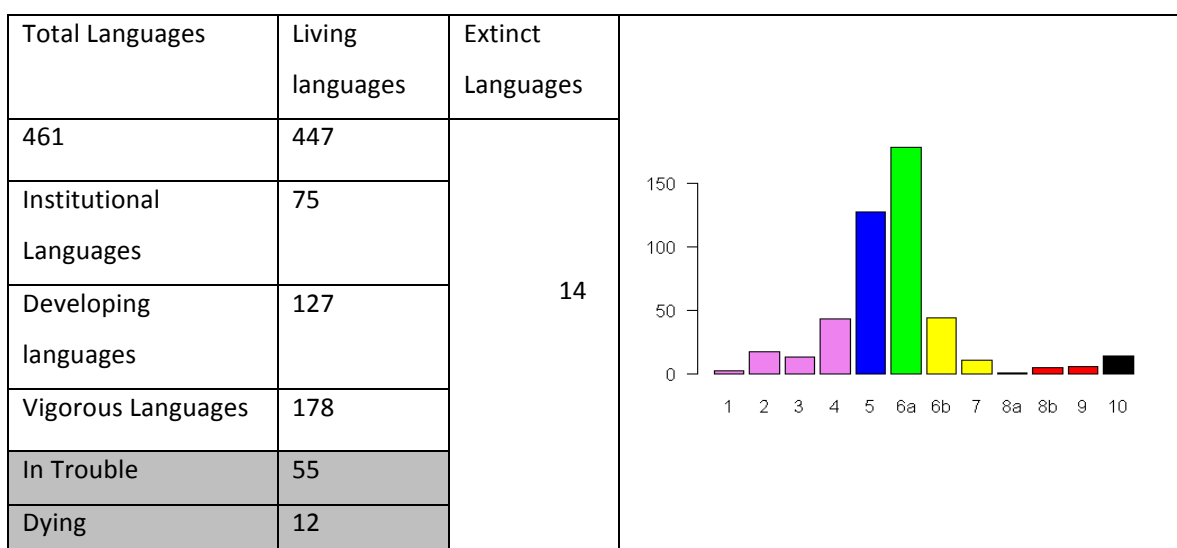


Figure 2. Language vigor based on EDGIS scale (Source: <http://www.ethnologue.com/country/IN>)

This graph shows status of language endangerment vs. language development in India. The horizontal axis represents the estimated level of development or



endangerment as measured on the EGIDS scale. The height of each bar indicates the number of languages that are estimated to be at the given level.

## 5. Corpora development efforts for Indian Languages

In India considerable progress has been made in terms of the generation of the text corpus with reference to Indian languages, this is true only for the widely spoken languages however. And work on spoken corpora is still in its primary stage even for the widely spoken above languages. Work on lesser-spoken languages is still in its infancy even for text corpus, leave alone speech – it is true at least in the open literature. If we compare the number of corpora with number of endangered languages in India, situation seems to be gloomy. And of course many of these corpora, especially speech corpora, are still in their raw form.

Excluding some foreign bodies, some of the main Indian organizations that are engaged in the corpora work for Indian languages are: CIIL Mysore, Karnataka; TDIL, New Delhi; IIIT Hyderabad; IIT Kanpur; IIT Mumbai; IIT Guwahati; ISI Kolkata; JNU, New Delhi; University of Hyderabad, Hyderabad; Annamalai University, Chennai; MS University, Baroda; Thapar University, Patiala; Punjabi University, Patiala; ER&DCI, Trivandrum; C-DAC, Pune & Noida; Utkal University, Bhubaneswar.

Available Text Corpora: Assamese, Bengali, Hindi, Marathi, Punjabi, Urdu, Tamil, Telugu, Oriya, Malayalam, Kashmiri, Kannada, Gujarati, Maithili, Manipuri, Indian English, English, Sanskrit, Nepali, Konkani, Kodava, Yarava, Dogri, Bodo.

Available Speech Data: Assamese, Bengali, Hindi, Marathi, Punjabi, Urdu, Tamil, Telugu, Oriya, Malayalam, Kashmiri, Kannada, Gujarati, Maithili, Manipuri, Some varieties of Indian English, Nepali, Konkani, Kodava, Dogri, and Bodo.

## 6. Planning data collection and corpus design

Before starting speech data collection process for any endangered language, it is advisable to keep following points in consideration:

a) *Level of endangerment*: If linguistic variety in question qualifies to III group – which means it, is extremely endangered and almost beyond revival – then it is necessary to record as much data as possible. Quantity becomes very important in such cases. A little delay in such cases can leave you with void that will never be filled.

b) *Speaking population*: How many people speak this language? As life of a language is usually directly proportional to the number of its speakers. Is this language being passed on to the younger generation or has become limited to the older ones.

c) *Embedded foreign language*: An endangered or under-resourced language is likely to be influenced by one or more neighboring linguistic varieties. It is more often the variety, which is considered as standard or institutionalized, and is used for the purpose of imparting education in schools and other modes of official communication.

d) *Orthography*: Whether the language has its own script for writing or it is written in some other language's orthography?

e) *Literature*: Is there some popular literature on this linguistic variety? If yes then is it available in form of manuscripts? These manuscripts may be available with the local administration, libraries, and museums, or with the town's head, or with the person/people who oversee religious ceremonies.

f) *Presence over the World Wide Web*: Do you know if there are some websites, blogs, e-portals, etc. which are exclusively written in this language or dedicated to the people using this linguistic variety.

f) *Versatility of use*: For corpus to be fully representative of the language, versatility of communication is another point to be kept in mind. Since an endangered language is usually limited in its domains and scope of use, you may have to provide a little training to your language consultants on what you want them to speak during recording.

Kanauji is under-resourced<sup>6</sup> variety of Hindi and falls within the group I – which signifies that it is rapidly going out of use. We collected 18 hours of speech data for it. Kanauji, including its all sub-dialects, has approximately six million speakers. Hindi is so embedded with Kanauji for younger generation that it does not even realize if there is any difference between Hindi and Kanauji. Most of them confirmed that they did not even know if they spoke Kanauji. The level of mixing may differ from one area to another depending on its proximity to city and towns, exposure to the education, economic status of people, etc. Like other varieties of Western Hindi, Kanujai uses Devnagari script. There is little literature available in Kanauji language; only to mention of mid 17<sup>th</sup> century authors, from Tikampur/Tikawanpur town of Kanpur district, such as *Chintamani Tripathi, Matiram Tripathi, Bhushan Tripathi and Nilkanth Tripathi* (see Keay 1920; Upadhyaya 1934). Folk Kanauji songs, which are based on famous valor stories of two brothers named Alha and Udal, are sung during festive seasons (Russell 1916). Kanauji has little presence over web. Some blogs and website are found which either talks about the history and culture of the region. Nearly all younger Kanauji speakers speak Hindi in their day to day communication with persons outside of their family.

For our corpus we collected the data from 12-15 domains of daily common use of language (Table 1). The Consultants' age, education and their contact with the other languages are also very important factors to be considered, as they together significantly influence the choice of vocabulary, mixing of codes and pronunciation factors. Most often young and educated speakers tend to use the language (pronunciation and words) according to neighboring standard form of the language. In contrast old and uneducated speakers usually retain the original form of the language.

No.	Domain	Explanation
1	Basic wordlist	equivalent words from Swadesh list and variety of domains.
2	Demographic description	regarding population, educational status, income etc.
3	Cuisine	relating to various dishes and cooking procedure
4	Family communication	day-to-day family communication among family members
5	Games	games played by children, teens and adults
6	Culture and Traditions	temples, festivals, marriages, ceremonies etc.

<sup>6</sup> Under-resourced languages, in our case, are the languages which don't have a wide presence on the web and/or don't have sufficient speech transcribed data.

7	Flora and Fauna	farmlands, gardens, types of crops, seasons etc.
8	Mythological stories	mythological stories, local deities, beliefs etc.
9	Daily life activities	usual working of children, teens and men and women
10	Children stories	popular stories among the children
11	Number systems	number systems of Kanauji
12	Free discourse	free discourse from variety of life activities
13	Minimal pairs	minimal pairs
14	Representative sentences	sentences providing information about tense, gender & aspect
15	Group conversation	group conversation on a given topic

Table 1. List of recorded items

The participants who were recorded had Kanauji as their first language/mother tongue. Since, like it happens with speakers of an endangered language, younger generation (12-25 years) is rapidly shifting to official standard variety, i.e., Hindi, most of them grow to become fluent bilingual. In contrast, older people (40-60 years) preserved the original form. Mid-aged people (25-40 years), as expected, fell somewhere between these two extremes. The factors that catalyze this process are education, employment and biased linguistic attitude. Information regarding these factors has been summarized in the given Table 2.

Male 15	Female 15	Age	Education	Bilingualism (Kanauji-Hindi)
8	8	12-25 years	10 <sup>th</sup> -undergraduate	Fluent bilingual
5	5	25-40 years	Illiterate to undergraduate	Intermediate to Fluent
2	2	40-60 years	Illiterate	Mostly monolingual, speaking Kanauji only

Table 2. Language consultants based on age, education and gender

Here are three samples extracted from the corpus that can give you the clue as to how Kanauji is getting overshadowed by the Hindi.

Audio transcription of the samples is done in Devanagari and each sentence is enclosed in a pair of angle brackets; IPA transcription of the sample is given and English translation of the sample is provided. Instances of three dots (...) between the pair of angle brackets <>...<> refer to the fact there are one or two sentences in between the

presented sentences have been omitted for they did not fit our purpose and also to save space. The code 'bjj and Hindi' are ISO 693-3 codes for Kanauji and Hindi, respectively. Its use in the beginning and end of the sample marks that entire conversation (syntax) has been in Kanauji.

### Sample 1: Record Entry No.121215-03

In the first sample, language consultant 1 is a fifty-eight years old man who sells milk for the livelihood and language consultant 2 is a 38 years old man who runs a small shop in the village. Except for some occasional visits, both of these men usually don't go out of village. Language consultant 1 narrates a personal story about his ghost experience in a recording session with data collector. Language consultant second 2 provides some inputs as an active listener. Only few sentences have been presented here for our purpose.

#### Language Consultant 1:

< bjj < पीपा लदो दूध क्यार >  
<bjj<pipa ləḍo dūdh<sup>h</sup> kyar>  
< Can of milk was loaded.>

#### Language Consultant 2:

<bjj <दूसर कउनो आदमी हात तो फेंक के भाग ठाढ़ हाता।> <कोउ न रुकी,कि कोउ पकल लीन्हेस> bjj>  
<bjj < dūsər kəʊno aḍəmi hwaṭ̪ t̪o pʰɛk ke bʰag t̪ʰaɽʰ hwaṭ̪ >> k̪oʊ n̪ə r̪u:ki, ki k̪oʊ pəkəl li:nh̪es > bjj>

If there was someone other than you, he would run away. No one would dare stop thinking that something could have caught him.

#### Language Consultant 1:

<bjj < पीपा उठाओ औ भुँई मा सारे का पटको।>> < तौ, इस्टैंड (बब्बी) तरे का चलो गा > bjj>  
<bjj < pipa utʰao ɔ bʰu:ĩ ma sare ka pət̪ko >> t̪o, ɪst̪ɛnd̪ t̪ərə ka t̪ʰəlo ga > bjj>  
< (I) raised the can of milk and turned him down to earth.> < Then, stand (of bicycle) went down.>

In this piece of conversation, both of the language consultants fluently use Kanauji without any hint of Hindi coming in their way of communication. Only the word “/ɪstənd/ - <stand>” is borrowed from English via Hindi into Kanauji and has got adapted phonologically adapted to Kanauji by inserting a schwa between /d/ and /m/ sounds. This insertion is due to the fact that Kanauji disfavors st\* cluster, a vowel is inserted in the beginning of the word to break the cluster. Words like / b<sup>h</sup>ɔ̃ / and /t̪<sup>h</sup>aɾ<sup>h</sup>/ are typical to Kanauji vocabulary and not found in Hindi.

### Sample 2 (Recording Entry No. 121211-0024)

This sample is from a twenty years old girl who is in the first year of her university education. Data collector asks her to narrate him a story in Kanauji that she stills remembers from her childhood textbooks.

<bjj <एक राजा राहे> <तीन रानियाँ राहे>...<राजा सिकार ख्यालें जात राहे>...<मतलब आइटम बना-बना खवाओ करिबे>...<उई सुन लीन्हेन, उई तीनो ना र.रानि से सादी करि लीन्हेन>... <तीनो रानिन से जब सादी करि लीन्हेन तो उनके एक लड़का औ एक बिटिया भे> bjj>

<bjj <ek radʒa rahe>. <tin raniġā rahē>...<radʒa sikar k<sup>h</sup>jalē jaṭ rahē>...<mətləb aiṭəm bāna-bāna k<sup>h</sup>əwao kəribe>...<ɔi sʊn lin<sup>h</sup>en, ɔi t̪ino na rə ..rani se saḍi kəri lin<sup>h</sup>en>. ...<t̪ino ranin se dʒəb saḍi kəri lin<sup>h</sup>en t̪o unke ek ləɾka ɔ ek biṭija b<sup>h</sup>e> bjj >.

<There was a king> <There were three queens>... <The king was on his way to hunt>...<Means, (I would) have (him) different types of cuisines prepared to eat>... <He heard this and married all three queens>...<After he married all three queens, one baby-boy and baby-girl were born to king>

This sample shows how Hindi words such as nouns (e.g., /sɪkar/ ‘hunting’, /saḍi/ ‘marriage’, /rani/ ‘queen’, ləɾka ‘boy’), numbers (/ek/ ‘one’), verbs (sʊn ‘to hear’), conjunctions (se ‘with’, t̪o ‘then’); items from English (aiṭəm ‘item’) and Urdu (mətləb ‘meaning’), etc. are frequently borrowed and used by young Kanauji speakers. In some instances borrowed items are phonologically adapted; others are borrowed in their

original forms, however. For example, IPA transcriptions for Hindi words ‘marriage’ and ‘hunting’ is /ʃaḍi/ and /ʃikar/, respectively. But in these words /ʃ/ is replaced by /s/ since /ʃ/ is not found in native Kanauji phonemic inventory. However, Hindi conjunction /t̪o/ ‘then’ and noun /l̪əɾka/ ‘son’ are adopted in their original forms despite of the fact that corresponding words for the same meanings in Kanauji have very similar phonotactics – /t̪o/ and /l̪əɾika/, respectively. One striking feature is that apart from the borrowing of lexical items from Hindi, a morphological process of ‘pluralization’ is also seemed to be borrowed. For example, plural of /rani/ is /ranin/ in Kanauji in contrast to /ranijā/, which is a Hindi plural form of /rani/.

### Sample 3: Record Entry No. 12121-001

In this sample, language consultant is 28 years old man who works as a “Pandit (a kind of priest)”. He often has to visit city and towns for performing rituals. On being asked about his nature of work and how he makes money for living. Language consultant shares some details:

<bjj <जइसे हमारा जजमान है> <जइसे तुम हमार आदमी हो> <तुमका कुछ करवावें का है तो तुम हमते बात कीन्हेव> तुमका कुछ बता दीन अगर कि तुम अपन आदमी हो> ... <दूसर आदमी आओ, पइसे वाला है – ठीक ठाक>

bjj>

<bjj<ɟəɪse h̄amara ɟəɟman h̄e> <ɟəɪse t̪om h̄amar aɟ̄ami ho> <t̪omka kut̪ʰ k̄arwawē ka h̄e t̪om h̄am̄t̪e baɟ̄ kin̄ʰew>. <t̪omka kut̪ʰ b̄əɟ̄a ɟ̄in əḡər kɪ t̪om əp̄ən aɟ̄ami ho>... <duɟ̄ər aɟ̄ami ao, p̄əise wala h̄e - t̪ʰik t̪ʰak> bjj>

<Like it’s my customer, like you are my man>. <And you want something to get done> <So, you talked to me>

<I would already tell you the things if you are an acquaintance>...<in case an unknown person visits and he is rich enough>...

Giving a closer look to this piece of conversation reveals that a few words and phrases are either borrowed from Hindi or are those which are very close to Hindi. For example, Hindi phrase /h̄amara ɟəɟman h̄e/ is used in place of Kanauji phrase /h̄amar ɟəɟman aje:/. Words like /aɟ̄ami/ and /kut̪ʰ/ and eco-reduplicated form like /t̪ʰik-

t<sup>h</sup>ak/ is also from Hindi. This conversation does not have a word that is exclusive to Kanauji and not found in Hindi some modified form.

**Sample 4: Record Entry No. (121215-0031 and 121214-0028)**

This sample is from a 15 years language consultant who studies in high school. Like any other children of the current generation, girl has come out be fluent bilingual speaking Hindi and Kanauji. While children learn Kanauji from their parents and society, Hindi is the main language of education and cinema. English is also taught in the schools but due to lack of proper training, most of them don't gain more than basic skills. In this piece of recording, language consultant is first asked to tell something about her family. Language consultant answers in Hindi. Then, language consultant is asked to share some story that she could have ever heard from her grandmother. And language consultant retells an entire story in Kanauji without a hitch. Only few sentences from both parts have been taken for our purpose.

**(Part-I, introduction)**

< hin <हमार नाम X<sup>7</sup> है > <हमार मम्मी का नाम Y<sup>8</sup> है> <पापा का नाम है Z<sup>9</sup>> <हमारे पड़ोस में दादी दुकान करती हैं। बाबू भी दुकान में दुकान में रहते हैं। और लक्ष्मी दीदी पुलिस में नौकरी करती हैं > hin>

<hin < hāmar nam X hε> <hāmar mām̄mi ka nam Y hε> <papa ka nam hε Z> <hāmare pəɾos me dād̄l̄ d̄ukan kəɾt̄l̄ h̄ε̄> <babu b<sup>h</sup>i d̄ukan me rəh̄t̄e h̄ε̄> <ɾ lək̄ʃmi d̄īd̄l̄ p̄ulis me n̄ok̄əri kəɾt̄l̄ h̄ε̄>hin>

<I am X> <My mother's name is Y> < Father's name is Z> < grandmother runs a shop in neighborhood> <grandfather also contributes in it> <sister Laxmi has a job in police>

<sup>7</sup> Personal information of language consultants such as their first name, family name, names of the place they work or study at has been kept confidential.

<sup>8</sup> Personal information of language consultants such as their family members, relatives, etc. has been kept confidential.

<sup>9</sup> Personal information of language consultants such as their family members, relatives, etc. has been kept confidential.



**(Part II, storytelling)**

< **bjj** < एक भाई, एक बहन रहें> <तो बहन कि पहले से सादी होइ गे राहे> <भाई छोटें रहें> <तो उई एक बार काहत हैं कि अम्मा बहिनी से हम मिलि आवन> <तो अम्मा कहेन कि बउआ अबै रुकौ कहे कि तुम्हरी बहिनी तो बहुत दूर राहती हैं > **bjj** >

<**bjj**<ek b<sup>h</sup>ai, ek bəhen rahē>. <to bəhen ki sadi pəhle se sadi hoī ge rahe>. <b<sup>h</sup>ai ʈʰoṭe rahē>. <to ʊ ek bar kahət hē ki əmma, apni bəhini se hām milī awən> <to əmma kəhen ki bəʊa abe rʊkə kəhe ki tʊm<sup>h</sup>ri bəhini tʊ bəhoṭ dʊr rahti hē>**bjj**>

<There was a brother and a sister> <Sister was already married> <And brother was younger one> <So, once he asked his mother if he can go and meet his sister> <Mother said “Dear Lad, not now as you sister lives far from this place”>

In ‘Part-I, introduction’ the girl fluently speak in Hindi; while in ‘Part II, storytelling’ she goes on speaking in Kanauji without a hitch. Girl, however, code-mixes some Hindi words and phrases such as ‘ek b<sup>h</sup>ai, ek bəhen’, ‘se’, ‘to’ and ‘dʊr’ like modern generation of Kanauji.

**7. Recording specifications**

Speech data was collected using two channels simultaneously: first, by Olympus LS-100 96kHz/24 PCM linear recorder which is outfitted with two built-in 90° stereo condenser microphones, second using Sony Digital Flash Voice Recorder (ICD-PX312). The major specifications of the recorder are listed below in Table 3.

Channel	Frequency response	Sensitivity	SPL	Pick-up pattern
Ch-1	20-20000Hz	high	140dB	Cardioids
Ch-2	75-20000Hz	mid	--	Mono

Table 3. Recording specifications

## **8. Speech data recording**

The language consultants were asked to maintain an approximate distance of 10-14 inches from the microphones and keep their speaking style consistent and natural. The speech was digitized at a sample rate of 44.1 KHz/24 bit. It was all stored in wave format. Since speech data collection is far more sensitive than text collection process, continuous monitoring is required for any unaccepted deviation in pronunciation, styles, rate of speech, etc. However, in this case, data collector himself was native speaker of Kanauji and therefore it was easy to keep track of the mistakes. An onsite cursory review of the recordings was performed however. As mentioned earlier, collecting maximum amount of speech data should be the primary goal while working for an endangered language. It has two benefits: first, it provides us with the potential possibility of the phonetically rich data (covering the frequency count of the rarely occurring phones), second, more data usually cover the more diverse instances of linguistic events. Corpus has both types of speech data, i.e., read and spontaneous. Read speech data comprises of grammatical sentences, whereas spontaneous data is full of grammatical mistakes and speech disfluencies. Therefore, the corpus can further be customized for different goal oriented speech applications.

## **9. Establishing phoneme inventory**

Establishing phonemic inventory is one of the most important things for a speech corpus. Depending on the requirements the researcher may choose biphones, triphones, syllables, words or sentences as basic speech unit. However, Indian languages are syllable centered, where pronunciations are mainly based on syllables and therefore it can be the best unit for Indian Language Speech Synthesis (Kiruthiga & Krishnamoorthy 2012).

Vowels in Kanauji are also represented by glyphs and consonants to get combined to form ligatures. Its phonemic inventory consists of 53 phones: 33 consonants and 20 vowels. Hence, the ideal number of possible biphones and triphones should be 2809 and

148877, respectively. However, due to phonotactic constraints this number will be considerably lower. A comparison is given below in Table 4:

Phones	53		
Biphones	53x53	2809	< 2809
Triphones	53x53x53	148877	< 148877
Syllables	Not counted	Not counted	
Words	Not counted	Not counted	

Table 4. Analysis of phones in Kanauji

## 10. Phonetic transcription and annotation

The phonetic characteristics of a speech corpus play a key role in the robustness of the future speech corpus (Stănescu *et al.* 2012), and therefore phonetically transcribed text is needed to provide the complete phonetic coverage (Murtoza *et al.* 2011). Transcription is often achieved with the help of some phonetic dictionary, but, in case of unavailability of resources, one needs to design the appropriate automatic grapheme-to-phoneme converter. Task of phonetically balanced sentences cannot be achieved unless every single phrase is phonetically transcribed.

The unedited speech data Kanauji lasts for more than 16 hours. Speech data contains instances of repetitions, code mixing and switching, and another non-speech sounds such as laughter, coughing, background noise, etc. It requires a manual cleaning and editing. As it is presupposed, data tends to reduce by several hours. Annotation is supposed to be of two types: basic and extensive. The basic annotation refers to the annotation of regular pronunciation such as boundaries of phones, syllables, words and sentences; whereas extensive annotation refers to the annotation of the actual speech. Since, work on transcription and annotation is still to be done, no conclusive remarks can be made on it. Speech files will be annotated using PRAAT in text grid files. Depending on the purpose, EAGLES, ELRA, ICAME, LDC etc. may be used as guidelines for transcription and annotation.

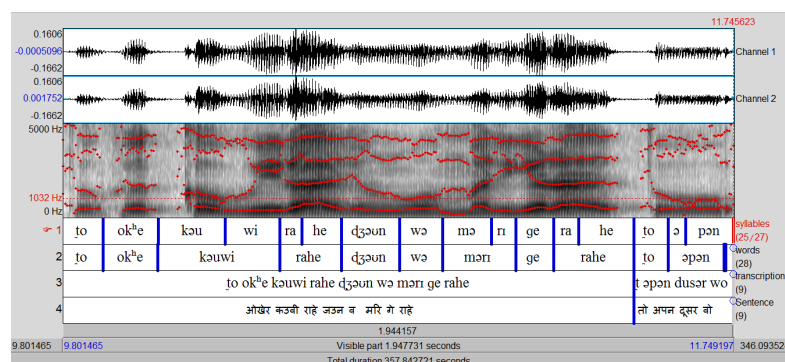


Figure 2. Annotation snapshot of Kanauji speech database

## 11. Spoken text markup

Usually text corpora are marked up using standard markup languages like XML and SGML. However, depending on the purpose behind the collection of speech corpus and availability of economic and manual resources, level and type of marking up (STML, JSML, SSML, etc.) could differ from one speech corpus to another one. While for a corpora to be used for the purpose of language teaching, a general purpose and content mark up is done; corpora to be used in speech applications require more extensive and intensive level of content markup that can cover minute features such as intonation, prosody, types of noise, pause or long pause, repetition and so on. This mark up usually has to go thorough further post editing and normalization process. Starting the general purpose markup, to be categorized as being standard, a certain level of markup is a must.

Listening to our speech database of Kanauji, we have noted down quite a few transcripts. Though, this corpus is still in its initial stage and not ready to be used in some application process, we are still exploring on what would be an appropriate level and type of markup for it. Endangerment of a language has little to do with its application markup and therefore we are not particularly engaged to provide details on it.

## **12. Advantages of spoken corpora over text corpora**

Speech is primary in nature than writing. It is ever evolving and more contemporary in nature. Speech contains far more linguistic information than that of text. Speech applications can be very useful even for the people who cannot read and write. Therefore, it becomes even more useful in case of people speaking endangered language, as they are often those who belong to backward section of the society. A good audio recording, accompanied by video, can provide articulatory and acoustic information. Speech data can be collected with minimum second party interference. Spoken corpora become more important when the language in focus is in endangered state because one may not have time to properly document it at once. The next attempt may lead to certain amount of setback as few elderly speakers may pass away. Endangered languages often don't have the written scripts and lack the written texts; speech corpora are particularly useful in these kinds of cases.

In the language teaching and revitalization process speech corpora can become a useful tool for learners' integration into the concerned speech community. Speech corpora present the people with near-to-immediate natural linguistic environment and almost authentic form of language. It can be utilized as an authentic effective tool or aid serving as a richer language resource than that of text materials. Brown & Yule (1983) claim that it is essential to give learner the opportunity to acquire some of the idiosyncrasies of the native natural speech in order to fit in to the culture of the target language.

Spoken corpus, which is robust and balanced in size, can be a very useful resource for those working on dialect study. It provides the evidence from the phonetic perspective and helps understand the first and second language acquisition theory. Evidence, as spoken form of the language provides finer nuances of a language. It also helps in the tasks such as modeling of articulation and prosody.

### 13. Challenges in constructing Kanauji spoken corpus

The spoken corpus consists of two kinds of speech samples: near-to-spontaneous and text read speech. Though we had wanted to create a phonetically balanced corpus yet things came out be different as project moved ahead. Earlier goal of the project was to record maximum amount of natural, unscripted and spontaneous speech, i.e., everyday conversations of the people. But as it happens, members of the communities were uneasy to have their personal conversations recorded in the database. Therefore, it was decided to resort to individual and group interview type of sessions and exercises. However, later on some youths agreed to let their speech recorded during their personal conversations. Some enthusiastic early age teens also had let their speech recorded. Another problem that came up was of 'NOISE'. In fieldwork, however, this is a common problem. Due to Kanauji being a variety of Hindi and unavailability of written records, transliteration of the speech was done in Devanagari script. During the process of the phonetic transcription, it appeared that there are noticeable prosodic differences in some speech samples. And these differences perhaps are due to intra-dialectal differences. However, our transcription is limited to broad level and no analysis has been done for now. As it usually happens with data collection process (especially spontaneous speech), language consultants often switched to Hindi and therefore recordings were full of the instances of code-mixing and code-switching. Sample 5 will throw a bit light on it.

#### **Sample 5 (Recording Entry No. 12121402)<sup>10</sup>**

This sample is from a 34 years man who works as a primary teacher in a school. The medium of instruction in his school is Hindi. Notational conventions are same as in Sample 1. In this piece of conversation primary teacher starts with requesting to data collector if he (data collector) can help him (primary teacher) to get a particular literary work from the city and then this teacher goes on informing the data collector about the

---

<sup>10</sup> Sample 5 has previously been used as a reference example in one of our other research reports and in a seminar presentation.

training he (primary teacher) is currently undergoing in his school. A look into the conversation (given below) tells how the teacher switches from Kanauji to Hindi and then vice versa. Recording shows us that the teacher does not even realize this shift and goes on talking in the same manner for about ten minutes.

<bjj < एत्ता काम करे जैव > <एत्ती चिन्हारी अपन देहे जैव > ...< सात साल होइगे पढ़ावत > bjj> < hin <अभी जो ट्रेनिंग चल रही है उसमें जब ये चीज आई तो दिमाग हमारा चकरा गया > hin> <bjj <हेई सासा का पढ़े वाला लरिका आयो > <नीतू का जानत है > bjj> < hin <जब उसने चालू किया तो हमने कहीं पढ़ा था ये > < दिमाग में था लेकिन होई कुछ > <लेकिन वो बेसिक से ही आधारित है तो उसका ज्ञान आवश्यक है> <ट्रेनिंग एक दिन के लिये थोड़ी न होती है> < आल लाइफ के लिये होती है> hin >

<bjj < etta kam kare dzew >. <etti tjinhari apn dehe dzew >...<sat sal hoige parhawet > bjj> <hin < abhi dzo trenin tjal rahi he usme dzab je tjidz ai to dimag hamara tjakra gaja. > hin> <bjj < hee sasa ka pade wala larika aje> <nitu ka dzanet he. > bjj> <hin < dzab usne tjalu kija to hamne kahni parha tha je. <dimag me tha lekin ho kuthe >... <lekin wo besik se hi adharit he. to uska gjan awasyak he> trenin ek din ke liye thodi na hoti he. > <al laiph ke lije hoti he. > hin>

< please do this much favour to me>. < It will remind me of you >...< I have been teaching for seven years> < When I came across this thing in the current training, my mind went blank> < He (the trainer) comes from Sasa and knows the Neetu > < As soon as he started, I knew it as I read about it somewhere>. < I was in my mind >. But then, may be something>... < But, it starts from basic concepts. So, it's necessary to know it >. < Training doesn't help only for a day (Meaning) > < It's for whole life>

This sample provides us with a fair idea that Kanauji speakers frequently code-mix (underlined words and phrases) and code-switch from Kanauji to Hindi and vice versa. Words 'training' and 'basic' and phrase 'all life' from English language is also used in different statements.

Embedded Foreign Languages: When the target language is under-resourced or low resourced it is likely to be influenced by one or more embedded foreign languages. The development of speech applications requires a corpus which is adequately developed (having good amount of transcribed speech data) and which in turn need good amount of text data for the construction of prompts, language models, pronunciation dictionaries, etc. Gathering all these components of an under-resourced language is quite a difficult task.

## 14. Conclusion

There is a dire need for multilingual and multipurpose speech corpora for Indian Languages. Despite many efforts, the number and varieties of such corpora are not yet enough to facilitate the academic research beyond their developers. It is fact that the majors' efforts were made for the larger languages of India. However, for corpus-based studies on smaller or endangered languages, even theoretical in nature, it is still the responsibility of the concerned researcher to develop the corpus of that language. This is a challenging, but important task ahead of any search studies.

These corpora would serve two perspectives: to provide language documentation and to make the benefit of information technology available to all sections of the society. Speech corpora can wonderfully work in language revitalization process by assisting in the process of product development for endangered languages. Considering the fact that India is one of the most linguistically richest countries in the world but hundreds of its languages are on the verge of being lost, a need for preservation and maintenance of linguistic heritage becomes all the more important and thereby role of language corpora also become very significant.

## References

- ALLWOOD, J. (2006) "Language survival kits", in A. Saxena & L. Borin (eds.), *Lesser-Known languages of South Asia: Status and policies, case studies and applications of information technology. Trends in linguistics studies and monographs*, vol. 175, Berlin: Mouton De Gruyter, 279-292.
- BIRD, S. & G. SIMONS (2003) "Seven dimensions of portability for language documentation and description", *Language*, 557-582.
- BORIN, L. (2006) "Supporting lesser-known languages: The promise of language technology", in A. Saxena & L. Borin (eds.), *Lesser-Known languages of South Asia: Status and policies, case studies and applications of information technology. Trends in linguistics studies and monographs*, vol. 175, Berlin: Mouton De Gruyter, 317-337.
- BROWN, G. & G. YULE (1983) *Discourse analysis*, Cambridge: Cambridge University Press.



- COX, C. (2009) "Corpus linguistics and language documentation: Challenges for collaboration", in John Newman, R. Harald Baayen & Sally Rice (eds.), *Corpus-based studies in language use, language learning, and language documentation*, Amsterdam: Rodopi, 239–264.  
<Retrieved on October 23, 2013 from <http://www.ualberta.ca/~aac12009/PDFs/Cox2009AAACL.pdf>>
- CRYSTAL, D. (2000) *Language death*, Chicago: Ernst Klett Sprachen.
- SOMAN, S. (2013) *India lost 220 languages in last 50 years*, (August 9, 2013), Times of India reports. <Source: <http://timesofindia.indiatimes.com/india/India-lost-220-languages-in-last-50-years-survey-finds/articleshow/21720601.cms>>.
- JANSEN, J. B. & S. B. SØRENSEN (2005) "Voices of the world: The extinction of language and linguistic diversity [motion pictures]", USA: Final Cut Productions. <Retrieved from <http://digital.films.com/play/2VTLH2> on October 17, 2013>
- JOHNSON, H. (2004) "Language documentation and archiving, or how to build a better corpus", *Language documentation and description*, 2, 140-153.
- KEAY, F. E. (1920) *A history of Hindi literature*, Mysore City: Wesleyan Press.
- KIRUTHIGA, S. & K. KRISHNAMOORTHY (2012) "Design issues in developing speech corpus for Indian languages—A survey", in *2012 International Conference on Computer Communication and Informatics*, 1-4.
- MCENERY, T., & N. OSTLER (2000) "A new agenda for corpus linguistics-working with all of the world's languages", *Literary and linguistic computing*, 15(4), 403-420.
- MURTOZA, S. M., F. ALAM, R. SULTANA, S. A. CHOWDHUR & M. KHAN (2011) "Phonetically balanced Bangla speech corpus", in *Proc. Conference on Human Language Technology for Development 2011*, Alexandria, Egypt, 87-93.
- PAUL, L. M., G. F. SIMONS & C. D. FENNIG (eds.) (2013) *Ethnologue: Languages of the world*, Seventeenth edition, Dallas, Texas: SIL International [Online version: <<http://www.ethnologue.com>> accessed: August 4th, 2015].
- RUSSELL, R. V. (1916) *The tribes and castes of the central provinces of India*, vol. 3, London: Macmillan and Company, limited.
- STĂNESCU, M., H. CUCU, A. BUZO & C. BURILEANU (2012) "ASR for low resourced languages: Building a phonetically balanced Romanian speech corpus", *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, Bucharest, Romania, IEEE.
- STEPHENSON, P. (2013) *Living Language*, Faculty Q&A With Linguist John McWhorter (February 20th, 2013): <<http://news.columbia.edu/content/living-language-faculty-qa-linguist-john-mcwhorter>> (accessed on August 20, 2015).

- TOGNINI-BONELLI, E. (2010) "Theoretical overview of the evolution of corpus linguistics", *The Routledge Handbook of Corpus Linguistics*. Abingdon: Routledge, 14-27.
- TROSTERUD, T. (2006) "Grammatically based language technology for minority languages", in A: Saxena & L. Borin (eds.), *Lesser-Known languages of South Asia: Status and policies, case studies and applications of information technology*, vol. 175, Berlin: Mouton De Gruyter, 293-315.
- UPADHYAYA, A.S. (1934) *The origin and growth of Hindi language and its literature*, Patna: Patna University.

*Received 28 July 2015.*

*Accepted 12 October 2015.*

## **TRANSLATING DIALECT EMBEDDED IN NAJDI PROVERBS INTO ENGLISH**

Sayed M. ISMAIL

Sattam bin Abdualziz University\*

sayedism@gmail.com

### **Abstract**

The study attempts to provide an insight into the problematic issues arising from translating dialect embedded in common sayings and proverbs. It proposes solutions to the problems of translating dialects. Owing to the fact that dialect has a cultural singularity and social privacy, it is only understood by the people who use it in their daily life. Thus, when an Arab translator attempts to translate dialect, he will be faced by a wide spectrum of problematic issues. For example, he/she cannot understand the meaning of the dialectal words in the same source language. In addition, he will not be knowledgeable about the cultural, social, and historical factors that change the meaning of these words. In addition, dialect is a form of oral speech which is delivered from a generation to generation through daily life contact. Accordingly, many words are subject to rapid changes and other words can be corrupted and distorted. These issues represent a great challenge for translators working in such a field. Thus, the study attempts to provide a translation approach which may constitute a good starting point for dealing with such problems.

### **Keywords**

dialect, polysystem theory, intralingual translation, cultural understanding

## **TRADUCCIÓN AL INGLÉS DE DIALECTALISMOS DE LOS PROVERBIOS EN NAJDI**

### **Resumen**

El estudio intenta penetrar en los temas problemáticos que surgen al traducir dialectalismos incluidos en refranes y proverbios habituales. Propone soluciones a los problemas de traducción de

---

\* Prince Sattam Bin Abdulaziz University, Alkharj Community College. Abu Dajan Street- Al-montzha District. P.O. BOX: 1866, Alkharj 11942, Riyadh, Kingdom of Saudi Arabia.

dialectos. Puesto que el dialecto tiene una singularidad cultural y privacidad social, es sólo entendido por los hablantes que lo usan en su vida cotidiana. Así, cuando un traductor árabe intenta traducir el dialecto, se enfrenta a un amplio espectro de cuestiones problemáticas. Por ejemplo, no puede entender el significado de las palabras dialectales en el idioma de origen y no está bien informado sobre los factores culturales, sociales e históricos que cambian los significados de estas palabras. Además, el dialecto es una forma de discurso oral que pasa de una generación a otra a través del contacto de la vida diaria. En consecuencia, muchas palabras están sujetas a cambios rápidos y otras palabras pueden ser corrompidas y distorsionadas. Estas cuestiones representan un gran desafío para los traductores que trabajan en este campo. Por lo tanto, el estudio intenta proporcionar un enfoque de traducción que puede constituir un buen punto de partida para abordar estos problemas.

**Palabras clave**

dialecto, teoría del polisistema, traducción intralingual, comprensión cultural

**1. Introduction**

Translating dialect is a controversial issue as dialect is distinct from both classical Arabic and Modern Standard Arabic. Dialect is known for its narrow locality. In other words, dialect is spoken by a small group of people or members of the same community who share in common certain linguistic, ethnic, and cultural features. Thus dialect is mainly associated with slangs, vernacular speech, daily conversations, and informal speech and so on. Traditional forms of literature and folklore attempt to keep dialects and vernacular speech surviving and vice versa. Thus, the Saudi culture draws unrivalled attention to the issue of dialect protection because it represents a major channel that not only protects the inherited tradition from loss or oblivion but also keeps it alive for the next generations. Accordingly, Saudi modern and pre-modern literature abounds in dialects and vernacular speech.

The varieties of dialects in Saudi are influenced by various factors such as its tribal affiliations, geographical or regional affiliations, socioeconomic conditions, the modern urbanization, the spread of education and the mass media in most parts of the Kingdom, and the waves of immigrations from neighboring Arab and the Asian countries for work.

## **2. The significance of research**

Saudi is divided into two main regions, Hejaz region and Najd region. Each main region includes subsidiaries regions with a variety of dialects, slangs, traditions and tribes. The locality and the linguistic specificity of each dialect may sometimes cause miscommunication and misunderstanding not only between Saudi people and non-Saudi Arab but also among the Saudi people themselves. Thus, the idea of rendering such dialect into English may cause a major problem of understanding for the Arab translators, whether they are Saudi or non-Saudi. In addition, the development of dialects is associated with various social, religious-geographical and cultural factors.

## **3. Research problems**

The research deals with various problems related to translating dialects. Firstly, dialect is different from classical Arabic and modern standard Arabic (MSA). These differences can be phonological, lexical, grammatical, etymological and cultural. Accordingly, when an Arab translator attempts to translate dialect into English, which is not in MSA or different from his special dialect, he will be faced by a problem of lack of understanding to variety of dialectal terms and expressions used in the source text. Thus, the translation problematic in such a case is twofold; the first is related to understanding within the same source language and the second is associated with rendering the source language into English. Secondly, another problem which faces the translator of dialect into English is that how he/she could preserve and keep such dialect notion of the source language unchanged in the target language. In other words, how can he/she render the dialect to another dialect, not formal language? Finally, because dialects represent vernacular speech and slangs, different from the modern Standard Arabic and Classical Arabic, they are vulnerable to various changes that range from semantic, lexical, and etymological and so on, which may bring about difficulties in translation.

#### 4. Methodology

The research will use both the analytical and comparative methodology. The analytical methodology will be used to analyze linguistically some examples taken from the forms of literature and proverbs of Najd region that demonstrate to what extent the dialect can affect the process of translation/interpretation into English. The linguistic analysis of Najd dialect being presented in these literary forms will be coupled with a cultural analysis. It will compare these examples to both Modern Standard Arabic and their equivalence in English culture.

#### 5. Review of literature

The studies written in English dealing with translating the Saudi dialects into English are a few in number. In most cases, these studies focused on the equivalence theories of translation. The dialectical expressions cannot be rendered simply through word for word or equivalence theories of translation because they are overloaded with long-standing historical tradition that fused with the present actualities.

In his unpublished PhD dissertation entitled, *A Critical and Comparative Study of Modern Najd Arabic Proverbs*, Muhammad Al-Suliman Al-Saudis (1950) did a marvelous effort in collecting a great deal of modern and classical Saudi proverbs widely used in Najd area and endeavored hard to translate them into English. Pursuing the translation of proverbs, he totally circumvented any reference to the dialect embedded in the body of these proverbs. Rather, he confined the translation process of the proverbs to the linguistic theories of translation. In addition, he failed to present the cultural connotations implicit in the original message intended by the author or the proverb narrator.

In his article, "The Golden Middle in Translating from Arabic, e.g. into English", Bahaa M. Mezid (2008) argues for presenting solutions for translating the texts written in classical Arabic language or in any of the colloquial varieties of the Arabic language. He proposed that the intranligual approach be used as a better method to translate the

texts embedded with dialects. However, he did not explain how the translator can use the intralingual approach for rendering classical words whose connotations have been changing overtime.

## 6. Classical Arabic, modern standard Arabic (MSA) and Najd dialect

C. A. Ferguson explains that Arabic speaking language is characterized by diglossia, which is defined as “two or more varieties of the same language are used by some speakers under different conditions” (1959: 325). He divides Arabic into the following varieties: Classical Arabic, Modern Standard Arabic and Vernacular speech. Modern Standard Arabic (MSA) is used in sermons and other formal contexts. Vernacular speech or slang is spoken in everyday conversations and other informal encounters. According to Thompson & Werfelli (2012), modern standard Arabic (MSA) is defined as follows:

This is the language used in the Arabic-speaking world for most written texts – it is the language of newspapers, of the media, and of the school system. In addition to this form of Arabic, also well-studied in the Islamic intellectual tradition is Classical Arabic, which is the form of Arabic found in the Qur’an, the Muslim holy book (Thompson & Werfelli 2012: 2).

MSA is different from dialect at different levels. These differences can be represented in vocabulary, syntax, morphology, and spelling. For example, dialects use their own lexical choices for articulating their concepts. The concept corresponding to *Oryd* ('I want') is spoken as *Awz* in Egyptian, *Abgy* in Gulf, *bdy* in Lebanon and Syria. Some examples of words and expressions that vary from one colloquial variety of Arabic to other are as follows: (1) 'very much': *barshabarsha* (Tunisian), *waajid* and *waayid* (Gulf, UAE, Kuwait), *ħeel*, *marrah* (Gulf, Saudi), *biz-zaaf* (Moroccan), *qawi* and *?awi*, *jiddan* and *giddan* (Egypt); (2) 'man, guy': *raajil* and *raagil*, *gadaa* (Egypt), *zalamih* (Levant, Syrian, Lebanese), *zoul* (Sudanese); (3) 'now': *dilwa?ti*, *dilwakiit*, *dilawakiiti*, and *dilwak* (Egypt), *al ħiin* and *da ħiin* (Saudi), *hassih* (Iraqi, Jordanian); (4) 'thus, as such':

*kida*,; *kdi* and *kidahawwiiti* (Egypt), *heek* (Syrian, Lebanese), *tshi* (UAE), *hakka* (Tunisian).

In addition to the lexical differences between MSA and dialect, there are also phonological differences. For example, these differences can be reflected through the glottal stop. The word /raʔs/ (head) as opposed to the vernacular /ra:s/ (Versteegh 1997: 99). In addition, the uvular /q/ is used in Standard Arabic, in Saudi dialects it is replaced by [g], [k], and [ʔ]. For example [ħagi:ga] is the /ħaqi:qa/ (truth). On the other hand, in some Egyptian and Syrian dialects /q/ is pronounced as glottal [ʔ], thus /ħaqi:qa/ is realized as [ħaʔi:ʔa].

Dialect is defined as “a language variety which is used in geographically limited part of a language area in which it is roofed by a structurally related standard” (Hinskens & Auer 2005: 1). The environment of the dialect is a part of larger environment that encompasses various dialects. Such comprehensive environment is called language. Classists sometimes defines it as a language. They also define it as distortion of speech (Anis 1965: 15). Each dialect has its distinct linguistic features and cultural singularity that makes it different from other dialects. “A dialect typically displays structural peculiarities in several language components” (Chambers & Trudgill 1998: 5). Each dialect has its own separate phonological features which are different from the other dialects. For example, Bani Assad tribe pronounced *Sakra* ‘drunk’ as *Skra’na*. In addition, a part of Tamim tribe, used the word *Madayoun*, indebted, instead of made. Differences are not only limited to phonological differences but also lexical. Some Arabic dictionaries explained that Hejazi people defines word *Hagrs* as ‘monkey’ and Tammi tribe defines it as ‘fox’ (Anis 1965: 16).

## **7. Translating Najd dialect in proverbs**

The process of translating the proverbs and the literary works embedded with Najd dialects would generate various problematic issues. First of all, the meaning of the dialectal words is largely influenced by the surrounding culture. The translator may find a further difficulty if he/she tries to understand or to get the meaning of the dialectal



words without linking them with their broader context. The dialect words cannot be separated from the things that produce them to the outside reality. Peter Behnstedt (2009) puts it clearly in the following lines:

It simply means that one has not only to study “the words”, their meanings and history, but also the “things”, the cultural artefacts and concepts, which are designated. The concept is meanwhile well-established, and to come right to the point: “a lamb within an Arab sheep –rears’ world is not simply a lamb”, it might be “weaned”, “fat”, “have two or four new incisors”, it might be “one year old or two” and consequently have different names. “A well” in Arabic dialects is not simply a well, it may be deep, shallow, narrow, broad, dry in summer, have much water, have little water, it may be walled or not, with bricks or something else, etc. Accordingly, different designations may be used (Behnstedt 2009: 63)

Translating a dialectal word into English is not a simple issue as in many cases “the lexicological issues reflect the interpretation of a given culture” (Cappuccio 2005: 47). Thus, dialects need a deep understanding of the cultural elements embedded in them that have changed their meaning across time and space. In other words, a dialectal word in itself receives its meaning from its cultural context. For example, according to Najd dialect the slang word *mokhanfas* means ‘easily provoked’ or ‘angry man’ (Algebali 1990: 130). However, according to the Egyptian dialect, *mokhanfas* means ‘long-curly haired man’. In Najd dialect, the Arabic word *akal* means ‘to spoil’, ‘to damage or to corrupt something’. This is explained in the following saying, *albard akel alrebei*, which can be rendered into English as follows: ‘cold damages the seeds of the wheat’. In the modern standard Arabic language, *akel* means ‘to eat something’ (Alqweai 1994: 80). In the Arabic dictionaries, the Arabic word *alrebei* means ‘blossoming’ and also means ‘the son of a young man’. Thus, the historical element and external cultural influences changed the meaning of the word *alrebei*, which means ‘the son of a young man’, into ‘the seeds of the wheat’. This may draw our attention toward an important conclusion that the words used in dialects acquired new connotations to be different from the formal and modern standard Arabic language. Dialect is deeply associated with its

cultural environment.

Another example which may explain how the dialectal word could receive a totally distinct meaning from its surrounding culture is the word *jard*. This word means 'cavalry'. People in Najd uses it in the following context: *ghaza fellan bee sard and jard* (Hamed 1990: 33). It can be rendered into English as follows: 'Mr. X, invades by shields and cavalry'. However, in formal Arabic dictionaries, the word *jard* has almost different meanings and distinct connotations such as a 'inventory', 'waste land', and so on. Accordingly, the dialectal words or vernacular speech share the same orthographic letters with the formal Arabic language but the differences can be shown clearly in the pronunciation, meaning, and cultural connotations.

Being informed of the lexical, semantic and phonological differences between the formal Arabic language and the dialect of Najd, one can admit that the translation theories based on the word for word would fail to provide an accurate and precise translation. Thus, the question being addressed in this context as follow: how could we translate such terms? Which translation methodology is recommended to render Najd dialectal terms embedded in proverbs and literary genres such as poetry, drama and so on? Taking into consideration the idea that dialect is largely used in oral speech rather than the written texts.

## **8. Translation theories and translating dialect**

Obviously, the issue of dialect translation is such a complicated and intricate due to the aforementioned reasons. Indeed, the previous studies recommended to use the intralingual approach for better understanding in the same source language (Mezid 2008, Dickins 1990). Roman Jakobson in his article "On Linguistic Aspects of Translation" (1959) explains that message being communicated through translation is always incomplete, as the linguistic equivalence never totally communicates the full message of the original text. However, there is always something missing and incomplete in the translated text. He uses the word 'cheese' in English as an example. This word is not identical to the Russian word or code-unit "syr" since the latter does not include the

concept of cottage cheese (Venuti 2000: 114). When the linguistic terminology fails to provide an accurate and precise translation, according to Jakobson, the translator should resort to using loanwords or loan-translations, neologisms or semantic shifts (Venuti 2000: 115). Thus, Jakobson's substitution policy on the level of the word is not sufficient for providing a good translation because the translator is still turning around the vicious circle of the word substitution policy. However, the major influence does not rest upon the linguistic rendering of the dialectal word but in the cultural influence that changes the meaning of the word across time to be a newly born word that is different from both Modern Standard Arabic and classical Arabic.

In his book, *Toward a Science of Translating* (1964), Eugene Nida argues for dynamic equivalence that can fill in the cultural gap between two different languages. The dynamic equivalence follows the cultural standard of the target text; it is away to adapt the source text to the culture of the target text. The translation process, however, is limited to rendering word for word. According to Munday (2001: 38), Nida focused the translation process on the context of the target text.

In *Comparative Stylistics of French and English: A Methodology for Translation* (1995), Jean-Paul Vinay and Jean Darbelent provide two translation strategies for handling the problematic issues arising from translating culturally embedded texts. These two translation strategies are direct and oblique strategies. They confined the problematic issues arising from translating culturally embedded texts to "structuralism parallelism" and the "metalinguistic parallelism". Thus, the direct and oblique strategies deal with the syntactic, lexical and stylistic issues. However, there is a remarkable parallelism ignored by such a theory: cultural parallelism. The linguistic theories of translation such as Catford's (1965) formal correspondence and textual equivalence, Katharina Reiss's (1971) theory of text type model, Vermeer's (1989) Skopos theory partly fails to deal with the problematic issues arising from translating texts embedded with dialectal terms and vernacular speech.

Accordingly, the convenient approach for rendering the dialectal terms embedded in Najd proverbs and literature is such an approach that considers translation a bridging vehicle between cultures. In other words, translated works should imitate the original literary work; it has to mimic the forms of literary text in the target language. Thus,

translated text has to be culturally and linguistically manipulated in order to be relevant to the target cultural system and its aesthetic and poetic forms. In this way, translation has to be employed as an act of representation. Itamar Even-Zohar clearly puts it as follows:

The idea that semiotic phenomena, i.e., sign-governed human patterns of communication (such as culture, language, literature, society), could more adequately be understood and studied if regarded as systems rather than conglomerates of disparate elements has become one of the leading ideas of our time in most sciences of man. Thus, the positivistic collection of data, taken *bona fide* on empiricist grounds and analyzed on the basis of their material *substance*, has been replaced by a functional approach based on the analysis of *relations*. Viewing them as systems made it possible to hypothesize how the various semiotic aggregates operate (Even-Zohar 1990: 10).

According to Even-Zohar, the translation process of a given text has to fully focus on the whole system affecting the text and producing it to the reality. The cultural, linguist, historical and social elements should be taken into consideration while translating a literary text into English. Furthermore, “being placed in this way in a larger sociocultural context, ‘literature’ comes to be viewed not just a collection of texts, but more broadly as a set of factors governing the production, promotion and reception of these texts” (Baker & Malmkjaer 1998/2001: 176-177). Thus, adopting Even-Zohar theory of translation can be helpful in handling the problematic issues arising from translating the vernacular speech and slang language”. “Translation is no longer a phenomenon whose nature and borders are given once and for all, but an activity dependent on the relation within a certain cultural system” (Even-Zohar 1990: 51).

Accordingly, the researcher proposes the following steps for translating dialect embedded in Najdi proverbs. Firstly, applying the intralingual concept of translation in order to understand, explain and interpret the vernacular speech and dialect words whose meaning and cultural connotations are ambiguous and clumsy in Arabic. Secondly, the translator has to study the role of the cultural impact on the texts being translated. This cultural influence should trace the historical and social changes and

their association with the development and change of meaning of dialectal terms. Thirdly, it questions whether dialectal words are fixed in meaning or changed over time. Fourthly, it tries to adapt and appropriate the translated text to fit the poetic and literary forms of the target culture. Finally, the process of translation is not only confined to search for linguistic equivalence; rather, the translator has to render his/her text against the grains of the entire tradition and culture. In other words, translation process should start from the smaller units to the larger units and the vice versa. That is to say, what is not clearly stated by focusing on the smaller units can be revealed by examining the larger units.

## **9. Samples and analysis**

First, the researcher not only attempts to translate the Najdi proverb but also explains the process of their translation, focusing on the problematic issues arising from encountering this translation. Translating the proverb, *la teferha be soraat omak ala al-tanour*, the translator does not face any problem on a linguistic level. The language is simple and the words can be easily rendered into English. The question posed is this: can linguistic equivalence provide us with the intended meaning of the proverb? A purely linguistic approach translates this proverb as 'Do not feel happy about your mother's quick performance at the oven'. This translation appears meaningless and does not communicate any kind of message to a Western reader. It has nothing to do with the core value of the proverb, which is intended to communicate a clear and intelligible message to the addressed party.

A cultural interpretation of the Arabic proverb is needed. In other words, the proverb has to be understood and interpreted within its broader context. The historical and socio-economic realities associated with the text have to be investigated prior to translating it. In this proverb, the speaker is warning the listener about the risks of being overly optimistic or from holding unrealistic expectations. The historical context, in which such proverb was produced, is a prerequisite for providing true and precise translation. In this context, the addressee may feel pleased at the expectation that

his/her mother is preparing a delicious meal for a few moments time. However, summoning up the historical context in which such a proverb was said can help us provide a true translation and this can communicate the core message of the original text without misrepresenting it. The historical context in which this proverb was said relates to scarcity of food and the lack of essential ingredients for cooking a good and delicious meal. As such it speaks of the mother cooking a meal very quickly because of a lack of resources" (Abdelmehssn 1997: 6).

Simple linguistic equivalence is irrelevant, unnecessary and does not communicate the invisible and concealed message of the proverb. Therefore, it is necessary to translate the invisible message of the proverb or the authorial or the speaker's intention. The speaker's intention or the authorial intention may be common among the members of Najd community, as it is a wise old saying with a clear association and a recognized situation. The English receiver lacks these cultural elements that would help reveal the ambiguity in meaning and clarify it. The task of the translator is neither simple nor easy because he/she has to transfer the paralinguistic elements of the source text to the target text. Subsequently, the unseen elements, invisible cultural notion, should be translated into seen elements and visible cultural notion in order to deliver a precise and clear meaning to the receiver. The wisdom of this saying is that one should not raise one's expectations to high and unachievable levels as this will lead to disappointment. The translator has to presuppose the historical background of the text, which can be achieved through understanding the contextual situation out of which the proverb was born and used.

Furthermore, the translation process needs to be culturally and linguistically adapted to the target culture. The translator has to search for an English equivalent which conveys the same concept as the Arabic. One possible translation of the abovementioned proverb could be 'blessed is he who expects nothing, for he shall never be disappointed' (Oxford Dictionary of Proverbs). Translating the following proverb, *Emdhani waa Khod Abyati*, the translator has to avoid using the linguistic equivalent or literal translation, 'praise me and take my cloak'. This proverb is similar to the Egyptian saying, *laqeeni wa la taghdeeni*. In most idiomatic translation, this proverb is translated as follows: 'I'd rather be well received than well fed' (Better warm welcome than being

invited to lunch). This translation is far beyond the authorial intention or the true meaning of the saying. In the Arab culture, this saying urges the addressee to welcome his guest warmly. The proposed method for translating this proverb is to search for the English proverb which communicates a similar message. Thus, it can be translated as follows: 'First impressions are the most lasting' (*Oxford Dictionary for Proverbs*).

'He is as tall as palm tree but his mind is as foolish as the kid' is literal translation of the following proverb, *altwool tool nakhlawaa akalaki sakhla*. This translation is void of poetic and atheistic language and does not agree well with the cultural values of the receiver, as 'the kid' is not the same symbol of foolishness in English culture. It does not convey a clear or understandable message for the reader, as it does not render the true meaning of the proverb. A proposed equivalent in English could be 'a little body does often harbor a great soul' or 'little things please little minds' (*Oxford Dictionary of Proverbs*).

## 10. Conclusions

The research reached a conclusion that the Dialect of Najd is translatable. However, the process of translating dialect requires an understanding of the nature of dialect as an oral speech, which is changeable, and renewable. A major problem facing the translators, who are concerned with rendering dialect, is their inability to understand the dictions and the words used in dialects. In addition, translating dialect cannot be done through resorting to literal translation or translating word for word. However, the process of translating Najd dialect necessitates understanding the world out of which the dialect has been produced to real-life. Dialect derives its meaning from its cultural realities and from its historical context. Since it is an oral form of speech is always changeable and derives its meaning from its larger context. That is to say, the translator never understands the language of dialect but the historical and cultural context which help produce them.

This study presents the following recommendations:

a) The dialectal words should be linguistically analyzed from time to time in order to make sure whether or not these words preserve their original meanings or not. In case of any change to the meaning, socio-linguistics studies should be done in order to investigate the reasons of such changes.

b) The translator who is concerned with translating dialect should culturally belong to the same region of the dialect in order to be fully aware of the ambiguous and ambivalent meanings of the dialect.

c) The cultural concept of translation has to be employed when translating dialects into English.

## References

- ABDELMEHSSN, K. (1997) *A Study of Najd Dialect*, Riyadh: Alnebras Publishing House.
- ALGEBALI, A. (1990) *Letters from the Tradition*, Riyadh: AlmahrajanAlwatni le torath and thaqfa.
- ALQWEAI, M. (1994) *A Reading of the Tradition*, Riyadh: King Fahd Publishing house.
- AL-SAUDIS , A. (1950) "A Critical and Comparative Study of Modern Najd Arabic Proverbs", PhD thesis, University of Leeds.
- ANIS, I. (1952) *Arabic Dialects*, Cairo: Anglo Publishing House.
- ARMSTRONG, N. (2005) *Translation, Linguistics, Culture: A French English Handbook*, Clevedon: Multilingual Matters.
- BAKER, M. & K. MALMKJAER (1998/2001) *Encyclopedia of Translation Studies*, London: Routledge.
- BEHNSTEDT, P. (2009) "Word and Things", *Arabic Dialectology*, Netherlands: Brill Publishing House, 60-72.
- Cappuccio, D. (2005) *Translating Neapolitan Dialect in Theatre: Problems of Cultural Transfer*. California: California Scholarship Press.
- CATFORD, J. C. (1965) *A Linguistic Theory of Translation*, Oxford: Oxford University Press.
- CHAMBERS, J. K. & P. TRUDGILL (1998) *Dialectology* London: Cambridge University Press.
- DICKINS, J. (1990) *Thinking Arabic Translation*, London: Routledge.
- EVEN-ZOHAR, I. (1990) "Polysystems Studies", *Poetics Today*, 11, 1, 9-26.
- FERGUSON, C. A. (1959) "Diglossia", *Word*, 15, 35-401.
- HAMED, A. (1990)<sup>2</sup> *The Arabic Dialects: Establishment and Development*, Cairo: Al-Azhar Publishing.



- HINSKENS, F., P. AUER *et al.* (2005) *Dialect Change Convergence and Divergence in European Languages*, London: Cambridge University Press.
- JAKOBSON, R. (1959) "On Linguistic Aspects of Translation", in Reuben A. Brower (ed.), *On Translation*. Harvard: Harvard University Press, pp. 232-239.
- MEZID, B. (2008) "The Golden Middle in Translating from Arabic, e.g. into English", *1<sup>st</sup> In't Conference on Translation/Interpretation and the Impact of Globalization, Nov. 18-20, 2008-UAE*, United Arab Emirates University.
- MUNDAY, J. (2001) *Introducing Translation Studies: Theories and Application*, London: Routledge.
- NIDA, E. (1964) *Toward a Science of Translating*, Netherlands: E. J. Brill.
- REISS, K. (2000) "Type, kind and individuality of text: decision making in translation", in L. Venuti, *The translation studies reader*, London: Routledge.
- THOMPSON, Ellen & Sawsan WERFELLI (2012) "The Position of the Subject in Spoken Saudi Arabic: A Processing Perspective", *Tucson*, 19, 20-30.
- VENUTI, L. (ed.) (2000) *The Translation Studies Reader*, London: Routledge.
- VERMEER, H. (1989) *Skopos and translation Commission*, Heidelberg: Universitat.
- VERSTEEGH, K. (1997) *The Arabic Language Study*, USA: Columbia University Press.
- VINAY, P. & J. DARBELENT (1995) *Comparative Stylistics of French and English: A Methodology for Translation*, Netherlands: Benjamin Publishing House.

Online sources:

*Oxford Dictionary of Proverbs* < <http://www.oxforddictionaries.com> > (Retrieved August, 13, 2014).

Received 27 August 2015.

Accepted 19 December 2015.

## PASSIVE CONSTRUCTIONS IN ILAMI KURDISH. A ROLE AND REFERENCE GRAMMAR ACCOUNT

Amir KARIMIPOUR & Vali REZAI

University of Isfahan\*

am.karimipour@fgn.ui.ac.ir / vali.rezai@fgn.ui.ac.ir

### Abstract

As far as obfuscation of the truth is taken into account, passivization seems to be one of the most prominent processes by which the agent of an action may become evasive (by being omitted or demoted to non-obligatory prepositional constituent) in certain contexts. Cross-linguistically, these constructions may not necessarily be formed in an identical fashion, as in some languages such items may be constructed syntactically, morphologically and/ or lexically. In this paper, through employing Role and Reference Grammar framework, we attempt to represent the nature of passive constructions in Ilami, a southern variety of Kurdish language. In pursuit of this goal, we enjoy Ilami data to represent how passivization could be explained in the RRG [Role and Reference Grammar] framework. Analyzing our Kurdish data, we concluded that Ilami gains advantage of the main types of passivization strategies. The use of *strict morphological*, *periphrastic* and *impersonal passives* was well-attested, and additionally, it can be said that the two stages of passivization suggested in RRG are done in Ilami passives.

### Keywords

Role and Reference Grammar, Ilami Kurdish, Passive Construction, PSA [Privileged Syntactic Argument]

---

\* University of Isfahan, Hezar Jarib Street, Isfahan, Isfahan Province, Iran.

## CONSTRUCCIONES PASIVAS EN KURDO ILAMI.

### UNA CONSIDERACIÓN DESDE LA GRAMÁTICA DEL PAPEL Y LA REFERENCIA

#### Resumen

En lo que respecta a la obstrucción de la verdad, la pasivización parece ser uno de los procesos más prominentes por los cuales el agente de una acción puede llegar a ser evasivo (omitiéndose o reduciéndose a un constituyente preposicional no obligatorio) en ciertos contextos. Lingüísticamente, estas construcciones no necesariamente pueden ser formadas de manera idéntica, ya que en algunas lenguas tales aspectos pueden construirse sintácticamente, morfológicamente y/o léxicamente. En este trabajo, a través de la utilización del rol y la gramática de referencia, tratamos de representar la naturaleza de las construcciones pasivas en Ilami, una variedad meridional de la lengua kurda. Para lograr este objetivo, usamos los datos Ilami para representar cómo la pasivización podría explicarse en el marco RRG [Gramática del Papel y la Referencia]. Al analizar los datos kurdos, concluimos que el Ilami obtiene ventaja de los principales tipos de estrategias de pasivización. El uso de pasivas morfológicas, perifrásticas e impersonales estrictas ha sido bien atestiguado, y, además, puede decirse que las dos etapas de pasivización sugeridas en la Gramática del Papel y la Referencia se aplican en las pasivas en Ilami.

#### Palabras clave

Gramática del Papel y la Referencia, Kurdo Ilami, Construcción pasiva, PSA [Privileged Syntactic Argument]

#### 1. Introduction

Passive is a term used in the grammatical analysis of voice, referring to a sentence, clause or verb form where the grammatical subject is typically the recipient or 'goal' of the action denoted by the verb, e.g.

(1) *The letter was written by a doctor.*

It is contrasted with active, and sometimes with other forms, e.g. 'middle' (as in Greek). A full linguistic statement of the constraints affecting these relationships is a complex matter. In English, for example, there are active sentences that do not have passive counterparts, for example:

(2) *The boy fell, They have a car*

And passive sentences which have an unclear active counterpart:

(3) *The house was sold.*

In addition, there is the problem that the central type of passive construction (using the verb *to be*, e.g. *She was pushed*) is closely related to other types of construction (cf. *She got pushed, She was interested*), and a boundary line is sometimes difficult to establish (Crystal 2008: 353).

Keenan & Dryer (2007) classifies passives as:

a) *Basic passives*: what makes them distinct from other passives is (i) no agent phrase (e.g. *by Mary*) is present, (ii) the main verb in its non-passive form is transitive, and (iii) the main verb expresses an action, taking agent subjects and patient object. Here is an example of this sort:

(4) *John was slapped*

b) *Strict morphological passives*: which are constructed through affixation (like infixing in Tagalog and internal vowel change in Arabic and Hebrew). Look at the following example from Sre (Mon-Khmer; Manley (1972)) representing a passive formed by suffixing:<sup>1</sup>

(5) a. Cal paʔ mpon

wind open door

*The wind opened the door.*

b) Mpon gə-paʔ mə cal

door PASS-open by wind

*The door was opened by the wind.*

---

<sup>1</sup> Abbreviations used: A = subject of a transitive verb, logical subject; DEF = definite; INDEF = indefinite; O = object of a transitive verb, logical object; OBJ = objective; PASS = passive.

c) *Periphrastic passives*: which could be constructed according to one of the following ways:

(i) *The auxiliary verb is a verb of being or becoming.* (Example from Persian.)

(6) a. Ali Ahmad-ra koj-t

Ali Ahmed-OBJ kill- 3SG/PAST

*Ali killed Ahmad.*

b. Ahmad kojft-e jod

Ahmed kill- OBJ become- PASS- PAST

*Ahmad was killed.*

(ii) *The passive auxiliary is a verb of reception* (e.g. *get, receive* or even *eat*). (Example from Welsh.)

(7) Cafodd Wyn ei rybuddio gan Ifor

get Wyn his warning by Ifor

*Wyn was warned by Ifor.*

(iii) *The passive auxiliary is a verb of motion* (e.g. *go, come*). (Example from Persian)

(8) a. Ali logat-ra be kar bord.

Ali word-do to work take

*Ali used the word.*

b. logat be kar raft.

word to work went

*The word was used.*

(iv) *The passive auxiliary is a verb of experiencing* (e.g. *suffer, touch, even 'experience pleasantly'*). (Example from Thai.)

(9) Mary th'uuk (John) k'oot

Mary touch (John) embrace

*Mary was embraced (by John).*

(Examples taken from Keenan & Dryer 2007)

In this study, we aim to represent the way passive constructions are formulated in Ilami Kurdish. By adopting Role and Reference Grammar approach, we attempt to answer the following questions:

1. How passive constructions are made in Ilami Kurdish, morphologically or syntactically?

2. Which passive-related stages (proposed in RRG) are done as far as Ilami examples are taken into consideration?

## **2. Review of literature**

Although we have been unable to find a prominent and relevant work in Ilami, there is lots of research done in other varieties, among others, we point to the following cases: Nolan (2005) examines impersonal passive constructions in Irish. He argues that impersonal passive construction in Irish has two forms. The impersonal passive form occurs with all verbs of Irish, across all tenses, whether intransitive or transitive. In the second form, the impersonal passive form is to be found productively with the substantive verb (one of the two verbs of 'to be' in Irish) across all tenses. It does not, however, under any circumstances occur with the copula verb. He then claims that the impersonal passive construction has an indefinite actor at the level of the semantics and that the impersonal passive verb expresses this as a third person indefinite pronoun in the syntax via a synthetic post-verbal suffix rendered on the matrix verb.

Adopting RRG framework, Rezai (2010) assesses passivization in Persian language. He claims that there are two strategies to make passives in that language, one is constructed via a syntactic operation: Objective form of the verb + *jodan* called "basic passive" and the other through the 3<sup>rd</sup> person plural inflexion of the verb with a covert

subject. Based on RRG criteria for passives, he shows that “transitivity alternation” should be differentiated from regular passivization, as such constructions are not syntactically made; however, they are semantically similar to passives.

Windfuhr (2009) briefly points to the construction of passive in Kurdish. Based on his analysis, passive stems are derived from present stems by *-r-e/r-a* « *\*-r-ad*): *kui-re-lkuz-ra-* 'be killed'. They are inflected with intransitive subject markers: *a-kui-r-e-m* 'I am, will be killed'; *kui-r-iJ-.w-im* 'I have been killed'; *agar bi-kui-r-e-m* 'if I am killed'; *agar bi-kuz-r-ii- m-ii-y-a* 'if I had been killed'. In conclusion, the author suggests no other way to construct passive in Kurdish dialects.

Jügel (2009) believes that in a prototypically ergative language, there is no need for a passive because O is the primary actant anyway, and A the secondary. The passive is, so to speak, inherent in the active construction in a prototypically ergative language. Hence an active of an ergative language can be interpreted as an active or as a passive of an accusative language depending on the context.

Leisiö (2006) assesses the passive constructions of Nganasan which is one of the varieties of Uralic language family. She claims that the characteristics of the English passive assigned by Foley and Van Valin (1984: 110-111) hold true for the Nganasan passive, as well. Passivization in this language permits the undergoer to occur as a syntactic pivot. The demoting of the actor follows as a result of the main operation of promotion of the undergoer. Thus, the actor may be either non- specified, or specified and omitted, or explicit-expressed by the lative in morphologically marked passive constructions and by the genitive *or/* and a possessive suffix in unoriented-participle constructions. While the unoriented participles and participles of passive verbs retain the aspect characteristics of their base verbs, the passive participles in Nganasan have a resultative meaning. Consequently, the pivot of the passive participle is typically a true patient – it is affected or effected in the result of the action and has more nouny and fewer verbal characteristics than other participles.

Esteban (2012) provides conclusive evidence to decide on the existence or absence of passive in two Native American languages, namely Lakhota and Cheyenne, which exhibit a similar behavior in this respect. In this paper, it is shown that there are two different types of languages with respect to this parameter and that this distinction

has a bearing on the existence of two different types of passive, whose combination will be exhibited by the English-style passive.

### **3. Role and reference grammar**

Introduced by Foley & Van Valin in 1984, Role and Reference Grammar was inspired by both typological and theoretical concerns. The motivating questions for RRG were:

1- What would a linguistic theory look like if it were based on the analysis of languages with diverse structures, such as Lakhota, Tagalog, Dyirbal and Barai (Papua New Guinea), rather than on the analysis of English?

2- And 'how can the interaction of syntax, semantics and pragmatics in different grammatical systems best be captured and explained?'

These two questions contain both theoretical and descriptive content. On the one hand, they both emphasize the importance of taking account of typologically diverse languages in the formulation of a linguistic theory, and on the other, they indicate that the resulting theory will be one in which semantics and pragmatics play significant roles. (Foley & Van Valin 1977, 1984, Van Valin 1977, 1981).

RRG maintains that a theory of clause structure should capture all of the universal features of clauses without imposing features on languages in which there is no evidence for them. This assumption rules out, for example, VP as a universal feature of clauses. RRG rejects both grammatical-relations-based representations, such as those in RelG and LFG, and X-bar-type constituent-structure representations, because, it is argued (see Van Valin & LaPolla 1997), neither type is universally valid. Despite the great diversity of human languages, there are universal features of clause structure: all languages distinguish structurally between predicating and non-predicating elements, on the one hand, and, among the non-predicating elements, between those that are semantically arguments of the predicating element and those that are not. This may be represented schematically as in Figure 1.



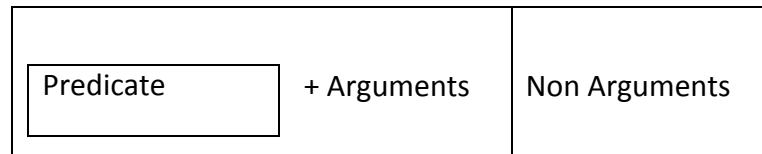
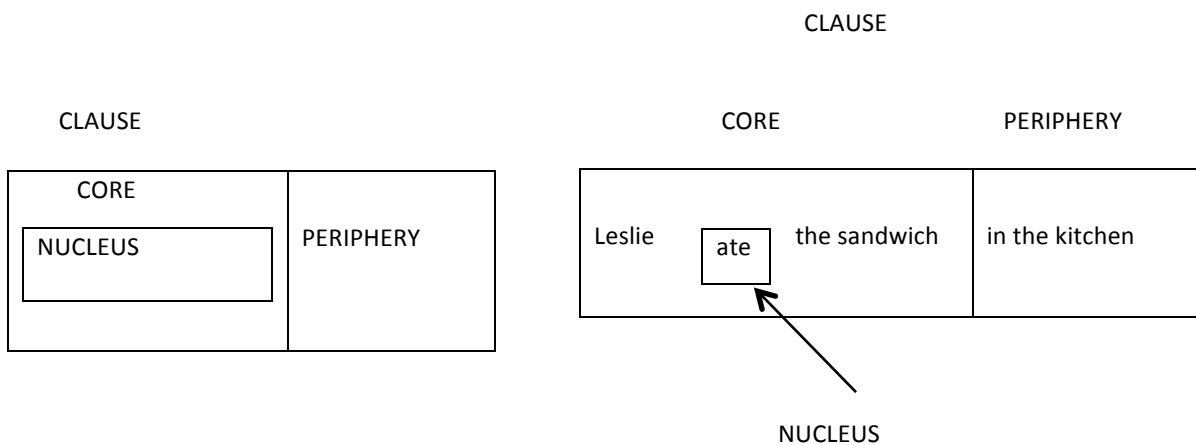


Figure 1. Universal oppositions underlying clause structure

The syntactic unit containing the predicate is termed the *nucleus*, the unit containing the nucleus plus the arguments of the predicate in the nucleus is called the *core*, and the unit encompassing the non-arguments (adjuncts) is labeled the *periphery*. Clauses are thus conceived of as having a *layered* structure, each layer being motivated semantically.



(Van Valin 2004: 205-206)

### 3.1 Organization and representations

There is a direct mapping between the semantic and syntactic representations of a sentence, unmediated by any kind of abstract syntactic representations; this excludes not only derivational representations as in e.g. the Minimalist Program, but also the use of abstract structures as in LFG. There is only a single syntactic representation for a sentence, and it corresponds to the actual form of the sentence. RRG does not allow any phonologically null elements in the syntax; if there's nothing there, there's nothing there. RRG posits a very concrete syntactic representation, and this constrains the theory significantly; this rules out derivations and therewith movement rules, however

they are formulated, and also abstract representations like relational networks in Relational Grammar or f-structures in LFG. Many of the descriptive and theoretical devices in theories with abstract syntactic representations are not available in RRG.

The organization of RRG is given in Figure 2:

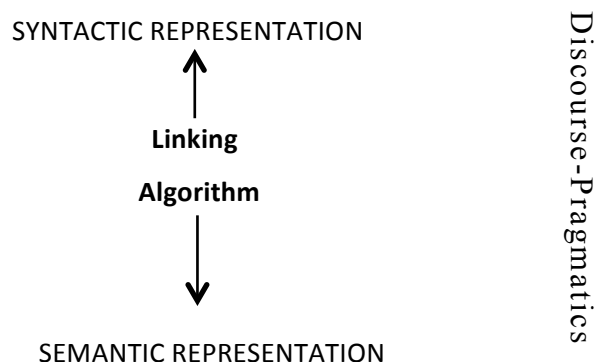


Figure 2. The organization of RRG

The syntactic representation is linked via the linking algorithm to the semantic representation. It consists of a lexical decomposition representation of the meaning of the predicator, along with its arguments. The last component in Figure 2 is labeled ‘discourse-pragmatics’, and it is parallel to the linking algorithm. What this depicts is the fact that discourse-pragmatics, primarily as realized in information structure, plays a role in the linking between syntax and semantics. Crucially, however, exactly what role it plays can vary across languages, and this variation is the source of important cross-linguistic differences among languages (Van Valin 1992: 3-4).

### 3.2 Semantic Roles

The RRG theory of semantic roles is rather different from that of other theories. There are only two macroroles, actor and undergoer, corresponding to the two primary arguments in a prototypical transitive relation. They are called ‘macroroles’ because each subsumes a number of specific thematic relations. The first are specific thematic relations, the traditional (since Fillmore 1968 and Gruber 1965) notions of agent, theme, patient, experiencer, etc. The second are generalized semantic roles called ‘semantic

macroroles'; they were introduced in Van Valin (1977) and have no exact analog in other theories, although Jackendoff's 'action tier' and Dowty's proto-roles bear some resemblance.

The role of the subject of an active voice transitive verb and the object of *by* in a passive construction is actor, and the role of the direct object of an active voice transitive verb and the subject of a passive verb is undergoer. Actor and undergoer are thus generalizations across the thematic relations. The single argument of an intransitive verb is either an actor, as with verbs like *run*, or an undergoer, as with verbs like *die*. The relationship between the macroroles and the argument positions in LSs is captured in the actor-undergoer hierarchy in Figure 3.

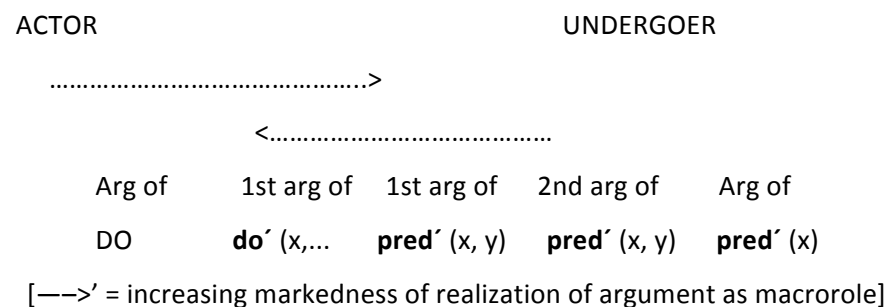


Figure 3. Actor-undergoer hierarchy

(Van Valin 2005: 13- 16)

### 3.3 Privileged Syntactic Argument

RRG has a very different view of grammatical relations from the other theories, because the theory does not attribute cross-linguistic validity to the traditional grammatical relations of subject, direct object and indirect object, and therefore does not employ them as theoretical or analytical constructs. Rather, it adopts a construction specific conception of grammatical relations and postulates only a single one, which is called the 'privileged syntactic argument'. 'Construction specific' means that a privileged argument may be identified for each construction; (Van Valin 2004: 212). So, PSA could be defined as a restricted neutralization of semantic roles and pragmatic functions for syntactic purposes. The other arguments in a clause are characterized as direct or

oblique core arguments; PSAs may be characterized functionally as controllers or pivots.  
Have a look at the following illustrations:

a. The tall man<sub>i</sub> hit William<sub>j</sub> and then.....i/ \*j ran away.

CONTROLLER                      PIVOT

b. William<sub>j</sub> was hit by the tall man<sub>i</sub> and then.....\*i/j ran away.

CONTROLLER                      PIVOT

a. Billi persuaded the tall man<sub>i</sub> [.....-i/j to visit Sam].

CONTROLLER PIVOT

b. The tall man<sub>j</sub> was persuaded by Billi {..... -i/j to visit Leslie}.

CONTROLLER                      PIVOT

Pivots are canonically the missing argument in a construction, as in the above-mentioned examples, while controllers prototypically supply the interpretation for a pivot (Van Valin 2005: 23).

### 3.4 *Passivization in Role and reference Grammar*

There are usually (but not always) two facets of a passive construction, the occurrence of a marked privileged syntactic argument choice, and the omission of the actor or its appearance as an oblique element in the periphery. Van Valin & LaPolla (1997) present the universal configuration of basic voice oppositions as follows:

a) PSA modulation voice: permits an argument other than the default argument to function as the privileged syntactic argument.

b) Argument modulation voice: gives non-canonical realization to a macrorole argument.

An important motivation for factoring voice constructions into these two parts is that they occur independently of each other in some languages.

(Van Valin & LaPolla 1997: 116)

#### 4. Methodology

Examples are taken from Ilami native speakers who have lived in the region for a long span of time. It might be worth noting that the participants were all above 40 years old with and without formal literacy. In order to analyze our data, we gained advantage of their linguistic intuition throughout the research wherever needed. Needless to say, the final decisions were made based on the authors' analysis.

#### 5. Ilami Kurdish

Kurdish as a new western Iranian language has speakers dispersed within broad regions of Iran, from west (Kurdistan, Kermanshah and Ilam) to the east (Khurasan), This language has three main dialect groups: The northern Kurdish dialects are usually given the term *Kurmanji* spoken in northwestern Iraq (Gunter 2004, xxv-xxvi). The central Kurdish dialects embrace *Mukri*, which is spoken in Iran, to the south of Lake Urmiya, and *Sorani*, to the west of Mukri, in the province of Erbil, in Iraq. The southern Kurdish dialect group includes *Kermanshahi*, *Ardalani*, *Laki* (and also *Ilami*) (Mackenzie 1963; Oranskij 1979: 35-36; Asatrian 2009: 12).

Each of these dialects, moreover, includes infinity of variants so that it is possible to say that every tribe and every valley has its own dialect. This is a phenomenon common to all mountain peoples. In any case it is nothing to be surprised at since, when the national dialects are used, the Arabs of the Maghreb make themselves understood with difficulty by the Iraqis, and the Egyptians do not understand the Lebanese very well. The Kurdish vocabulary is basically Iranian, but it has been influenced by Arabic as has Persian and Turkish, particularly in the department of religion since these peoples are all Muslims for whom the language of the Koran is essential (Bois 1966: 111-112).

Ilami, a less well-documented dialect of the Kurdish language, is widely spoken in Ilam by about 200000 speakers. It shares some features with Kermanshahi in most of its linguistic (e.g. syntactic, morphological and lexical) modules, but shows some

idiosyncratic characteristics too.

Ilami Kurdish, like many Iranian varieties, is marked as a no- gender/ no- case system in nouns and pronouns. This can be considered as a sharp distinction in comparison with the owning varieties of Kurmanji which have preserved gender and/ or different cases (Bynon 1979). In contrast, Ilami has pronominal affixes used to construct case relations which are not usually observed in northern dialects of Kurdish.

## 6. Data Analysis

In this section, through presenting Ilami data, passive constructions will be exemplified. For simplicity's sake, in addition to giving non-literal translations, English grammatical glosses are also given for each example.

### 6.1 Privileged Syntactic Argument in Ilami Kurdish

In an intransitive clause (a clause containing an intransitive verb), the PSA could be either actor or undergoer regardless of the tense of sentence:

(10) ej gæn qwtən- e.  
s/he bad cough-3SG PRES  
*S/he coughs so much.*

(11) ej gæn dəqwt-an  
s/he bad cough-3SG PAST  
*S/he coughed so much.*

(12) kæf-əm æ xwar  
fall- 1SG PRES to down  
*I fall down.*

- (13) kæft-əm æ xwar  
 fall- 1SG PAST to down  
*I fell down.*

Examples (10) and (11) depict contexts in which the syntactic agreement (person and number) is held between the subjects (as actors) and verbs of the sentences; however, in the next two examples, (12) and (13), the agreement is held between the subjects (as undergoers) and the verbs. Based on this fact, it can be said that in Ilami Kurdish, the syntactic agreement is a subject- verb agreement type and additionally the default PSA is the actor of the sentence.

### 6.2 Strict Morphological Passivization

In order to make passive sentences, sometimes Ilami speakers use a specific construction. Have a look at the following examples:

- (14) a. ej kwəl-ə hadʒæt-el-æ fur-t.  
 s/he all-POSS dish- PL- DEF wash- 3SG PAST  
*S/he washed all the dishes.*  
 b. kwəl-e hadʒæt-el-æ fur-ija.  
 all-POSS dish- PL- DEF wash- PASS PAST  
*All the dishes were washed.*
- (15) a. rusæm dæ nam dʒængan suraw kwəf-t.  
 Rusam into battle Suraw kill-3SG PAST  
*Rusam killed Suraw in a battle.*  
 b. suraw dæ nam dʒængan kwəf-(ə)rija.  
 Suraw into battle kill- PASS PAST  
*Suraw was killed in a battle.*
- (16) a. bætf-el-æ qæza-(g)æ xwæ-n.  
 child- PL- DEF food- DEF eatPRES-3PL  
*The children eat the food.*

b. qæza-(g)æ xwær-e.  
food- DEF eat- PASS PRES  
*The food will be eaten.*

(17) a. æli mɑʃin-æ ʃə|pən-e.  
Ali car- DEF shatter- 3SG PRES  
*Ali will shatter the car.*

b. mɑʃin-æ ʃə|p-e.  
Car- DEF shatter- PASS PRES  
*The car will be shattered.*

Examples (14) through (17) show active and passive counterparts in different semantic contexts. As we can see in all the examples mentioned so far, a given morphological affix (*-ija* in the past and *-e* in the present or future tense) is attached to the verb root in order to make the passive form of the verb, hence, they are portmanteau morphemes with the function of time and voice denotation. Actor as the default PSA, which is present in all examples, tends to be omitted in the passive constructions and undergoer (here direct object) adopts the PSA role and appears in the subject position.

If we take example (14) into account, “*ʃurijan*” as the passive form of the verb is in agreement with the marked PSA which is “*kwæle hadʒætælæ*” and it shows that morphologically, there is no longer agreement between the verb and the unmarked PSA (that is “*ej*” in the active construction) and this comes true regarding other examples too.

Respecting unmarked PSA (actor), it seems that, unlike English, Kurdish passive constructions usually push away any unnecessary element out of the core structure. This is why we can see the previously privileged argument (actor) in none of the passive examples mentioned thus far (even in the periphery). This is not clear why Ilami passive constructions often drop the actor out of the core structure, however an assumption could be related to the minimization of the actor role which is reflexed through the deletion of this element, or they adhere to the economy principle of the language. We



should keep in mind that in a few cases, the appearance of unmarked PSA (in the periphery) is optional which is believed to be a non-local pattern entered Ilami Kurdish and could be functionally used in a more or less style:

- (18) suraw ( wæ dæsə        **rusæm**) kwəf- (**ər**)ija.  
 suraw with hand- POSS rusam    kill- PASS PAST  
 ACTOR

*Suraw was killed by Rusam.*

- (19) mafin- æ wæ dæs- ə        **æli**    fə|p-e.  
 car- DEF with hand- POSS    Ali    shatter- PASS PRES  
 ACTOR

*If Ali is the owner of the car, it will be shattered.*

Importantly, native speakers make a distinction between the two examples mentioned above. In fact, the first example is a declarative sentence which optionally preserves the unmarked macro role which is “Rusam” in the periphery, while the latter represents a (semantically) conditional sentence meaning “if Ali is the owner of the car, it will be shattered”.

The logical structure of the aforementioned example could be so:

- (20) suraw ( wæ dæsə        **rusæm**) kwəf- (**ər**)ija.  
 Do (Rusæm, [do' (Rusæm, Ø)]) CAUSE [BECOME **dead'** (Suraw)]

It should be hinted that the definite marker æ attached to the undergoer is determinant in the construction of sentence meaning. Should we detach the definite marker æ from the undergoer in example (19), the action done by *Ali* would be inferred as a habitual action and consequently the passive counterpart seems to be highly unnatural and marked, if not ungrammatical:

- (21) \*<sup>2</sup>maqjin jəlp-e.  
car shatter- PASS PRES  
*car will be shattered.*

### 6.3 Periphrastic passivization

So far, we represented how morphologically an active sentence could change to a passive one through adding certain morphological suffixes to the verb root. Now let us see if periphrastic passivization also exists in Ilami. By analyzing Kurdish data, we notice that some sentences could change to periphrastic passives and here are some examples of this kind:

- (22) a. mukət- æ qəjtji kər-d-əm.  
carpet- DEF cut- PAST-1SG  
*I cut my hand.*  
b. mukət- æ qəjtji(j)-aw bö.  
carpet- DEF cut become- PASS PAST

In example (22), *əm* as the actor of the sentence occupies the PSA position and is omitted in the passive sentence, as the PSA role is taken by *mukətæ* which is now at the center of attention. Remind that Keenan & Dryer (2007) claimed a basic periphrastic passive consists of an auxiliary verb plus a strict morphological function of a transitive verb and one class could occur when “the auxiliary verb (of the sentence) is a verb of being or becoming”. Based on this classification, the above illustration could be categorized as a periphrastic passive. It seems that Ilami Kurdish does not tend to use a purely syntactic operation to construct such sentences; instead a morphosyntactic rule seems to be involved in this operation. This claim is based on the fact that in such examples where “X- *böjən*”(becoming) is the verb of the periphrastic constructions, there is a morphological suffix *-aw* attached to the first element of the verb.

---

<sup>2</sup> \* is used to represent markedness and/ or unacceptability.

Even though it may not be the case in all Kurdish verbs, a morphosyntactic rule could be generalized at least to those verbs owning *kərdən* as the head of the compound verb. In other words, when the verb of the active sentence is “X+ *kərdən*”, this sentence is prone to be morphosyntactically constructed. Have a look at the following examples:

- (23) a. *zenaw ləbas-el-æ kwət kwət kərd.*  
 Zenaw clothe- PL- DEF tear- 3SG PAST  
*Zenaw tore the clothes into pieces.*
- b. *ləbas-el-æ kwət kwətaw bön.*  
 clothe- PL- DEF cut become- PASS PAST  
*The clothes were/ become torn.*

- (24) a. *zenaw ləbas- el- æ dər-i.*  
 Zenaw clothe- PL- DEF tear- 3SG PAST  
*Zenaw tore the clothes into pieces.*
- b. *ləbas-el-æ dər-ija-n.*  
 clothe- PL- DEF cut- PASS PAST- 3PL  
*The clothes were/ become torn.*

It might be worth mentioning that without knowing the active version of the sentences, the passive counterparts seem to be ambiguous, because it is unclear whether an activity has performed upon the undergoer or not. In other words, we are unable to judge whether the verb describes a state or a dynamic action.

#### 6.4 Impersonal Passive Constructions

There are some Kurdish constructions which are specific in some facets. Actually, these sentences are “impersonal passive constructions” as they lack an overt actor in their syntax. This fact is worth taking a look at:

(25) 'tə|ɑ bər-d-ən.

gold rub- PAST- 2/3PL

*(you/ they) stole jewelry.*

(26) 'mɑ|-æ da-n-æ dæm-ə kwəftijaw.

house- DEF give- PAST- 2/3PL mouth-POSS wrestling.

*(you/ they)made the house messy.*

The verbs of the mentioned examples are conjugated in plural 2<sup>nd</sup> and 3<sup>rd</sup> person and clearly there is no agreement with the undergoer of the sentence. These constructions are referred to as impersonal passive constructions lacking a definite actor but the passive voice of the sentence is inferable. It is worthwhile mentioning that in Ilami Kurdish, 2<sup>nd</sup> and 3<sup>rd</sup> person morphemes are pronounced identically and in this regard, Ilami is different from other languages in which the verb of impersonal passive constructions takes 2<sup>nd</sup> and 3<sup>rd</sup> person morphemes.

As far as accentuation is taken into consideration, it is a must for such constructions to give the undergoer of the sentence the main stress which we have shown it here with an acute stress. Based on this criterion, it can be said that undergoer in such cases is focalised because of conveying new information in the sentence and correspondingly indefinite actor is deleted due to conveying unnecessary information. For example, in example (25) and (26) 'tə|ɑ' and 'mɑ|' as the undergoers of the sentences have been focalised as the new information of the construction. Additionally, the undergoer may (or may not) take a definite marker, like "æ" for 'mɑ|' and nothing for "tə|ɑ" respectively.

## **7. Concluding remarks**

In this paper, we attempted to show the way passive forms are made in Ilami Kurdish. Analyzing our data, we found out that passive construction in Ilami is the role of morphology and its interface with syntax. In other words, in the morphology module the

main type of passive form which is adding an affixal morpheme to the verb root is constructed and many verbs could be passivized in this manner. It was also shown that in the respective construction, actor is usually deleted out of the core construction which is a type of argument modulation. On the other hand, impersonal passive constructions could be made via a syntactic operation which is the deletion of PSA (that is actor), though, the verb inflexion remains intact which is a type of argument modulation. Although PSA role is not handed to the undergoer of the new construction, these forms should be considered as passive forms as the actor of the sentence is prone to drop and is morphologically conflated onto the main verb. Last but not the least, regarding the periphrastic passives, it was noticed that this sort of passive does exist in Ilami Kurdish and they are made through a morphosyntactic operation. Transferring the above findings to RRG framework, we get the following information: Firstly, regarding two phenomena of the universal formulation of basic voice oppositions (PSA and argument modulation), it can be said that in all forms of Ilami passives including strict morphological passives and periphrastic passives, both stages are done. In fact, in these cases, the actor is omitted or is rarely placed in the periphery and the undergoer becomes the PSA of the sentence. This tells us that in Ilami passives the default PSA is the actor and in specific contexts (passive voice), undergoer is permitted to function as the PSA.

## References

- ASATRIAN, G. (2009) "Prolegomena to the Study of the Kurds", *Iran and the Caucasus*, 13, 1-58.
- BYNON, T. (1979) "The ergative construction in Kurdish", *Bulletin of the School of Oriental and African Studies*, 42(2), 211-224.
- BOIS, T. (1966) *The Kurds*, Beirut, LB: Khayats.
- CRYSTAL, D. (2008)<sup>6</sup> *A Dictionary of Linguistics and Phonetics*, Malden, MA: Blackwell Pub.
- ESTEBAN, A. C. (2012) "Does There Exist Passive Voice in Lakhota and Cheyenne?", *Revista de Lingüística y Lenguas Aplicadas*, 7, 77-96. <<http://dx.doi.org/10.4995/rlyla.2012.1126>> (accessed 11 February 2015).

- FILLMORE, C. (1968) "The Case for Case", in E. Bach & R. Harms (eds.), *Universals in Linguistic Theory*, New York: Holt, Rinehart & Winston, 1-88.
- FOLEY, W. & R. D. VAN VALIN Jr. (1977) "On the Viability of the Notion of 'Subject'" in *Universal Grammar*, *Proceedings of the Third Annual Meeting of the Berkeley Linguistics Society*, 3, 293-320.
- FOLEY, W. & R. D. VAN VALIN Jr. (1984) *Functional Syntax and Universal Grammar*, Cambridge: Cambridge University Press.
- GUNTER, M. (2004) *Historical Dictionary of the Kurds*, Maryland: Scarecrow Press Inc.
- GRUBER, J. (1965) *Studies in Lexical Relations*, Unpublished Ph.D. dissertation, MIT.
- JÜGEL, T. (2009) "Ergative Remnants in Sorani Kurdish?", *Orientalia Suecana*, LVIII, 142-158.
- KEENAN, E. L. & M. S. DRYER (2007)<sup>2</sup> "Passive in the World's Languages", in Timothy Shopen (ed.), *Clause Structure, Language Typology and Syntactic Description*, vol. 1, Cambridge: Cambridge University Press, 325-361.
- LEISIÖ, L. (2006) "Passive in Nganasan", in W. Abraham W. & L. Leisiö (eds.), *Passivization and typology: form and function*, Amsterdam: John Benjamins, 213- 230.
- MACKENZIE, D. N. (1963) "Kurmandzi, Kurdi, Gurani", *Narody Azii I Afriki*, 1, 162-170.
- MANLEY, T. (1972) *Outline of Sre Structure*, Hawaii: Oceanic Linguistics Special Publications.
- NOLAN, B. (2005) "The Role of Definiteness in the Impersonal Passives of Modern Irish: Towards an RRG Characterization", *Paper presented at the 2005 International RRG Conference*, Taipei: Academia Sinica.
- ORANSKIJ, I. (1979) *Iranskie Jazyki v Istorčeskom Osveščeenii*, Moskva: Nauka.
- REZAI, V. (2010) "Passive Constructions in Persian: a New Perspective", *Research in Linguistics*, 2 (1), 19-34.
- VAN VALIN Jr., R. D. (1977) *Aspects of Lakhota Syntax*, Unpublished Ph.D. dissertation, University of California, Berkeley.
- VAN VALIN Jr., R. D. (1981) "Grammatical Relations in Ergative Languages", *Studies in Language*, 5, 361-94.
- VAN VALIN Jr., R. D. (2004) *An Introduction to Syntax*, Cambridge: Cambridge University Press.
- VAN VALIN Jr., R. D. (2005) *A Summary of Role and Reference Grammar*. <<http://linguistics.buffalo.edu/research/rrg.html>> (accessed 9 February 2015).
- VAN VALIN Jr., R. D. (1992) *An Overview of Role and Reference Grammar*, University of Buffalo, the State University of New York.

VAN VALIN Jr., R. D., & R. J. LAPOLLA (1997) *Syntax: structure, meaning & function*, Cambridge: Cambridge University Press.

WINDFUHR, G. (2009) *The Iranian Languages*, New York and London: Routledge.

Received 4 January 2016.

Accepted 30 October 2016.

## **“REAL” DOWN-HOME SOUTHERN ENGLISH: COMPARING AN OKLAHOMAN’S REAL AND IMITATED DIALECTS<sup>1</sup>**

Justin T. McBRIDE

Northeastern State University\*

mcbrijed@nsuok.edu

### **Abstract**

The Research on the Dialects of English in Oklahoma (RODEO) project offers a view of local perceptions of dialects in Oklahoma, USA. “Beth” (Female, 46, Watts, Oklahoma) spontaneously read the same passage in both her local, Southern-influenced English variety and a pretend, “real down-home Southern” variety. Spectra from both performances were analyzed to determine pitch, intensity, and formant values of stressed vowels and the length of words containing them. Beth’s local and exaggerated Southern performances were compared using paired t-tests. While many differences were insignificant, Beth’s imitated variety was characterized by FACE vowel onset centralization, THOUGHT vowel raising, and increases in time and intensity. Sociolinguistic interview data was also analyzed to reveal attitudinal evidence for the contrast. Beth appears to reveal weak perceptual associations with stereotypic Southern English phonological features but exhibits a complicated attitudinal relationship to the local speech community and an awareness of registers within the local repertoire.

### **Keywords**

Southern American English, Oklahoma, language regard, perceptual dialectology, performance

---

<sup>1</sup> An earlier version of this article appeared in the conference proceedings the University of Central Oklahoma's 2012 Language and Linguistics Student Conference, Edmond, Oklahoma, as follows: McBride, Justin T. (Winter 2013). “‘Real’ down-home southern English: Comparing an Oklahoman’s real and imitated dialects”, *The ScissorTale Review: The Language and Linguistics Student Conference*, 101-107.

\* 3100 E New Orleans St, Broken Arrow, OK, USA 74014.



**INGLÉS DEL SUR “REAL” CASERO:  
COMPARANDO LOS DIALECTOS REALES E IMITADOS DE UN NATIVO DE OKLAHOMA**

**Resumen**

El proyecto de investigación sobre los dialectos del inglés en Oklahoma (RODEO) ofrece una visión de las percepciones locales de los dialectos en Oklahoma, EE.UU. “Beth” (mujer, 46 años, Watts, Oklahoma) leyó espontáneamente el mismo pasaje tanto en su variedad local inglesa influenciada por el sur como en una pretendida variedad “real casera del sur”. Se analizaron los espectros de ambas interpretaciones para determinar los valores de tono, intensidad y formante de las vocales acentuadas y la longitud de las palabras que las contenían. Los resultados locales y exagerados sureños de Beth se compararon mediante t-tests emparejadas. Mientras que muchas diferencias fueron insignificantes, la variedad imitada de Beth se caracterizó por la centralización de la iniciación de la vocal, elevación de la vocalidad y incremento en tiempo e intensidad. Los datos sociolingüísticos de las entrevistas también fueron analizados para revelar evidencias actitudinales en el contraste. Beth parece revelar asociaciones perceptivas débiles con rasgos fonológicos estereotípicos del inglés del sur, pero muestra una complicada relación actitudinal con la comunidad del habla local y una conciencia de los registros dentro del repertorio local.

**Palabras clave**

inglés americano del sur, Oklahoma, conciencia lingüística, dialectología perceptual, actuación

**1. Introduction**

Dialect research frequently relies on structural descriptions of language varieties, but individuals’ perceptions and behaviors relating to dialect varieties also deserve attention, mostly as a means of gauging linguistic attitudes. Generally speaking, perception is the mechanism by which individuals internalize the language experience, and behavior—both conscious and less than conscious—is the mechanism by which individuals externalize it. However, the two are merely the most tangible components of a larger system of language regard characterized by beliefs about and attitudes toward language varieties, their speech communities, and members of those communities. Preston argues that such a language regard system “interacts with or, better, influences language production and comprehension, particularly the latter” (Preston 2011: 10). If so, then beliefs and attitudes are at the very heart of the language experience, not only

as passive mental filters but as active drivers of both the perceptions and behaviors of individuals and groups alike. Preston distinguishes these two kinds of language regard tenets, belief and attitude, as follows: “[B]eliefs are not necessarily evaluative, and evaluation is taken to be a necessary component of attitude” (Preston 2011: 10). Additionally, Garrett describes attitudes as “having a degree of stability that allows [them] to be identified” (Garrett 2010: 20). Attitudes are, then, essentially value judgments—positive or negative, strong or weak, conscious or less than conscious—as situated within particular social contexts against which stability<sup>2</sup> can be assessed through analysis of the perceptions and behaviors they presumably influence.

Consider the dialectal context of English in present-day Oklahoma, a largely rural state located in the southern Great Plains region in the central U.S. The dialect situation in Oklahoma is not well documented in the scholarly literature, though structural descriptions of English varieties in the state (e.g., Bailey, Wikle, Tillery & Sand 1993; Bakos 2013; Labov, Ash & Boberg 2006; Tillery & Bailey 1998; Weirich 2013; Wikle & Bailey 1997) are more readily available than attitude-oriented studies (e.g., Bakos 2013; McBride 2015). As it stands, structural research demonstrates that the state exhibits certain features—mostly phonological and lexical—of both the South and Southern Midlands dialect regions. In terms of phonology, this includes fronting of the GOOSE, FOOD, and GOAT vowels,<sup>3</sup> a feature associated with both regions; occasional reversal of the DRESS/FACE onsets, a Southern but not Midlands feature; and merger of LOT/THOUGHT vowels, a Midlands but not Southern feature. Yet, the exact details of this dialectal mix, especially speaker attitudes about these and other features, are poorly known. For this reason, Dennis Preston founded the Research on Dialects of English in Oklahoma (RODEO) project at Oklahoma State University to collect and analyze dialect and dialectal attitude data from across the state. It is hoped that the

---

<sup>2</sup> Garret (2010) defines attitudes in terms of stability, yet Bassili & Brown (2005) problematize stability thusly: “One may think, for example, that feelings that are accessible, that are held with certainty, that are considered important to the self, and that are felt with intensity, ought to resist the influence of suggestions contained in an attitudinal query. Under most circumstances, however, this is not the case” (Bassili & Brown 2005: 547). Nevertheless, a full treatment of stability is somewhat tangential to the discussion at hand insofar as attitudes, while clearly malleable over time, appear capable of some sort of assessment through contextualized behavior on a moment-by-moment basis.

<sup>3</sup> Note that vowels in this article are represented using the Standard Lexical Sets for English vowels (cf. Wells 1982).

RODEO project will help to clarify understanding of dialects and their speakers' perceptions in Oklahoma.

One of the early subjects in the RODEO study is "Beth," a then 46-year old single female living and working as an artist and homeless shelter supervisor in Tulsa, OK. Beth is of Anglo ethnicity and has a baccalaureate degree. One of the things that makes Beth of great interest is that her isolated, rural hometown of Watts, where she lived until finishing high school, is located along a theorized line of isoglosses that Labov, Ash, & Boberg (2006: 129) suspect divides the Southern Midlands dialect region from two proper South dialect regions, the Inland South and the Texas South (see Figure 1). If these isoglosses truly exist then she may be expected to exhibit specific perceptions and/or behaviors of either Southern or Midlands dialects that may differ from those who grew up farther from this proposed border.

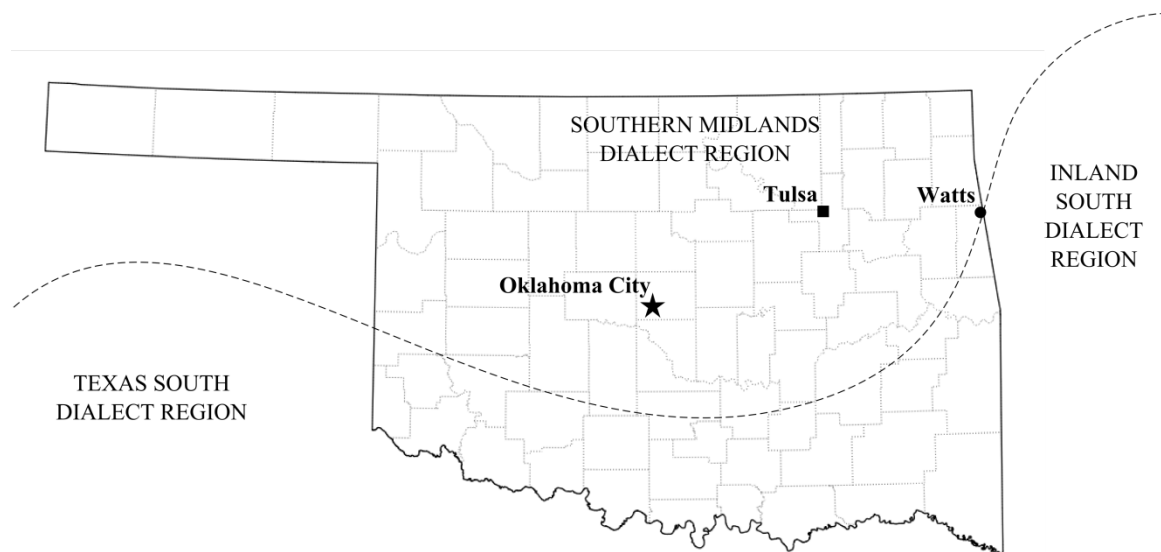


Figure 1. Watts and Oklahoma dialect borders (adapted from Labov, Ash & Boberg 2006)

Another intriguing fact about Beth occurred during her sociolinguistic interview, which a RODEO field linguist digitally recorded under unknown conditions at Beth's Tulsa home in September of 2009. The interview included a brief reading passage that Beth read twice. After first reading it through using her natural dialect, she then volunteered to read it through a second time using an English variety that she termed

“real down-home Southern Oklahoma” (see Appendix). Her decision to do this offers the unique opportunity to examine not only the structure of both her real and imitated varieties (henceforth termed *Real* and *Imitated* to distinguish the performances from the common adjectives), but potentially her attitudes towards a Southern—using her term of preference—perceptual prototype. Thus, by examining both of Beth’s reading passage performances we can presumably learn a great deal about how she situates herself within Oklahoma’s dialect context, both in terms of reception and production.

The purpose of this study, then, is to investigate three problems involving Beth’s reading passages. They are as follows:

1. What exactly are the behavioral differences between the two performances?
2. What do these differences tell us about Beth’s perceptions of “real down-home Southern Oklahoma” English?
3. What do those perceptions and behaviors tell us about Beth’s attitudes about the English varieties in her local environment?

## **2. Method**

This study employed both quantitative and qualitative methodologies to present a fuller picture of Beth’s dialectal behaviors in both her Real and Imitated performances and her perceptions, as far as they can be determined.

### *2.1 Quantitative Procedures*

The quantitative component consisted of two parts, non-instrumental and instrumental comparisons of the two performances. The non-instrumental comparison was limited to phonological reduction and lexical substitution phenomena that could be detected by simply transcribing the performances and making note of differences. The most obvious feature of interest in this phase was [ING] realization. The instrumental comparison involved the use of PRAAT (Boersma 2001) to analyze performance data. The original digital sound capture was imported into the program, resampled at a

sampling rate of 10,000 Hz, and then analyzed word-by-word in terms of the following primary variables: Word length (s), mean pitch (Hz), mean intensity (dB), and vowel quality, including F1 and F2 frequencies (Hz) and various monophthongization and diphthongization concerns. For the instrumental comparison, only words containing stressed vowels were of interest (N = 87). Two-tailed paired sample t-tests were then used to determine the statistical significance of the differences between the two performances.

## *2.2 Qualitative Procedures*

The qualitative component consisted mostly of classical content analysis of the interview (Bauer 2000). The purpose of this analysis was to identify themes as evidence of Beth's attitudes toward either the local dialect variety—i.e., Real—or her “real down-home Southern Oklahoma” variety—i.e., Imitated. To accomplish this, the entire interview of approximately 33 minutes was transcribed (a close-vertical but not phonetic transcript was used for this purpose) and then analyzed with thematic categories emerging organically from the process. Qualitative findings will be addressed in the Discussion section below.

## **3. Results**

### *3.1 Non-Instrumental*

Table 1 represents the primary results of the non-instrumental comparison of the two performances for each of three variables, lexical or phonological reduction, lexical substitution, and [ING] realization as a special case of phonological reduction.

Feature	Real	Imitated
Phonological/lexical reduction	<i>them</i>	<i>'em</i>
	<i>remembered</i>	<i>'membered</i>
	<i>garage</i>	<i>g'rage</i>
	<i>that</i> [relative pronoun]	$\emptyset$
Lexical substitution	<i>(the) Wal-Mart</i>	<i>Wally World</i>
[ING] realization	<i>planning</i>	<i>plannin'</i>
	<i>shopping</i>	<i>shoppin'</i>
	<i>going (to)</i>	<i>gonna</i>
	<i>baking</i>	<i>bakin'</i>

Table 1. Non-Instrumental performance comparisons

Nine types were found to vary between performances, one token per type. The Imitated performance exhibits more reduction and substitution, and less realization of [ING]. Note that at no time in the reading passage that [ING] occurred did Beth realize it as *-ing* in Imitated, and at no time in Real did she reduce it. Be also aware that all of these phenomena are associated with informal registers of American English.

### 3.2 Instrumental

Figures 2 through 4 represent the results of instrumental comparison of the stressed words from the two performances. In each graph, the blue line represents Beth's Real variety, and the red line represents her Imitated variety. Figure 2 shows the comparison of mean F0 across the stressed vowel in each word token, listed as W1 through W87. While there were scattered areas of difference between the two performances, the overall difference was not found to be significant (Real: M = 181.439 Hz, SD = 44.594; Imitated: M = 187.117 Hz, SD = 41.133;  $p = 0.117$ ).

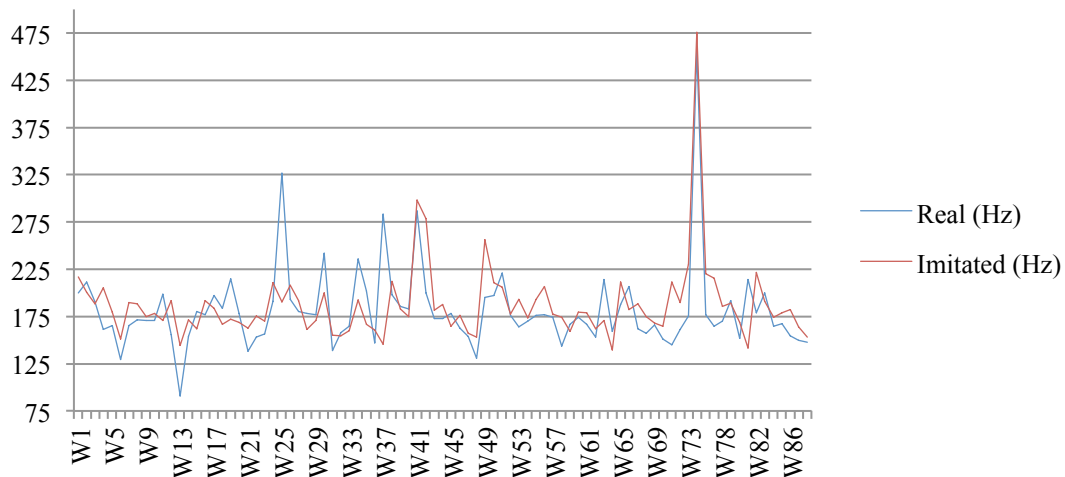


Figure 2. Mean F0 of stressed vowels (n.s.)

Figure 3 shows the comparison between the lengths of stressed words in each performance. Here, the visual impression of difference corresponds to a statistically significant increase in Imitated word length ( $p < 0.01$ ), perhaps mimicking the so-called Southern Drawl. On average, the Imitated words are around 3 ms longer (Real:  $M = 0.316$  s,  $SD = 0.140$ ; Imitated:  $M = 0.349$  s,  $SD = 0.148$ ;  $p = 0.002$ ). So great is the difference that the Imitated performance is approximately 2 s longer than the Real.

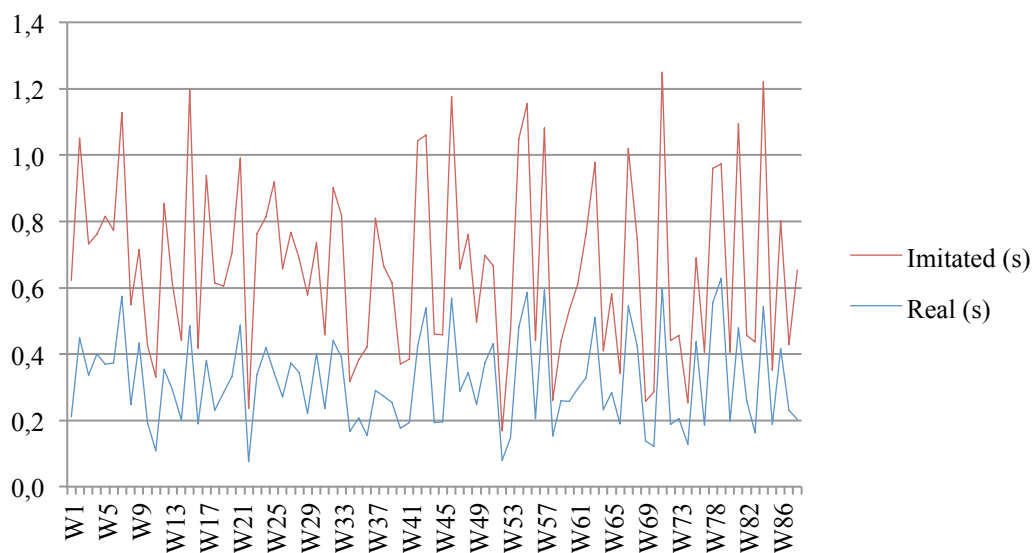


Figure 3. Stressed vowel word length ( $p < 0.01$ )

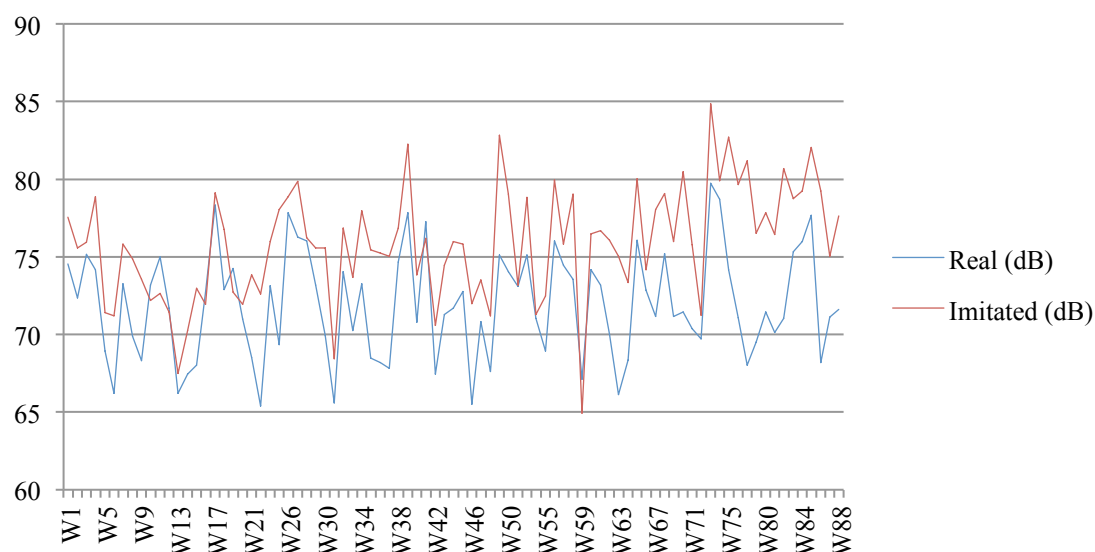


Figure 4. Mean intensity of stressed tokens ( $p < 0.001$ )

Figure 4 shows the distribution of mean intensity of stressed vowel tokens. Note that two tokens, W23 and W53, were not sufficiently stressed with respect to their external environments to allow for the calculation of intensity. Again, Imitated tokens show a visually discernible and significant increase ( $p < 0.001$ ). On average, the difference between individual pairs from each performance was nearly 4 dB more forceful in Beth’s imitation (Real:  $M = 71.890$  dB,  $SD = 3.432$ ; Imitated:  $M = 75.857$ ,  $SD = 3.654$ ;  $p = 1.607 \times 10^{-20}$ ). Be aware, however, that the recording reveals Beth’s noticeable animation before and during her Imitated performance presumably due to her excitement in offering an unsolicited interpretation of what she regards as a “real down-home Southern” variety.

VOWEL	N	Real F1		Imitated F1		F1	Real F2		Imitated F2		F2
		$M^a$	SD	$M^a$	SD	$p$	$M^a$	SD	$M^a$	SD	$p$
TRAP	6	624	63	<b>645</b>	93	0.714	1863	201	<b>2009</b>	141	0.250
LOT	10	649	43	<b>670</b>	55	0.196	1078	133	<b>1090</b>	82	0.677
DRESS	13	570	44	<b>573</b>	47	0.805	1734	187	<b>1801</b>	283	0.275
FLEECE	9	423	35	<b>449</b>	82	0.218	<b>2408</b>	123	2378	238	0.570
KIT	14	<b>531</b>	39	531	61	0.973	<b>1865</b>	327	1863	305	0.987
THOUGHT	4	<b>713</b>	16	701	15	$0.002^b$	<b>1061</b>	24	1056	31	0.783



FOOT	3	561	32	<b>579</b>	32	0.156	<b>1453</b>	142	1400	59	0.464
STRUT	8	663	63	<b>674</b>	47	0.650	1439	177	<b>1457</b>	174	0.374
GOOSE	2	<b>428</b>	58	406	2	0.698	1645	684	<b>1670</b>	761	0.721
NORTH	2	452	21	<b>545</b>	34	0.252	790	32	<b>904</b>	81	0.181

Notes: *a* Larger means between pairs are shown in boldface for clarity  
*b* Statistically significant ( $p < 0.01$ ); italicized for clarity

Table 2. Monophthongal vowel quality (F1 and F2 in Hz)

As far as vowel quality is concerned, one of the most noteworthy aspects of both of Beth's performances was how similar they were. Indeed, most vowels showed no significant difference whatsoever, in either F1 or F2, from one performance to the next. Tables 2 and 3 show comparisons of mean monophthongal and mean diphthongal vowel quality, respectively. The only significant monophthongal difference (Table 2) is a reduction in F1 frequency for the THOUGHT vowel in the Imitated variety ( $p < 0.01$ ), indicating raising from Real.

Likewise, the diphthongs (Table 3) show almost no change from one performance to another. Nevertheless, there is a difference with respect to the F2 of the FACE vowel onset ( $p < 0.01$ ) wherein Imitated experiences a decrease, indicative of a centralization of this vowel with respect to Real. Several other diphthongs, specifically, the PRICE onset and offglide and the GOAT offglide, approach significance ( $0.10 < p > 0.05$ ) but do not reach the threshold. Perhaps a larger number of tokens in a longer reading passage would have borne out this difference; the PRICE vowel, after all, such as in the word 'price,' is typically realized as [p<sup>h</sup>ɹa:s] in advanced Southern Shift speech.

VOWEL	N	Real F1		Imitated F1		F1 <i>p</i>	Real F2		Imitated F2		F2 <i>p</i>
		M <sup>a</sup>	SD	M <sup>a</sup>	SD		M <sup>a</sup>	SD	M <sup>a</sup>	SD	
<i>Onsets</i>											
PRICE	5	629	93	<b>707</b>	45	0.086	1218	241	<b>1277</b>	112	0.524
MOUTH	3	617	58	<b>632</b>	68	0.259	1798	208	<b>1823</b>	244	0.353
FACE	6	625	32	<b>640</b>	37	0.154	<b>1831</b>	228	1717	220	<i>0.010<sup>b</sup></i>
GOAT	6	545	94	<b>585</b>	60	0.170	1104	278	<b>1192</b>	342	0.552

Offglides											
PRICE	5	573	32	<b>625</b>	86	0.237	<b>1915</b>	197	1715	329	0.082
MOUTH	3	651	30	<b>666</b>	36	0.710	1230	94	<b>1297</b>	72	0.161
FACE	6	<b>483</b>	41	465	29	0.405	2263	98	<b>2308</b>	82	0.295
GOAT	4	488	60	<b>548</b>	31	0.057	1107	274	<b>1151</b>	254	0.809

Notes: a Larger means between pairs are shown in boldface for clarity  
 b Statistically significant ( $p < 0.05$ ); italicized for clarity

Table 3. Diphthongal vowel quality (F1 and F2 in Hz)

Figure 5 shows a combined, non-normalized plot of the vowel means from both performances. It is curious to note that the significantly different vowels (THOUGHT and FACE) appear very close to one another whereas those that are statistically insignificant are in many cases quite distinguished from one another. This is mostly a function of the low N number associated with each token. Perhaps in a larger sample of Imitated speech, Beth would have demonstrated consistently greater separation of vowels throughout the spectrum.

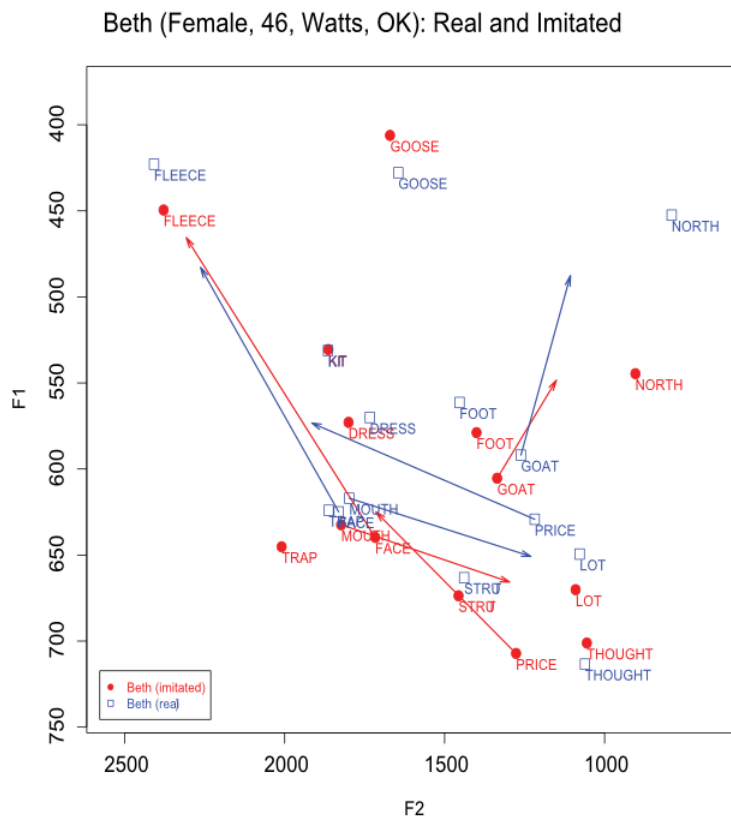


Figure 5. Combined plot of all vowel onset means from both performances

There are several other things to note about the general plot of vowel means. For starters, be aware of the general reversal of DRESS and FACE onset, indicative of the so-called Southern Shift (Figure 6), which is further characterized by the reversal of KIT and FLEECE onset, raising of TRAP, and additional movement of both PRICE and LOT. While Beth's TRAP vowel is noticeably raised (as is her non-significant Real PRICE, see above), and while her DRESS and FACE onsets have changed places, in neither performance do the other Southern Shift-participating vowels undergo characteristic movement. This would indicate that Beth is only a partial participant in the greater Southern Shift. Other Southern features in her vowels include fronting of GOOSE, GOAT, and even FOOT, but these have become widespread throughout the western U.S.

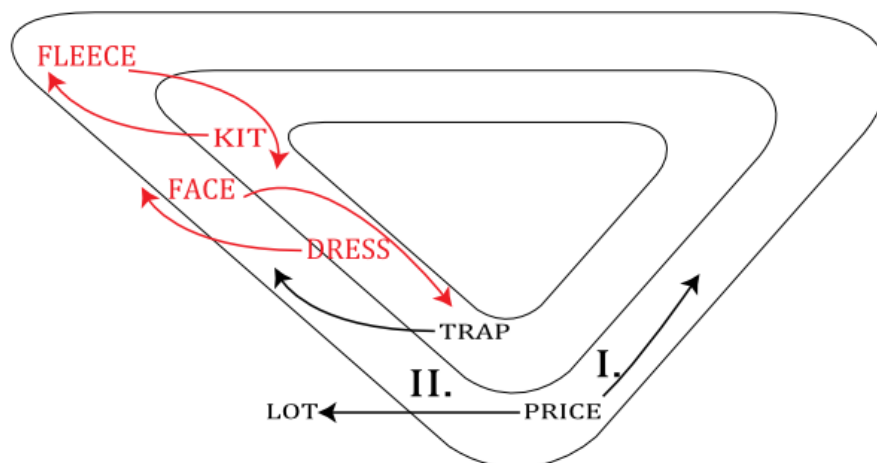


Figure 6. Southern Shift (adapted from Labov, Ash, & Boberg, 2006: 125)

Nevertheless, a careful glance at the vowel means exhibiting significant differences from one performance to another (Figure 7) reveals two interesting implications. First, the centralization of the FACE onset in the Imitated variety places greater diphthongal contrast between the onset and the offglide, with the onset moving closer to STRUT. This more central destination is the expected result of the Southern Shift of FACE. Second, the raising of THOUGHT in the Imitated variety increases the contrast between it and LOT. Merger of these two vowels is a feature of the Southern

Midland, not Southern proper dialects. Thus, the net effect of both of these significant differences is a movement toward what would be expected of Southern speech from one who is only a partial participant in its broader phonological vowel behaviors.

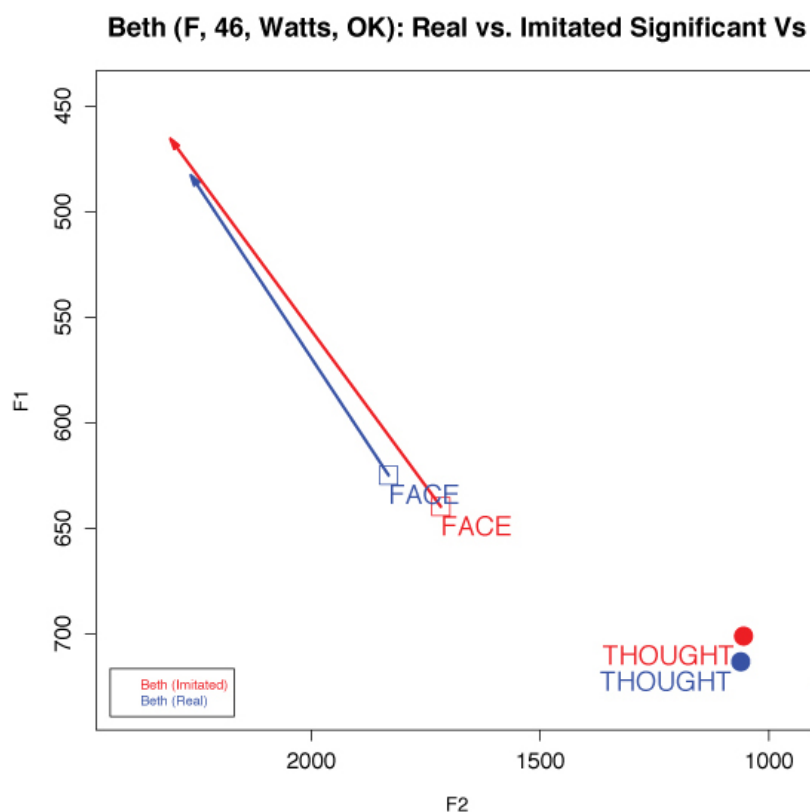


Figure 7. Combined plot of significant vowel onset means from both performances

## 4. Discussion

### 4.1 Quantitative Summary

The purpose of this study has been to examine the differences between Beth's two reading passage performances, including evidence of perception and attitudes about the dialects involved. So far, only the quantitative results have been discussed. Key findings from this component have included a number of features about Beth's "real down-home Southern" variety, such as the following: a 10.4% increase in mean word length; a 5.5% increase in mean intensity, 0% realization of [ING] as *-ing*; THOUGHT

raising; FACE onset centralization; and numerous reductions and/or substitutions (e.g., *gonna*, *'em*, *'membered*, *Wally World*). But what do these phenomena have to say about Beth's perceptions of either the Southern Midlands dialect or the relevant dialects of the South?

#### 4.1.1 Perception Implications

To begin with, the word length increase in the Imitated variety is probably a reflection of Beth's perception of the so-called Southern Drawl, which involves, in phonetic terms, the diphthongization of certain historical monophthongs, but which, to the layman, may simply sound like slower speech in general. This may be reinforced by the fact that the Southern monophthongization of PRICE is typically associated with a lengthening of the vowel. Similarly, her general F1 reversal of DRESS and the FACE onset and the subsequent centralization of the FACE onset in the Imitated variety both produce greater contrast between front onsets and offglides, a feature associated with Southern Shift movement and diphthongization of the front diphthongs and historical monophthongs. Her generally raised TRAP also participates in the Southern Shift of front vowels, as does her fronting of GOOSE and GOAT (as well as FOOT, though this is not historically typical of Southern English; cf. Wolfram & Schilling-Estes, 2006: 149). Her Imitated raising of THOUGHT, too, contributes to greater contrast between low back vowels, which have merged in much of the western U.S. but not in the South (Wolfram & Schilling-Estes, 2006: 147). All of these features would seem to indicate that Beth perceives herself as something of a Southern speaker, but capable of being much more Southern. It also demonstrates a subtle but nuanced understanding of certain Southern vowel phenomena.

Be aware, however, that other features of Beth's Imitated speech do not necessarily index Southern English. These features include her phonological and lexical reductions and substitutions, including her [ING] realization. Rather, they are usually associated with informal register. This suggests that Beth perceives Southern speakers as being more informal than at least Southern Midlands speakers. Additionally, Beth's Imitated variety was much more forceful (i.e., exhibiting a higher mean intensity) than

her Real variety. Perhaps this reveals perceptual associations with the South and loud talk or possibly even masculinity. More likely, though, this is probably an effect of her excited performance. It is a feature worth bearing in mind, at any rate.

#### 4.2 Qualitative Analysis

If the features described above imply certain perceptions about the dialects of Oklahoma, then one should be able to see reinforcement of the attitudes underlying the perceptions elsewhere in Beth's interview. Excerpts 1 through 4 present some such evidence. In sum, she ultimately displays somewhat contradictory attitudes about her local variety and what she identifies as Southern accents throughout her interview, ranging from the negative (Excerpts 1 through 3) to the positive (Excerpt 4). Note that in these excerpts, bolding is used to draw attention to key words and phrases.

With respect to negative attitudes about her normal English variety, Beth describes the situation in which, while living and working at an art gallery in Palo Alto, California, she first became conscious of it (Excerpt 1).

##### Excerpt 1. Experiences in California and in Oklahoma

1 Beth: It really never occurred to me until I moved out of state. I graduated from  
2 college, and went to California after I got my art degree, and I worked in,  
3 um, an art studio and framing shop. It was a gallery. And, um, I noticed  
4 when I would do customer service out front, people were looking at me  
5 strangely. And I really, honestly did not have a clue. Um, and finally one  
6 day, I just got really bothered by it and went back to the back to talk to my  
7 manager <laughs> about it. And he got really tickled, and he told me. He  
8 said, "You really don't <laughs> get it, do you?" And I said, "No. I mean,  
9 what?" <sniffs> And he's, he, he explained, you know, that, "Wow, I mean  
10 that you really have, um, a Southern accent!" And he was being nice  
11 about it, but, basically, he was telling me that I sounded like a hick.  
12 <laughs> And, uh, it- And I really never got it until I lived in California for  
13 a little while, and had been around Californians, came back here, and then,

14 um, moved back home, and heard how people were talking back here. And,  
15 wow! I thought, 'Oh, my God! That's how I <laughs> sounded to those  
16 people. I get it now.'

In Lines 4-5, Beth reveals evidence of her first negative attitude about her dialect, i.e., that it attracted undesirable attention. While it is at first not apparent to her what is causing this strange behavior of customers, she directly attributes it to her speech in Line 10. It is worth noting that she identifies her dialect here as specifically Southern—or at least she does so through the words of her manager. In the same line, she also describes how his calling her variety Southern was “being nice about it,” but that she in fact sounded “like a hick” (Line 11). In other words, she—not her manager—equates a Southern speech with hicks, a derogatory term for low status rural individuals. She also distances herself from the Californians to whom she sounded strange in Lines 15-16 by referring to them as “those people.” Taken together, this is a largely negative portrayal of her normal language variety, but it also seems to indicate a level of identity with Southern speech that may prove important later on.

#### Excerpt 2. English in Oklahoma

1 Lily: (continued from Excerpt 1) Uh-huh. Yeah, yeah. Ooh, wow. Okay. What do  
2 native Oklahomans sound like?  
3 Beth: <laughs>  
4 Lily: <laughs>  
5 Beth: <imitating accent> Well! <laughs> Oh, very hick-like! Oh, as an example,  
6 um—and I can say that because I am one—um, you, you know, some of  
7 the things that really stand out to me that I was just clueless about before,  
8 you know, people say things like- And, we just butcher the English  
9 language. Like, instead of 'doesn't' and 'wasn't,' we say 'dudn't' and  
10 'wudn't.' And even when I point that out to people who have lived here,  
11 uh<sniffs>, and are native Oklahomans, they <laughs> say, “No, we don't.”  
12 And I, And I catch them saying it, and point, and point it out to them,  
13 and they're like, “Oh, <laughs> my God! We do that!” “Yes, we do.”

Excerpt 2 also presents numerous negative attitudes about Beth's local English variety, this time much more direct and critical and but as conflicted. When asked by the researcher, Lily (a pseudonym), to describe her own speech and that of fellow Oklahomans, Beth uses harsh language, including "very hick-like" (Line 5) and "we just butcher the English language" (Lines 8-9). She also twice mentions that she "points out" examples of what she considers language butchery to other Oklahomans (Lines 10 and 12) in order to draw their attention to it. Perhaps her impetus for her desire to do so comes from the fact that she only became aware of her "hick-like" speech while living away from other speakers of her variety; her fellow Oklahomans may not have had such experiences and may be "just clueless" (Line 7), as she was before moving to California.

Again, though, notice her statement of identification in Line 6: "I can say that because I am one." Here, as with her use of "those people" to refer to Californians in Excerpt 1, she volunteers evidence of her identity as an Oklahoma speaker. While she does not identify Oklahoma directly with the South in this excerpt, she does so elsewhere. It is also worth pointing out that she equates membership in her speech community with the ability to comment on it using harsh language. This point gets indirect reinforcement from her statement in Excerpt 1 that her manager, presumably not Oklahoman, "was being nice" by calling her accent "Southern" when she interpreted his comment as meaning "that I sounded like a hick." Apparently, Beth makes a division between those who can perceive the "Southern accent" of Oklahoma (i.e., those outside the speech community—recall that, in these excerpts, she and other Oklahomans are not aware of it until pointed out) and those who can comment on it (i.e., those within the speech community).

Excerpt 3 is Beth's final negative critique of Southern speech in Oklahoma. It occurs immediately between the two reading passage performances and serves as her request to record the Imitated variety. Here, she specifically references her upcoming imitation as both "Southern" (Line 3) and "Oklahoma" (Line 5). She also specifically identifies this variety as being "like her brother-in-law" (Line 7), whom she mentions sneeringly in the recording. Furthermore, she expresses what appears to be a statement of expected disbelief in the existence of "dialects that are really severe" (Lines 11 and



13). This statement is ironic given the fact that, aside from nine instances of phonological or lexical reduction or substitution, her two performances—one normal and one “really severe”—require sophisticated instrumentation to distinguish from one another!

Excerpt 3. Do-over request

- 1 Beth: I was really hoping that you would ask me to read this, um, <laughs> the  
 2 way that I thought maybe real, um, <laughs> re- real, real, RE:al  
 3 down-home [Southern-  
 4 Lily: [Okay.  
 5 Beth: [Oklahoma <indistinct>  
 6 Lily: [Well, go ahead! Go ahead! <laughs>  
 7 Beth: Like maybe my brother-in-law, [or, you know-  
 8 Lily: [Oh, all right.  
 9 Beth: Okay. I would love to.  
 10 Lily: Okay.  
 11 Beth: Because there are actually people that really do sound like that, in, uh,  
 12 in my friend [“Shelly”]’s family an’—we were talking about that last night,  
 13 actually, about dialects that are really severe.

Perhaps the most important fact to take away from Excerpt 3 is that, while Beth clearly identifies as an Oklahoman, she does not identify as a “real down-home Southern” Oklahoman. Although this fact is not directly stated in the excerpt, it can easily be inferred from her descriptions of the variety, which are even harsher than her critical descriptions in Excerpt 2. Also worth noting is the fact that Beth, while not identifying as a “real down-home Southern Oklahoma” speaker, suggests by her willingness to perform the variety that she is capable of sounding like such a speaker by changing very little of her normal dialect behaviors. Once again, this would seem to indicate that Beth views this variety as a matter of informal register or style, and not a bona fide dialect in its own right.

There are a number of instances of attitudinally neutral identification with the broader context of Oklahoma dialects. Apart from those mentioned above, including her use of “those people” in Excerpt 1 to refer to Californians and “I can say that because I

am one” in Excerpt 2 to justify her criticism of Oklahoma speech, there’s also a brief exchange in which Lily asks Beth, “Do you think you talk like other Oklahomans?” Beth’s answer is simple: “Yes. I think pretty much, yeah.” While not obviously value-laden, these comments reinforce Beth’s membership in the Oklahoma speech community, which in turn indexes her fundamental identity as an Oklahoman.

On the positive end of the attitudinal spectrum, Beth does offer a heartfelt account of her upbringing that, while not specifically indexing the local speech of Watts, does reinforce her attachment to and identity with her small hometown of Watts (Excerpt 4).

Excerpt 4. Recollecting schooldays in Watts

- 1 Beth: Uh-huh. And we didn’t have any art classes in school. And I remember my  
2 teachers in that little school were just precious. Um, they saw something  
3 early on. <clears throat> And so, <sniffs> because we didn’t have any art  
4 classes, they would foster that, uh, in me. And so they would encourage me  
5 to make banners for pep rallies and ask me to do their bulletin boards, and  
6 then when, um, I, uh, got into high school, um, the teachers actually paying  
7 for me to have art lessons in another town.
- 8 Lily: Really?
- 9 Beth: They did.
- 10 Lily: That is just, that’s so wonderful.
- 11 Beth: They did. I could go on and on about that little town and how precious  
12 it was to me and those teachers and how caring they were.

Twice in this excerpt (Lines 2 and 11) Beth uses the word “precious” to describe the people and community of Watts, even saying that she “could go on and on about that little town ... and those teachers and how caring they were” (Lines 11 and 12). Clearly, Beth’s attitude toward Watts and its people is one of fondness, regardless of how “down-home” their accent may have been.

## 5. Conclusion

To summarize, Beth identifies as a Southern Oklahoma speaker, but not as a “real down-home Southern Oklahoma” speaker. The primary reason for this is that she appears to perceive Southern Oklahoma vernacularity as situated along a stylistic scale of more or less formality, the low end of which she terms “severity.” Vowel quality is not especially salient to performance on either end of the scale, but the more “down-home” register is associated with slightly more Southern than Southern Midlands dialect features, including front vowel diphthongization and movement as a result of the Southern Shift, back vowel fronting, and low back contrast. Beth also identifies an out-group/in-group distinction with respect to the Oklahoma speech community, one in which perception of the dialectal variety clearly does *not* equate with the right to evaluate it. In other words, whether or not the local speech is Southern, it is the Oklahoman’s exclusive privilege to comment on it.

## References

- BAILEY, Guy, Tom WIKLE, Jan TILLERY, & Lori SAND (1993) “Some patterns of language diffusion”, *Language Variation and Change*, V, 359-390.
- BAILEY, Guy, Jan TILLERY, & Tom WIKLE (1997) “Methodology of a survey of Oklahoma dialects”, *SECOL Review*, II, 1-30.
- BAKOS, Jon (2013) *A comparison of the speech patterns and dialect attitudes of Oklahoma*, Oklahoma State University doctoral thesis, ProQuest, UMI Dissertations Publishing. (3598873).
- BASSILI, John N. & Rick D. BROWN (2005) “Implicit and explicit attitudes: Research, challenges, and theory”, in D. Albarracin, B. T. Johnson & M. P. Zanna (eds.), *The handbook of attitudes*, Mahwah, New Jersey, USA: Lawrence Erlbaum Associates, 543-567.
- BAUER, Martin W. (2000) “Classical content analysis: A review”, in M. W. Bauer & G. Gaskell (eds.), *Qualitative researching with text, image, and sound: A practical handbook*, London: Sage, 131-151.
- BOERSMA, Paul (2001) “Praat, a system for doing phonetics by computer”, *Glott International*, V, 341-345.

- GARRETT, Peter (2010) *Attitudes to language*, Cambridge: Cambridge University Press.
- LABOV, William, Sharon ASH & Charles BOBERG (2006) *The Atlas of North American English: Phonetics, phonology, and sound change*, Berlin: Mouton/de Gruyter.
- MCBRIDE, Justin T. (2015) *Native American English in Oklahoma: Attitudes and vitality*, Oklahoma State University doctoral thesis, ProQuest, UMI Dissertations Publishing. (3735141).
- PRESTON, Dennis R. (2011) "The power of language regard—Discrimination, classification, comprehension, and production", *Dialectologia*, Special Issue II, 9-33. <<http://www.publicacions.ub.edu/revistes/dialectologiaSP2011/>>
- TILLERY, Jan & Guy BAILEY (1998) Yall in Oklahoma, *American Speech*, LXXIII, 257-278.
- WEIRICH, Phillip (2013) *A study of the pin/pen merger in Oklahoma*, Oklahoma State University master's thesis, ProQuest, UMI Dissertations Publishing. (1547181).
- WELLS, John C. (1982) *Accents of English I: An introduction*, Cambridge: Cambridge University Press.
- WIKLE, Tom & Guy BAILEY (1997) "The spatial diffusion of linguistic features in Oklahoma", *Proceedings of the Oklahoma Academy of Science*, LXXVII, 1-15.
- WOLFRAM, Walt & Natalie SCHILLING-ESTES (2006) *American English*, Malden, MA: Blackwell.

### Appendix: Paragraph Elicitation Text

#### *First Time Through: Real*

Mike was planning to throw a party on Tuesday night and decided to check his list one more time before he went shopping. He already had plenty of stuff to drink and he had enough plates and cups. His brother Dave was going to bring some fish he'd caught and maybe put them on the grill. Mike thought he should get some chips, pretzels, and a few other snacks to start the meal. He looked around to see if he had anything sweet but then remembered that his friend Linda was baking a cake. When he looked in the cupboard, he saw that he was out of coffee. He wrote it down on his list and hoped it was on sale. Then he went to the garage, got in his truck, and went to the Wal-Mart.

#### *Second Time Through: Imitated*

Mike was plannin' to throw a party on Tuesday night, and decided to check his list one more time before he went shoppin'. He already had plenty of stuff to drink and he had enough plates and cups. His brother Dave was gonna bring some fish he'd caught and maybe put 'em on the grill. Mike thought he should get some chips, pretzels, and a few other snacks to start the meal. He looked around to see if he had anything sweet, but then 'membered that his friend Linda was bakin' a cake. When he looked in the cupboard he saw he was out of coffee. He wrote it down on the list and hoped it was on sale. Then he went to the g'rage, got in his truck, and went to Wally World.

Received 3 November 2015.

Accepted 8 February 2016.

## DEL HORDIO A LA CEBADA: ISOGLOSAS LÉXICAS EN EL CASTELLANO NORTEÑO MEDIEVAL<sup>1</sup>

Emiliana RAMOS REMEDIOS

Universidad del País Vasco/Euskal Herriko Unibersitatea (UPV/EHU)\*

emiliana.ramos@ehu.es

### Resumen

En este trabajo se analizan algunas voces del léxico agrícola, extraídas de la documentación del monasterio burgalés de Oña (siglos X al XIII), comparándolas con el léxico de textos coetáneos de otras diatopías peninsulares, con el fin de establecer algunas áreas dialectales dentro del continuo lingüístico norteño que vengán a avalar la complejidad interna del castellano, lengua cuyos usos léxicos coinciden a menudo con los de las variantes occidentales peninsulares, pero también con los de las orientales.

### Palabras clave

dialectología histórica, geografía léxica, léxico agrícola, castellano norteño, siglos X al XIII

### FROM *HORDIO* TO *CEBADA*: LEXICAL ISOGLOSSES IN NORTHERN OLD CASTILIAN

### Abstract

In this paper, some words from the agricultural lexicon — extracted from the documentation of the Monastery of Oña (10th to 13th centuries) — are analyzed and compared with the lexicon of

---

<sup>1</sup> Este trabajo se integra en el proyecto FFI2012-36813 del Ministerio de Economía y Competitividad: *El castellano norteño en la Edad Media. Estudio lingüístico de documentación cántabra y burgalesa (orígenes-siglo XIV)*.

\* Departamento de Filología Hispánica, Románica y Teoría de la Literatura. Facultad de Letras. Universidad del País Vasco/Euskal Herriko Unibersitatea. Paseo de la Universidad, 5. 01006 Vitoria-Gasteiz.

contemporary documents of other peninsular areas with the aim of establishing some dialectal areas within the northern linguistic *continuum*. This tries to confirm the internal complexity of the Old Castilian language, whose lexical uses coincide sometimes with those of the western peninsular variants and others with those of the eastern ones.

### Keywords

historical dialectology, lexical geography, agricultural lexicon, northern Old Castilian, from 10th to 13th centuries

## 1. Introducción

El léxico relacionado con las labores agrícolas y con los cultivos aflora en los textos notariales de la Edad Media a través de la pormenorización de las heredades que se compran, arriendan y donan, pero, sobre todo, a través de la consignación de tributos en especie y de prestaciones de servicios señoriales, hecho que se intensifica a partir del XII. El análisis de este campo, más cercano a lo romance que a lo latino, en la documentación castellana más temprana resulta un instrumento idóneo para establecer algunas áreas dialectales dentro del continuo lingüístico norteño, áreas que permiten corroborar la complejidad interna del castellano, cuyos usos coinciden con mayor frecuencia con los de las variantes occidentales peninsulares, pero también, en ocasiones, con los de las orientales (Fernández-Ordóñez 2011a: 32-60; 2011b: 31-34), además de presentar usos propios del extremo norte central.

En este sentido, y a partir de un corpus de documentos notariales particulares del monasterio de San Salvador de Oña (Burgos),<sup>2</sup> datados entre los siglos X y el XIII –siglo del que tenemos el mayor número de escrituras–, describiremos una serie de usos léxicos en torno a cultivos, aperos, unidades de medida o labores agrícolas, con el fin de

---

<sup>2</sup> Unos doscientos cincuenta documentos particulares, recientemente reeditados en un corpus digital: *Corpus Histórico del Español Norteño [CORHEN]*, [en línea], dirigido por María Jesús Torrens, subcorpus de *CHARTA*, <http://corhen.es/>, cuyos contenidos obedecen a donaciones, permutas, compra-ventas, contratos de arrendamiento y encomienda, concesión y revisión de fueros, pactos o establecimiento de términos. Tendremos en cuenta, además, la edición que de esta documentación realizó Juan del Álamo en 1950. Las palabras y contextos del corpus que nos sirven de ejemplo en este trabajo van marcados con el año entre paréntesis. Seguimos para estos ejemplos la edición crítica del corpus, aunque en todo momento hemos tenido presente tanto la edición paleográfica, como las propias escrituras originales.

delimitar algunas isoglosas léxicas aisladas, insertas en el continuo norteño castellano; para lo cual, será obligado el cotejo con estudios análogos, con fuentes lexicográficas y, especialmente, con colecciones documentales coetáneas del centro peninsular. Con todo, debe considerarse que algunos de estos usos no estarán transmitiendo diatopía,<sup>3</sup> sino usos particulares de los amanuenses y del monasterio (Morala 2008: 212-213) o, podríamos añadir, del propio género discursivo del texto notarial.

Tal vez debido a este último factor, el léxico común de estos textos se sirve de numerosas voces con valor genérico (*árbores, pan, pomiferos*). De hecho, es perceptible en este tipo de corpus que el léxico latino responde a una mayor fijación y, si se quiere, a una menor riqueza, frente a la mayor variación en el léxico romance, a pesar de la cual, no es fácil encontrar términos romances matizados por un adjetivo o sintagma que precise el significado, al modo de *E Majolo Cascajoso* (1177), *el Peral Estranguadizo* (1271). Como en estos ejemplos, es la toponimia la que se encarga a menudo de aportarnos la riqueza y variación de la lengua hablada: denominaciones de un lugar como *El Maello* ‘manzano silvestre’ (1177, 1212) son las que nos informan de voces usadas por los hablantes, pero no especificadas en unos textos que, como se ha señalado, prefieren el uso de genéricos, como *pomiferos*; de voces como *La vit* (1182) o *La Vidiziella* (1177) solo sabemos, de hecho, a través de nombres de lugar. A partir del XII y del XIII será necesario contar también con la antroponimia, a través de las designaciones complementarias del nombre, como apodos o componentes toponímicos, para entresacar elementos léxicos (*Domingo Martínez Peral d’avuelo* (1276), *Gómez Pajar* (1202), *Nieto de Olmos* (1215)).

El orden de la descripción y comparación de las voces nos lo marca una carta de compraventa de 1254, en la que los bienes vendidos se agrupan en las tierras de cereal (*la heredit del pan*), las tierras de viñedo (*la heredit del vino*) y el huerto con sus frutales (*perales e maçanos*).

---

<sup>3</sup> Aunque la mayoría de los documentos hacen referencia a la zona norte de Burgos (las Merindades y la Bureba), la data tópica de algunos textos o sus propias referencias internas marcan además como área de influencia del monasterio el extremo occidental alavés, la costa cántabra y gran parte de la provincia de Palencia.



## 2. La heredad del pan

El cereal más habitual del corpus es el trigo, las voces latina (siempre *triticu*) y romance (*trigo*) son generales en los documentos onienses desde la última década del siglo XII.<sup>4</sup> Frente a ello, el cereal menos frecuente es el *centeno*, cuyas únicas alusiones se concentran en un documento romance de 1212,<sup>5</sup> constituyendo una de las primeras dataciones en el castellano, si atendemos al *DCECH* (s.v.), que ofrece este documento como testimonio:<sup>6</sup> «las Tejas, VI cuarteros de trigo e alia cum ea coniuuncta, III<sup>es</sup> modios de centeno y la de las Salzeras, IIII<sup>o</sup> modios de trigo e de centeno».

La voz romance *cevada* aparece en tres ocasiones; la más temprana es esta de 1168: «et non detis nisi decem panes in enfurcione, et sex dineros in carne, et quatuor quartalejos de vino, et una emina de cevada», junto a dos testimonios del XIII: «Et al mayordomo dar le edes un día en año pan, e vino, e cevada e lo que pudiéredes aver a él e a su compañía» (1254); «que dé a diez bestias una fanega de cevada» (1266); en todos ellos –con más claridad en el último– pudiera mantener el sentido de ‘pienso, alimento para el ganado’, pero también el de ‘cebada’, en tanto que grano más empleado en ese menester;<sup>7</sup> esta acepción queda patente en texto de 1261, donde se

<sup>4</sup> No hace falta señalar que es voz antigua, y que los derivados de *TRITICUM* son propios del castellano y del portugués, además de otras hablas consideradas marginales en la Romania como el sardo, ciertas hablas réticas y alpino lombardas, pues el resto ha seguido los tipos *blat/blé* o *frumento* (*DCECH* s.v.). No figuran derivados de *frumentu* en el corpus, voz muy poco habitual en la documentación peninsular occidental, salvo algunas apariciones en La Rioja (Salas 1986: 297; *Becerro Galicano Digital*), en León (Mínguez 1976: docs. 274 y 308, siglo X) o en Valpuesta (*agros frumentarios*: Ruiz Asencio et al. 2010: doc. 6, año 894). La frecuencia de aparición del trigo se explica porque era el cereal más valorado en la península, seguido de lejos por la cebada. Por el trigo podía llegar a pagarse al triple de precio que por el ordio o la cebada (Piedrafita 2005-2006: 124).

<sup>5</sup> Se trata de una posible copia, por lo que el testimonio habría quizá que retrasarlo a mediados del XIII.

<sup>6</sup> Navarra parece el límite por el este de esta voz; *centeno* es palabra hispánica, conservada solo en castellano, en el portugués *centeio* y en las hablas bereberes, el resto de romances, incluyendo el catalán, tienen derivados del lat. *SECALE* (*DCECH* s.v.). Documentada desde el X en León y Galicia, y posteriormente en Liébana (*LHP* s.v.). El *CORDE* testimonia *centeno* hacia 1134 en el Fuero de Marañón en Navarra, en 1210 en una carta de venta salmantina, en 1221 en el Cartulario de Silos –donde se opone a *ordeo* («medietatem de tritico et medietatem de ordeo vel centeno»)–, en 1248, en Santa María de Trianos en Salamanca, en el Fuero de Cáceres (1234-1275), etc.

<sup>7</sup> Es voz bien atestiguada en la Península desde el XI (*LHP* s.v. *cevada*); fue ganando espacio en el oeste peninsular, relegando a *ordio* hacia el este (Tabernero 1996: 239-240). Según el *DCECH* s.v. *cebo*, esta voz aún significa ‘pienso’ en el *Cid*, aunque acabará designando el cereal más empleado como alimento de los caballos, sustituyendo a los derivados de *HÖRDĒUM*, cuyos herederos solo se conservarán en el alto arag. y nav. *ordio*, en fr. *orge* o en cat. *ordi*; en cat., en arag. y en occ. *civada* pasó a significar ‘avena’, al ser este el cereal utilizado generalmente como pienso, de lo que es testimonio la propia toponimia (Vázquez 1989: 156, 159 y 164-165; Nortes 1979: 229, n.3; *DU CANGE*, s. v. *civata*). Un claro ejemplo de su sentido primitivo

alude ya a su uso como cereal panificable: «III almudes de pan, lo medio trigo e lo medio cevada, e un tocino el año que ovierdes carne».

Frente a ello, el latín *ordeum* y su correspondiente romance *ordio* se imponen cuantitativamente. De hecho, es la forma habitual en este corpus, junto a *tritico/trigo*, restando como anecdóticas las apariciones de *cevada* y *centeno*: «XII<sup>cim</sup> cuarteros, medietatem tritici et medietatem ordeï» (1198); «dos almudes de buen pan e limpio, lo medio trigo e lo medio ordio, a la medida de Tejada» (1279). Así, la distribución en los textos, desde el XII hasta bien entrado el siglo XIII, entre *cevada* –con la evolución semántica desde ‘pienso’ a ‘cebada’–, y *ordio* ‘cebada’ confirma que esta última no responde a una peculiaridad de la mitad oriental peninsular, a pesar de que en el propio *DCECH* (s.v. *horchata*) parece incidirse en un corrimiento desde el este hacia Castilla en el XIII; más bien, la convivencia de *cevada* y *ordio* compartiendo significado en documentación de características similares del XIII hace pensar en una sustitución progresiva de esta voz por aquella en el oeste peninsular, donde fue usual desde el X.<sup>8</sup> Con todo, la aparición de *centeno*, *cevada* y *ordio* en los mismos contextos, no deja de resultar compleja, pues no es posible deducir matices semánticos o hasta qué punto las voces de los documentos hacen referencia a tipos diferentes de cereales, dado el parecido existente entre ellos.<sup>9</sup>

Como es habitual, todos los tipos de grano se incluyen bajo el genérico *pan* ‘cereal’: «esta es la heredit del pan» (1254), «e fielmientre sacada la semient, partiremos lo ál por medio, tan bien del pan, como de las uvas e de lo ál que ý fuere»

---

‘pienso’ lo ofrece este ejemplo de San Millán: «permittimur ipsa cibata que solita erant dan equis nostris in quadragesima» (*Becerro Galicano digital* 1049, doc. 354).

<sup>8</sup> El uso fue también general en León, desde los documentos del X y parece llegar hasta el XV (Sánchez González de Herrero 2005: 43; Fernández Catón 1999 y 2002; Morala 2007: 386). Más ejemplos similares en la documentación toledana, en la que alterna con cebada (Vicente 2007: 1060). Pensado (1960: 86-88) recogía ya un testimonio asturiano, sin *-di-*, del XIII, *huerdo*. En Valpuesta, por citar un corpus muy cercano espacialmente a Oña, aparece solo una vez *ordeo*, en un contexto similar (Perdiguero *et al.* 2012: 61; Ruiz Asencio *et al.* 2010: doc. 87), si bien, llama la atención que se trate de un documento firmado por Olivarius en 1093, notario de origen navarro, como certifica su plural *carners*: «et III concilias de anona, II, scilicet, de tritico et una de ordeo, atque VII carners (*sic*)» (Ramos 2013: 258-259). Abundando en esto, los testimonios del *LHP* (s.v. *ordio*), del XI y del XII, son catalanes, aragoneses, navarros o riojanos; véase también para el uso en Aragón, Fort (1994: 60).

<sup>9</sup> En 1259 un documento de Las Huelgas (*CORHEN*) deja claro ya que se trata de cereales diferentes: «E dexámosvos X ochavillas de trigo, e VII ochavillas de centeno, e X ochavillas de cevada, e XIII ochavillas de comuña, e I ochavillas de avena; recogidos en el genérico pan: XXIIII ochavillas de pan, la meatat de trigo e la meatat de cevada, de buen pan seco e limpio».

(1250), valor metonímico atestiguado en Oña desde fines del xii: «Redditus panis semper in mense septembri tribuatur» (1199) y más frecuente que el etimológico de ‘alimento’, presente en la mención más antigua en el corpus: «et non detis nisi decem panes in enfurcione, et sex dineros in carne» (1168). El pan puede designar también el grano que se reserva para simiente: «xxvi almudes de pan sembradura» (1228). Las entregas de *pan* suelen incluir dos tipos de cereal, el trigo y el hordio: «det unoquoque anno abbati Onie in redditu l<sup>a</sup> modios de pan ad medietatem» (1208); «vendimus illud vobis pro xl<sup>a</sup> cuarteros panis, xxx<sup>a</sup> de trigo et x de ordeo» (1206); «dos almudes de pan, ell unu de trigo e ell otro de ordio» (1238),<sup>10</sup> voz sustituida ya por cebada en: «III almudes de pan, lo medio trigo e lo medio cevada» (1261). En el documento de 1212, el único en el que se recoge centeno y en el que no constan ni cebada ni hordio, las entregas responden a trigo y centeno: «las Tejas, vi cuarteros de trigo e alia cum ea coniuncta, iii<sup>es</sup> modios de centeno», «las Quintanas, xiii modios de trigo e de centeno». En ocasiones se recoge la unidad pluriverbal *pan limpio* o *buen pan limpio* para indicar el grano cribado de impurezas y listo para la molienda: «doze almudes de buen pan limpio al almut de Burgos, lo medio trigo e lo medio ordio» (1274) «dos almudes de buen pan e limpio, lo medio trigo e lo medio ordio, a la medida de Tejada» (1279).

La práctica habitual de la mezcla de cereales para harina –dada la condición de artículo de lujo del pan blanco, de trigo– explica que en algunas ocasiones el grano que se entregaba junto al trigo fuera la *comuña*: «Et damus tibi iiiii<sup>or</sup> tabulatas tritici, et iiiii<sup>or</sup> communie et iiiii<sup>or</sup> solidos» (1190); así, el pan que, como alimento, se entrega a los vendimiadores y labradores en varios documentos consistía en una o dos libras, mitad de harina de trigo y mitad de comuña: «dues libras de las del cuende don Sancho, la una de trigo e la otra de comuña» (1238); «vos dé una libra de pan, la media de trigo e la media de comuña, a cada uno de vós que ý labrare» (1266). La comuña es aún hoy el ‘trigo mezclado con centeno’ (DRAE, s.v.),<sup>11</sup> sin embargo, durante la Edad Media la

<sup>10</sup> Las características similares de un territorio vecino, como es La Rioja, explica que en su documentación los granos más citados sean también estos dos (Salas 1986, 297-298; *Becerro Galicano Digital*).

<sup>11</sup> También ‘aparcería, principalmente de ganados’, acepción propia de Asturias, según el DRAE y DGLA, pero no ajena al euskera: *comuñeru* ‘el que tiene ganado a comuña’, ‘antiguamente, el árbol plantado en terreno ajeno’, *acomuñar* ‘tomar ganado en comuña’ (DCECH s. v. común).

mixtura debió de hacerse con diferentes granos;<sup>12</sup> resulta difícil deducirlo de los ejemplos onienses, si bien, cabe deducir que fuera con trigo y cebada, la mezcla más habitual en el corpus, aunque hemos aludido ya al documento de 1212 en el que se combinan trigo y centeno exclusivamente. La voz parece estar bastante localizada en el centro y oeste peninsular, desde Navarra, La Rioja y Soria hasta Cantabria y Palencia.<sup>13</sup>

Sentido similar, si no idéntico, puede otorgarse a *cibaria/civera*, documentada aquí solo en la última década del siglo XII: «in enfurcione tres quartarios de cibaria, medietatem tritici et medietatem ordeï» (1191); «det vobis unoquoque anno illos modios de cevera quos nobis solet dare» (1192), ejemplo este último menos específico. Más difundida que comuña, es voz atestiguada desde los orígenes en castellano y relacionada etimológicamente con *cebada*, lo que explica los cruces entre ambas voces; su sentido es el de ‘grano que se muele para harina’ (*DCECH*, s. v. *cebo*).<sup>14</sup> Aunque varias fuentes lexicográficas apuntan al trigo como único integrante de la cibera y, en concreto, al trigo que se va echando a la tolva para cebar la rueda,<sup>15</sup> en el corpus oniense, el primer ejemplo que aquí ofrecemos evidencia que responde de nuevo a una mezcla de trigo y cebada, del mismo modo que en un documento de 1199 la referencia anafórica *cibaria pretextata* alude a una entrega de cereal previa, en la que se combinan a partes iguales trigo y cebada (*ordio*).<sup>16</sup> Cabe además que el genérico *pan*, arriba

---

<sup>12</sup> Ese parece ser el sentido de la *comuyna* en el Fuero General de Navarra ‘mezcla de trigo, centeno y otras semillas con que se hacía el pan’ (*DCECH* s. v. *común*).

<sup>13</sup> *LHP* (s. v. *comunia*). Una cata en el *CORDE* nos lleva a estos documentos de Oña, a una carta de venta relativa a Palencia de 1128-1129, al Fuero General de Navarra 1250 – 1300 y a los Fueros de la Novenera (c. 1150 y a. 1253). También en Valpuesta en 1112: «V quarteros trico et quartero de comunia et moio de sal» (Ruiz Asencio *et al.* 2010: doc. 127), en el centro de Burgos (Las Huelgas 1259, *CORHEN*) o en La Rioja (González Bachiller 2002: 188-189). Con este sentido, no lo localizo en documentación leonesa (Fernández Catón 1999, 2002). En el *NTLLE* (s.v. *comuña*) se vincula el sentido ‘trigo mezclado con centeno’ a Castilla la Vieja, pero solo hasta 1791, pues desde esa fecha y hasta las ediciones modernas se pierde la marca dialectal, aunque Terreros y Pando (1786) amplía el significado a ‘pan que se hace de salvado y algo de harina de trigo’, como propio de la Montaña. En cuanto a los testimonios gallegos que aporta J. L. Pensado (1961: 313), se recogen en el siglo XIX y pudieran ser castellanismos. Parece equivaler a la *mestura* de documentos altoaragoneses (Líbano y Villacorta 2013: 103).

<sup>14</sup> Del lat. *CĪBARIA* ‘víveres, alimentos’, plural de *CĪBARIUM* ‘alimento’, ‘harina grosera’; las formas con *i* son las menos habituales en la Edad Media (*DCECH* s. v. *cebo*). Es voz documentada desde Aragón hasta León (*LHP* s.v. *cevera*).

<sup>15</sup> Nebrija (*NTLLE* s.v. *civera*) o *Autor.*, s.v. *cibera*. Todavía hoy en *DRAE* s.v. *cibera*. Nortes (1979: 229-230) propone un sentido general de ‘cereales o grano para la alimentación, tanto humana como de caballerías’, pero también ‘cebada’.

<sup>16</sup> «Et damus vobis in vineis de Villiella v<sup>e</sup> arenzadas et media. Et insuper concedimus [...] centum et xvicim almudes, medietatem tritici et medietatem ordeï, in apoteca de Villiella. Redditus panis semper in mense

mencionado, pueda considerarse como sinónimo de *cibera* y *comuña*, esto es, ‘mezcla de grano destinado a la molienda’, de modo similar al tratamiento que estas voces reciben en áreas próximas como La Rioja (Salas 1986: 297-298), teniendo en cuenta además expresiones como la mencionada *pan limpio*.

Después de la trilla, la paja resultante se utilizaba en tareas como la de la construcción, a raíz del sentido de este conocido testimonio: «Et que nos dedes cad’año doze mostelas de paja restrojaza pora tejar las casa» (1275). Esta *paja restrojaza*, por su atado en mostelas ‘gavillas’, debe responder al sentido general románico de ‘paja larga’ y no a la ‘paja quebrantada’, esto es, la que resulta tras la labor de trilla, una de las acepciones, junto a la de ‘cascabillo’, que poseía PALĒA, y que se fue perdiendo frente al general románico de (DCECH s. v.).<sup>17</sup> En el adjetivo *restrojaza* destaca el sufijo *-aza*,<sup>18</sup> sobre el latín *-ĀCEUS* (var. *-ĀCIUS*), que presenta aquí la función clásica de derivar adjetivos de pertenencia ‘relativo a’ sobre bases sustantivas, y no los sentidos aumentativo o de nombres de golpes, propios del romance (Pharies 2002: 129); la paja restrojaza será, por tanto, la que se encuentra en los campos que acaban de ser segados, valor medieval y clásico de *restrojo*, y de *rastrojo*, evoluciones del antiguo *restojo*, quizá por el cruce con *rastro*, y generales en la Península desde el XIII (DCECH s.v. *rastrojo*)<sup>19</sup>.

Sobre la voz *mostela* ‘haz o gavilla’, voz usada todavía hoy en la Bureba (González Ollé 1964: 166), puede remitirse a los estudios de Pensado (1961) y Pascual (1976).<sup>20</sup> En

septembri tribuatur [...]. Postquam vero vos domna Sancia diem concluderitis extremum, *predictas vineas et cibaria pretexata sine exactione aliqua*». La cursiva es nuestra.

Un testimonio del Becerro de San Millán relativo a una donación al monasterio de Obarenes, cerca de Miranda de Ebro, en 1008, revela un sentido idéntico al de Oña: «Et ii<sup>os</sup> boves, uno asino, vii carros de cevera, inter tritico et ordio» (*Becerro Galicano Digital* doc. 420).

<sup>17</sup> *Pajar* figura como topónimo y como apodo: «A la faza del pajar super carrera» (1208); «de una parte filii Dominici Frutosi, et ex alia Gómez Pajar» (1202).

<sup>18</sup> Similar al de *casas paiaza* ‘pajiza, hecha o cubierta de paja’ del Fuero de Zorita de los Canes y de Sepúlveda (*CORDE*) o al *carro maderiço* ‘¿hecho de madera?’, en Las Huelgas en 1259 (*CORHEN*).

<sup>19</sup> Este era el sentido general de *rastrojo*, más que el moderno de ‘residuos de la mies tras la siega’, como acredita más claramente este documento datado en Saldaña en 1282: «E otrossi, si de alguno de nos leuar el pan del restroio ante que den a uos uuestro quinto e la meatud del diezmo[...] E si el uuestro omne non quissier uenir tomar el diezmo o el quinto, nos, que lo dexemos en el restroio e que nos leuemos el nuestro» (Fernández Flórez 1994: doc. 1285).

<sup>20</sup> Superada la propuesta etimológica sobre \*AMBÖSTĒLLA, fundamentalmente debido a que las formas gallegas *mostea*~*mosteia*, requieren una reducción de -MB-, poco habitual al gallego, solo cabe plantear el étimo MŪSTĒLA ‘comadreja’ para explicar todas las formas hispánicas, entendiendo un cambio semántico mediante un procedimiento de identificación metafórica animal-objeto, relativamente frecuente en el mundo agrícola. Junto a los testimonios modernos gallegos y castellanos (desde *Autor*. s.v.: ‘Nombre que dán en algunas partes a la gavilla de sarmientos’) y a los medievales gallego-portugueses, este ejemplo de Oña es citado recurrentemente para apuntalar la vitalidad medieval castellana de esta voz, junto con uno

el corpus comparte significado con *faces* ‘haces. Atado de mieses, lino, hierba, leña o cosas semejantes’ (*DRAE* s.v. *haz*):<sup>21</sup> «quitámosvos la paja, e el feno e los faces del ordio que nos solíedes dar» (1266), aunque ambas voces pudieron poseer matices semánticos diferentes. Un sentido similar presenta *cerro* ‘manojos de lino’, sobre un lat. *CIRRU* que dio resultados en arag. *ciervo* y en cat., it. y cast. *cerro* (*DCECH* s.v. *cerda*), siendo este de Oña el testimonio más antiguo castellano. Salvando la escasez de testimonios, quizá debiera considerarse voz propia de la mitad oriental peninsular.<sup>22</sup>

En cuanto a las medidas de áridos, muchas utilizadas también para la medición de la superficie de la tierra por ampliación semántica,<sup>23</sup> las más habituales en este corpus son los almudes (*almud/almut*) y tabladas (*tabulatas/tabulas/tabladas*):<sup>24</sup> «centum et xvi<sup>cim</sup> almudes, medietatem tritici et medietatem ordeï» (1199); «onze almudes de pan, lo medio trigo e lo medio ordio al almud que andido en Medina» (1271); «iiii<sup>or</sup> tabulatas tritici, et iii<sup>or</sup> communie» (1190); «nonaginta tabladas ad mensuram Onie, medietatem tritici et medietatem ordeï» (1208).<sup>25</sup> Aunque *tablada* como ‘terreno labrantío

---

latino de Valladolid de 1191 –mostelas de lino– y otro segoviano del XIII –mostelas de hierba– (*DCECH* s.v. *mustela*, *ambuesta*; *DD*, s.v. *mostea/mosteia*; *FEW* s.v. *müstēla*; Pensado 1961: 316-318; Pascual 1976). También en el *Tratado de Agricultura de Ibn Bassal*, anónimo de a. 1300 (*CORDE*). Cabe la posibilidad de que esta identificación metafórica se viera favorecida porque en la designación del animal la voz *MUSTELA* se fue viendo relegada por otras como *comadreja*, dejando herederos solo en catalán y en zonas de Galicia, junto con derivados afectuosos en el norte de León y el oeste de Cantabria, según los datos del ALPI (Pato 2011).

<sup>21</sup> Documentada desde el XIII con esta acepción (*CORDE*).

<sup>22</sup> No la localizo en documentos leoneses medievales y los resultados del *CORDE* corresponden al *Fuero General de Navarra* y al *Fuero de Tudela*, de mediados del XIII. Otro ejemplo en *Tesoro de la medicina (Tesoro de los remedios)*, anónimo de 1431 (*CORDE*). El *DCECH* (s.v. *cerda*) cita ejemplos italianos y catalanes (*cerro*), aragoneses (*ciervo*), castellanos y también leoneses (*cerro*, *cerras*), aunque estos de época reciente.

<sup>23</sup> Resulta habitual que las medidas se empleen tanto para la medición del fruto, en especial del cereal, como para la de las tierras agrícolas, dado que la tierra puede medirse tanto por cantidad de trabajo como por cantidad de grano (Kula 1980: 36-46; Salas 1986: 297): «illam terram que est in la serna de las eras, capientem seminatura de una tablada» (1211); «al molino de Mogada, i tierra que coge ix tabladas sembradura (1254); «un solar que coge iiiii tabladas» (1254).

<sup>24</sup> No figura el cahíz. Como tampoco es habitual en la vecina Rioja alta o castellana, frente a su presencia frecuente en la baja o navarra y en Aragón (Fort 1994: 243; Salas 1986: 297).

*Almud* es frecuente en documentos riojanos y aragoneses (Fort 1994: 243; García Andrevia 2010: 50; González Bachiller 2002: 102-103). Puede ser también medida de líquidos, si bien, con menor frecuencia (Morala 2007: 424).

<sup>25</sup> No se trataba, sin embargo, de medidas universales; por ejemplo, en cuanto a las tabladas, se tomaba como referencia la medida de diferentes lugares o de nobles, principalmente de Oña, pero también a la medida de Frías (1261), a la de Burgos (1274), a la de Tejada (1279), o «dos libras de pan de las del cuende don Sancho» (1266, 1268). El primer intento de regulación de las medidas castellanas no llegará hasta 1261 bajo el reinado de Alfonso X (Martínez Ezquerro 1997: 82-83; Sánchez Martín 2007: 952).

rectangular' es todavía frecuente<sup>26</sup> y ha dejado numerosos herederos en la toponimia peninsular, podemos considerar que, en esta acepción de medida de grano y simiente o de superficie, corresponde a un uso propio de la zona norte de Burgos, pues solo se documenta en esta documentación oniense y en otra relacionada con el entorno de Oña y la Bureba.<sup>27</sup>

### 3. La heredad del vino

Junto al cultivo de cereales, destaca la economía vitivinícola. El vino tenía el mismo rango que el pan en la dieta medieval de todos los niveles de la sociedad, a lo que se une el uso de ambos productos en la liturgia<sup>28</sup> y el hecho de que la viña fuera una planta muy resistente, adaptable a todos los terrenos. Por ello, las menciones de tierras dedicadas al cultivo de la viña son muy frecuentes en el corpus, con idéntico valor en documentos latinos y romances: «iuxta illam vineam de Onia, ut plantes eam et facias in ea vineam ad fundus terre sicut mos est istius patrie» (1205), junto a la voz *parral*, el terreno plantado de cepas alzadas artificialmente, apoyándose en árboles, normalmente en sauces, que son muy frecuentes en esta documentación:<sup>29</sup> «e que la plantedes vós

<sup>26</sup> Así, en Asturias (DGLA, s.v.). El DRAE lo considera voz palentina: 'cada uno de los espacios en que se divide una huerta para su riego'.

<sup>27</sup> El DCECH, s. v. *tabla*, la considera voz occidental, sin embargo consta también en documentación de la mitad oriental –al menos como medida de superficie–, así Enguita (1989: 145) registra en Tarazona en el xiv *tabla* 'faja de tierra labrantía comprendida entre dos árboles'. No obstante, las primeras documentaciones con este valor de medida, de fines del XII y del XIII, son estas de Oña (LHP s.v. *tablada*; DCECH s.v. *tabla*; CORDE). Además de los onienses, el CORDE registra ejemplos con esta acepción en *El Becerro de las Behetrías*, aunque las citas siempre corresponden a localidades de la zona norte de Burgos; también figura *tablada* como medida de sal en un documento de Alfonso X para regular el mercado de sal en las salinas de esta área, documento dado en Miranda de Ebro. No se recoge esta acepción en la documentación leonesa coetánea.

<sup>28</sup> Sobre los factores que generaron la expansión del viñedo a partir del siglo IX en el área del Alto Ebro, especialmente en torno a los monasterios, véase Ruiz de Loizaga (1988: 26-34) y García de Cortázar (2004: 67-69), también Piedrafita (2005-2006: 126-127).

<sup>29</sup> Muchos de los árboles plantados, frutales o no, pero especialmente los sauces, servían como postes de apoyo para el alzado de las parras: «Establecemos que plantedes y cad' año diez salzes e seis ceresos [...] e los salzes que sean todos pora vós pora la labor del parral» (1276); las cepas se elevaban con estacas y se enlazaban con mimbres, por ello es frecuente que los parrales estén situados cerca de sauces, para aprovechar sus ramas (González Bachiller 2004: 81). Esta vinculación entre árboles y viñedos ha sido muy habitual desde época clásica; tal vez por ello, estos árboles no frutales se mencionan en este corpus con bastante frecuencia, frente a la documentación de otras diatopías, como la leonesa, donde las referencias

luego viña assí que sea plantada toda estos tres años primeros que vienen, e desí que fuer bien plantada que alcedes parral» (1276); «e que lo plantedes lo que fuer para parral, parral; e lo que fuer para viña, viña, lo que más convinier en la heredat» (1271). Este sentido moderno –que el *DCECH* (s.v. *parra* I) no atestigua hasta 1400– es claro y frecuente en el corpus en los documentos del XIII; las escasas menciones del XII no son tan evidentes en cuanto a su significado y pueden hacer referencia al antiguo valor de ‘emparrado’ o ‘cercado’ al que alude el *DCECH*: «Al prato de las parras, en una terra sextam partem cum suos arbores» (1144).<sup>30</sup>

El majuelo –con formas diptongadas y sin diptongar–<sup>31</sup> presenta aquí el sentido moderno colectivo, propio de los romances hispánicos y galorromances, de viña nueva que ya da frutos y no el primitivo de cepa nueva –conservado aún en La Rioja (*Aut.*; *DCECH*; *DRAE*, s.v.)–: «III aranzcadas de viña en el majuelo novo» (1177), ejemplo tautológico aportado precisamente por el *DCECH* para dar testimonio de la extensión de esta acepción en los textos del XII;<sup>32</sup> sin embargo, en el mismo documento, el topónimo *E Majolo cascajoso* (1177), esto es, viña plantada en suelo de cascajo,<sup>33</sup> se presenta asimismo como un claro testimonio del sentido colectivo de este término.<sup>34</sup>

Las labores propias del cultivo de la viña, las *labores del vino* se relatan siempre en un orden similar que coincide aproximadamente con el cronológico, a partir de la

---

llegan a través de la toponimia o como apelativos usados para definir los términos de un terreno (Morala 2007: 397).

<sup>30</sup> *Parral* se documenta desde el XI y el XII en Navarra y La Rioja (González Bachiller 2004: 81), pero los testimonios más tempranos tampoco presentan un sentido claro.

<sup>31</sup> La única forma latina es *malleolum*, en un texto de 1177. En otro de 1202 se presenta una latinización de la forma ya romance: *et unum majolum circa lohannem Tomé* (1202).

<sup>32</sup> No obstante, la presencia del adjetivo *nuevo* quizá venga a indicar que el significado se había ampliado ya hasta el actual de ‘viña en general’ (Morala 2007: 385). El mismo sentido en documentos del centro de Burgos, como los de Las Huelgas: «dó una terra que es en Duratón a medias a poner majolo (1188), el primer pedaço de viña es [...] el majuelo del ero del carro» (1228) (*CORHEN*). Sobre su uso en Aragón, Nortes (1979: 200-201) o Fort (1994: 56). En los textos leoneses tiene valor colectivo desde las primeras apariciones en el X, pero siempre se diferencia de *viña* que, en todo caso, funciona como hiperónimo (Fernández Catón, 1999 y 2002); también en Oña *viña* y *majuelo* quedan bien diferenciados: «esto es el heredamiento del vino: un majuelo e una viña a Pezuelos» (1254), si bien *viña* puede actuar como hiperónimo: «Nos damus tibi illud *majuelo* del Val quod est in sulco defiliis de Diego de la Fuent et ex alia parte est terra de filiis de lohanne Domínico. Tu vero das nobis quantumcumque habes in illa *vinea* del Val, scilicet, tres pedacios et sunt in medio de *suprascripta vinea*» (1206), «et *vineam* quamdam scilicet *malleolum* de Covilla» (1177); la cursiva es nuestra.

<sup>33</sup> El idóneo para este tipo de cultivo. Por otro lado, este es un testimonio antiguo de un derivado de *casca* a juzgar por lo afirmado en el *DCECH* s.v.

<sup>34</sup> La vid, sin embargo, no consta como léxico común en el corpus, sino tan solo como topónimo: «illam villam quam vocitant la Vide» (1135), «don Díago de la Vidiziella» (1254).



vendimia<sup>35</sup>: «Et estas faças sobredichas que las lavredes bien cad’año de todas sus lavores, de escavar, e de podar, e de atar, e de cavar e de vinar» (1275). Destacamos la labor de *atar* los sarmientos de las parras que han quedado tras la poda, en invierno, al tratarse de una voz de la mitad occidental peninsular, documentada en el XIII en castellano; con el sentido de ‘atar los sarmientos’, solo figura en estos documentos de Oña (*CORDE*; *DCECH* s.v.).<sup>36</sup> Respecto a la voz *vendimia/vendimiar*,<sup>37</sup> siempre se registra con la yod del diptongo –no registrada hasta Juan Ruiz por el *DCECH*– y nunca con la variante *vendema*, frecuente aún hoy en Aragón, La Rioja, Álava, Navarra o la propia Bureba (*DCECH* s.v. *vino*; Vicente 2007: 1063-1064; González Bachiller 2004: 75, 87-88; González Ollé 1964: 127, 160 y 220).

En cuanto a las medidas de superficie para las viñas, la *arenzada* es la medida más frecuente, equivale a la extensión que puede ser arada por una pareja de mulas en un día (Ruiz De Loizaga 1988: 90).<sup>38</sup> «duas arenzadas de viña» (1144); «in vineis de Villiella v<sup>e</sup> arenzadas et media» (1199); «unam araçadam de vinea in vinea Ferrandi Echavídez, et mediam aranzadam in nava de Ferrero» (1220).<sup>39</sup> Aunque también se recogen medidas de superficie por cantidad de trabajo, en concreto, el trabajo diario realizado por un obrero: «a Pozuelos, I<sup>a</sup> viña de VIII obreros cavadura»,<sup>40</sup> «al Sernal, otra viña de X obreros» (1254). Lógicamente, frente al campo de cereal, que se mide por el trabajo de

<sup>35</sup> Es habitual en otras diatopías, como refleja, por ejemplo, la documentación toledana (Vicente 2007: 1063). Las enumeraba ya San Isidoro (Ruiz de Loizaga 1988: 53).

<sup>36</sup> La inclusión de este verbo sistemáticamente tras podar indica que se trata de ese tipo de atado del sarmiento del año anterior y no del atado en verde, esto es, de los sarmientos o pámpanos del año, tarea que se realiza a finales de primavera o en verano. En algunos documentos, esta labor no figura, pero se respeta el orden de las demás: *escavar, podar, cavar, vinar*.

<sup>37</sup> Solo en una ocasión la correspondiente latina *vindemie* (1229).

<sup>38</sup> En el corpus solo con este sentido; si bien, al igual que otras medidas, como la fanega, se documenta antes en los textos medievales como medida de capacidad –en este caso de vino– que como medida agraria para viñas (*LHP* s.v. *aranzata*). Bien documentada en Aragón (Fort 1994: 242). Su sentido original ‘lo que se puede comprar por un arienzo’ se documenta en textos del IX, por ejemplo en el Becerro de San Millán (García Andreva 2010: 55). El *arienzo* es moneda bien documentada en el corpus: *solidos de arienzos* (993), *solidos argenteos* (1054), *solidos de argento* (1045, 1056), *solidos de argenteos* (1056), *solidos argenti* (1107); en Navarra, sin embargo, el arienzo era la medida agraria propiamente, empleada especialmente para las viñas (Taberner 1996: 421-422). En Valpuesta, como medida de vino (Perdigueru *et al.* 2012: 73).

<sup>39</sup> Con la alternancia vocálica *a/e* muy frecuente en otra documentación (*LHP* s.v. *aranzata*). El ejemplo de *arenzada* lo trae el *DCECH* (s.v. *arienzo*).

<sup>40</sup> El sufijo de formación de sustantivos *-ura* se suele añadir a adjetivos verbales en *-tus* en el latín tardío (Nortes 1979: 177 y 220); además de este caso, véanse en el corpus *sembradura/seminatura, labradura, fiadura, arroturas, pasturas*... El sufijo *-ura* comienza a constatarse en el siglo XIII como variante del sufijo latino *-tura* y de su variante *-sura*; se emplea para derivar sustantivos de nombres de acción y de resultado a partir de verbos (Pharies 2004: 165-166).

los bueyes, el viñedo se mide por el trabajo humano (Kula 1980: 37; Salas 1986: 297).<sup>41</sup>

#### 4. El huerto y los árboles

Junto con la producción cerealística y vitivinícola, la economía agrícola dependía de las pequeñas tierras cercanas a las casas, de los huertos y herrenes, y de los árboles frutales que en ellas se plantaban y que dotaban de gran valor a la propiedad, lo que explica su presencia continua en los textos (Morala 2007: 391-392). Al lado del frecuente huerto: «con so palmiento e sos ortos» (1229 ca.), «un pedazuelo que es en fondón de so villa en cabo de los uertos» (1278),<sup>42</sup> es muy habitual el sinónimo *ortal*,<sup>43</sup> en textos latinos y romances, como apelativo y como topónimo, lo que puede atestiguar su vitalidad en la lengua hablada: «ipsas vineas, et terras, et casas, et ortales» (1011); «l faza que dicitur Ortal del Río» (1177), «Joán del Ortal» (1265).

La herrén (*ferragine, ferranne, ferraine, ferrenas*,<sup>44</sup> *ferrén, ferreñal*) no hace alusión al cultivo en sí, esto es, al cereal que se siega verde para servir como forraje al ganado, sino a un terreno cercado, en el interior del pueblo, dedicado a su cultivo.<sup>45</sup> «la nuestra herrén que está cerca la vuestra casa en Cameno» (1279). Con el mismo sentido metatético de ‘tierra en la que se siembra herrén’ es frecuente en los textos leoneses (Morala 2007: 391) –donde alternan las formas con o sin sufijo *-al* (Fernández Catón 1999 y 2002)– o en los aragoneses, donde suele presentar el sufijo (Fort 1994: 51; Nortes, 1979: 189); los textos riojanos también parecen decantarse por la forma sin sufijar (González Bachiller 2002: 296-297), como en Valpuesta o en estos textos de Oña,

---

<sup>41</sup> *Obradas de viña* es medida similar (*DCECH s.v. obrar*; Taberner 1996: 141).

<sup>42</sup> Diminutivo: *aliam terram ad Orteolum* (1202).

<sup>43</sup> Sigue el modelo de *casa/casal*, como en textos leoneses, riojanos o aragoneses (Morala 2007: 387 y 389; Fort 1994: 36; García Andruva 2010: 94; Nortes 1979: 197), además de en los castellanos (Perdiguer *et al.* 2012: 57), aunque Enguita (1989: 157) lo daba como aragonesismo.

<sup>44</sup> Convive con *ferrén* en un interesante documento de 1208, un borrador de una pesquisa, en romance, con mínimos elementos latinos.

<sup>45</sup> En principio, la herrén o herrenal se diferenciaba claramente del huerto, al convivir ambos en los mismos documentos; con el tiempo, su uso para el cultivo de hortalizas, además del de forraje, hizo que se identificaran con el huerto, a pesar de su mayor tamaño (Arribas 2012: 618-619).

donde solo una vez presenta el sufijo *-al (ferreñal)*,<sup>46</sup> lo que inclina a pensar que estas formas sin sufijo podrían responder a una tendencia central castellana.

Por otro lado, en la voz romance ha triunfado el vocalismo en *-e-*, como en el centro de Burgos o León, frente a Valpuesta, con formas en *-a- (ferrane)*, similares a las que hoy son propias de Cantabria, La Rioja o Álava (Sáiz Barrio 1991: 140; Goicoechea 1961: 99; López de Guereñu 1958: 268, 275; Ramos 1999: 71-72) y que coinciden en vocalismo con las orientales peninsulares.<sup>47</sup> La aparición de dos ejemplos con *-a- (ferranne)* en el documento de 822 –copia del XIII– es fácilmente explicable por tratarse de un texto relativo al monasterio de Tobillas, la zona occidental de Álava (Tobillas), de vocalismo en *-a-*. De los diferentes resultados de *ferragine* en la documentación más antigua y de la repartición primitiva de los mismos ya se ocupó Menéndez Pidal (1999 [1926]): § 14<sub>1</sub>, 112<sub>6</sub>) y García de Diego (1950: 109) describió la importancia de *ferragine* como delimitadora de una zona dialectal que abarca un espacio vasco de *larrain* y *arrain*, con la forma *rain* en Álava y *herrain* que baja por la Sierra de la Demanda hasta Soria.

Estos espacios cercanos al pueblo –huertos, herrenes y también viñas– se dedicaban frecuentemente a la producción de frutales:<sup>48</sup> «et una ferraine cum suos solares, cum suos pumiferos, cum exitu et introitu» (1067), «e la tierra que avemos en los huertos que dizen del Peral Estranguadizo» (1277),<sup>49</sup> «que lo plantedes lo que fuer para parral, parral; e lo que fuer para viña, viña, lo que más convinier en la heradat. E que plantedes ý fructeros» (1271). Destaca en este punto el doblete léxico

<sup>46</sup> «Et vineas, ferreñales, et hortales et fontes», en una donación de la iglesia de San Salvador de Barruelo a Oña, firmada en 1065 por el propio abad, don Íñigo: «Ego Éñeco scripsi et de manu mea hunc signum feci», por lo que puede inferirse aquí un uso personal del redactor, de origen aragonés, uso que casaría con nuestra propuesta de que la presencia del sufijo tiende a desaparecer en el centro peninsular. Llama la atención además que los elementos de la serie léxica en la que se halla *ferreñales* constituyan algunos de los escasos rasgos romances en un texto con un latín muy cuidado, debido seguramente también a la intervención del abad.

<sup>47</sup> El pueblo de Herrán, hoy burgalés, se encuentra a 15 km al suroeste de Valpuesta, en el actual límite entre Álava y Burgos. Oña es, por tanto, el límite por el norte y el este de la isoglosa de *-e-*.

<sup>48</sup> La función básica de los huertos –terrenos pequeños, cercados y aledaños a la casa–, posteriormente identificados con las herrenes, era la de la producción frutal, más que la de hortalizas y legumbres para el consumo de la casa (Arribas 2012: 620-621). En los textos riojanos es también habitual que la herrén constituya una ‘propiedad rústica con frutales’ (García Andrevia 2010: 75).

<sup>49</sup> *Estranguadizo*, tal vez sobre *estragnar* ‘asolar, destruir, echar a perder’, con *n* epentética, más un sufijo *-izo*, que forma derivados de particios con valor pasivo o activo, como ya sucedía en latín (Pharies 2002: 375-376), en este caso, ‘que se estraga o seca fácilmente’. En La Rioja *estragado* es ‘sediento’ (Pastor 2008: 299); en Asturias, *estragnar* ‘atragantarse’ (DGLA). Tal vez haya un cruce con otras voces, como las relacionadas con *estrangular*: *La Estranguada* es un alto de montaña, un paso estrecho, en Cantabria.

*pomiferos*~*pomares*, que convive en algunos textos con la voz *mazanare(s)*, y que debió de tener –como ya planteó Aebischer en 1948 para toda la Romania– un valor genérico de ‘(árbol) frutal’ y no el más restringido de ‘manzano’.<sup>50</sup>

No obstante, no es fácil determinar qué valor específico añadido pudo tener el romance *pomar*, que figura desde muy pronto en las series léxicas de pertenencia de Oña, de modo paralelo a *mazantar*: «in terras, in binias, in pumares, in ortales, in casas» (993), «ipsas vineas, et terras, et casas, et ortales, et mazanares, cum pratis» (1011) y que ha permanecido en la toponimia de la zona (Medina de Pomar). Los testimonios de Oña son poco clarificadores al respecto. En el resto de la documentación hispánica se observa, por un lado, un sentido genérico de *pumar* y *pomifero* en los documentos leoneses anteriores al XII, en los que tampoco *poma* posee aún el sentido específico de ‘manzana’ (Morala 2007: 392-393; Ruiz de Loizaga 1988: 94), de manera semejante a lo que se aprecia en los textos aragoneses anteriores al 1157 estudiados por Nortes (1979: 206), si bien, en Aragón pronto parece significar también un árbol frutal concreto, diferente al manzano (Fort 1994: 55). Por otro lado, la documentación hispana refleja asimismo un sentido relativamente temprano y específico de *pumar/pomar* como ‘frutal, similar o igual al manzano’, que se ve en el Fuero de Soria de c. 1196 («Qvi cortare arbol que leuare ffructo, como serual o pumar o mjespolar, por el tronco dos mr.», *CORDE*), en algunos textos de Santa María del Puerto de Santoña, especialmente en una pesquisa de 1210 en la que *pumar* se equipara a otros nombres de árboles frutales (*CORDE*) o en Valpuesta, con mayor nitidez: «cum quatuor mazanos et uno perare et duodecim pumares» (950) y «cum mea terra cum x mazanos et II pomares» (1054?) (Perdiguero *et al.* 2012: 62-63; Ruiz Asencio *et al.* 2010: 206 y 245), en fechas además bastante tempranas, todo lo cual nos habla de una posible difusión desde el centro peninsular de este sentido específico de *pomar*.

En realidad, *pumar/pomar* desaparece del corpus oniense en el siglo XI y solo permanece en el topónimo Medina de Pomar, mientras que *(im)pomiferos* deja de aparecer a mediados del XII, para ser sustituido por los derivados de la base *fruct-* (*fructiferos*, *fructuosis*), cuyo equivalente romance del XIII será *fruteros*. Esta base *fruct-*

---

<sup>50</sup> Lat. med. *pomaris* ‘lugar plantado de arboles frutales’ (*DU CANGE S. V.*).

/frut- es bastante más productiva en el corpus desde el siglo XI, anticipando la forma romance que triunfará como genérico para designar ‘fruto’ o ‘frutal’. A partir del XI, de modo paralelo, se incrementa el uso de las voces con sentido específico ‘manzano, manzanar’ sobre la base *mazan-/maçan-*.<sup>51</sup>

Los frutales más habituales en el corpus oniense son manzanos, perales, cerezos y nogales y, de entre los árboles sin fruto, los sauces: «e que plantedes cad’ año seis pies de ceresos» (1275),<sup>52</sup> «otra de III tabladadas con sos árboles, maçanos e perales e otros frutereros» (1254), «e plantar nos edes y cad’ año x árboles de maçanos, e ciruelos e salzes» (1254)... Cerezo, ciruelo y manzano, frente a lo que sucede en otras diatopías, como la leonesa, donde es frecuente que los nombres de frutales se construyan incrementando el nombre del fruto con el sufijo *-al, -ar*: *la cerezal* (Morala 2007: 394); en los documentos de Oña el sufijo se aplica solo a *nogar/nogal, peral, moral* y *figar*.

Llama la atención el frecuente *nogal*, que presenta la terminación en *-ar*, en textos latinos o híbridos, y en *-al* en los romances, continuando con la vinculación de estos sufijos desde el latín, donde *-ĀRIS* funcionaba como alomorfo disimilatorio de *-ĀLIS*, como ha señalado Pharies (2002: 91):<sup>53</sup> «et unam vineam trans palatium, et terciam partem de ipsis nogaribus» (1202), «et el nogar quod est in orto de medio et cum illa parte del nogar que fuit Dominico Petri» (1202), «con el terradgo de dos nogales» (1279). No

<sup>51</sup> En Valpuesta, a pesar de que ya hemos mencionado que *pomar/pumar* parece presentar un valor específico, además del genérico ‘frutal’, se sigue un proceso paralelo, pues a mediados del XI deja de emplazarse *pomares* (las últimas menciones de *pomiferos* son de fines del X) y se imponen los derivados de *manzana* y de *fruto* (Ruiz Asencio *et al.* 2010). Frente a ello, los documentos leoneses mantienen el uso de *pomar/pumar* por más tiempo, al menos hasta fines del XII –en Otero de las Dueñas (Fernández Flórez y Herrero de la Fuente 1999), monasterio de Sahagún (Fernández Catón 1999) o en la catedral de León (Fernández Catón 2002). Para alimentar la confusión, todavía hoy el *DRAE* ofrece las acepciones ‘fruta de árbol. Manzana. Casta de manzana pequeña y chata, de color verdoso y de buen gusto’; en Álava y otros lugares responde al ‘fruto del serbal’ (López de Guereñu 1958: 229).

<sup>52</sup> Pie de árbol ‘tronco’ o mejor ‘planta’. Siempre el plural *pies* y no *pieses* en los tres testimonios de nuestros corpus, de fines del XIII (1275 y 1277), a pesar de que para Corominas *pieses* es la única forma para ese siglo, apoyándose en las lecturas del metro de Berceo o del Mio Cid (*DCECH s.v. pie*); González Bachiller (2004: 82) aporta testimonios riojanos similares y anota este mismo uso del plural analógico desde el siglo XIII.

<sup>53</sup> Pharies (2002: 91) da cuenta de los numerosos dobles *-al/-ar* en la historia del castellano. Como se aprecia en los textos de Oña (*maçanar* frente a *nogar/nogal*), el sufijo latino *-ĀLIS*, aplicable a adjetivos, da lugar a *-al/-ar* en castellano, formante de sustantivos con sentido de ‘planta’ –tras la elipsis del sustantivo que originariamente acompañaba al adjetivo– o de ‘lugar donde abundan plantas’ –por metonimia–. A partir de adjetivos latinos como *NUCĀLIS* ‘similar a un fruto seco’, de *NUX, NUCIS* ‘fruto seco’ se entiende que *-al* dé lugar en castellano a sustantivos referentes tanto a la planta misma (*peral, moral*) como a abundanciales (*juncal*) (Pharies 2002: 57-59; 91-93; y 2004: 160-161); en Oña lo general es que estas formas se refieran al árbol ‘nogal’, pero en ocasiones parece tener valor abundancial: «al nogal, I tablada» (1254).

parece este un reflejo de variación diatópica, sino un uso propio de la escritura del monasterio o, más bien, de algunos escribanos, pues se concentra en cuatro documentos, datados en un corto espacio de tiempo: 1177, 1200 y 1202; puede deberse a un caso de distribución de uso entre lenguas o registros, considerando *-al* romance y *-ar* latino<sup>54</sup>.

*Nogal/nogar* alterna en el corpus con la variante *nozeda/nozedo*,<sup>55</sup> documentada a través de topónimos, que posee sentido colectivo ‘lugar plantado de nogales’ –al menos en el caso en femenino–, o específico y colectivo en el caso del masculino:<sup>56</sup> «faza que dicitur de los Nozados» (1177). En 1177 conviven en el mismo documento *nogar* y *nozedo*: «I terram que dicitur del Nogar iuxta viam publicam», «I faza que dicitur de los Nozados».

Por último, uno de los testimonios más destacable en este campo es el proporcionado por la toponimia: «et I terram al Maello iuxta vineam de Gonzalvo Díaz» (1177), en referencia al ‘manzano silvestre’, con la forma aún conservada en zonas de Cantabria, Palencia, Burgos, Soria, Álava y occidente de Segovia, Ávila y Salamanca (*DCECH s.v. maguillo*; *ALCYL* mapa 404), lo que confiere a esta voz un carácter central en la Península.<sup>57</sup>

---

<sup>54</sup> Varios casos de *nogar* en un documento de remembranza de 1300 de Santo Toribio de Liébana (*CORDE*), monasterio, por otra parte, donado a Oña en la primera mitad del XIII.

<sup>55</sup> En Aragón alternan *noguera* y *nocedo* como ‘nogal’ (Enguita 1989: 156 y 170). En el área de Oña se prefieren *nogal* y *nogala* para el árbol y *nogalera* para el colectivo (González Ollé 1964: 168-169). Hoy el término general en el centro peninsular es *nogal*, mientras que *noguera* queda relegado hacia el este –general en Soria– o hacia el oeste –*noguera* o *nogueira* en el extremo occidental de León, Zamora y Salamanca– (*ALCYL* mapa 399; Fernández-Ordóñez 2011a: 33).

<sup>56</sup> Del valor abundancial de *Nocedo* da cuenta Nortes (1979: 206) para el aragonés. Sobre su uso como nombre del árbol véanse los ejemplos riojanos de González Bachiller (2002: 364-365) y los de Valpuesta: «cum quinque mazanos et tres nocetos et uno perare» (950) (Perdiguero *et al.* 2012: 64; Ruiz Asencio *et al.* 2010: 209).

<sup>57</sup> También en La Rioja *maguilla* o *maila* ‘manzana silvestre’, voz general del valle del Ebro, y como topónimo, Maguillo (González Bachiller 2004: 79; Pastor 2008: 311). De hecho uno de los testimonios más antiguos (*maiello*) se recoge en San Martín de Albelda, 921 – 1108 (*CORDE*), junto a *maello* en el Fuero de Sepúlveda, del siglo XIII (*CORDE*).

Un caso similar en el corpus es el del topónimo *Pruno*: «et sinera ad Pruno» (822), que puede aludir a una variedad de ciruelo, de fruto negro, propia del norte de España (Morala 2007: 395), caso único frente a la voz común *ciruelo*, general en el corpus.

## 5. Conclusiones

La fracción del léxico que nos ofrecen los textos notariales ha de ser una mínima parte del empleado en la oralidad y, desde luego, distará mucho de lo que tuvo que ser la variación de la lengua hablada. Ciertamente es además que no es fácil ver en estos documentos áreas bien definidas, especialmente en determinados campos léxicos; no obstante, este de la vida agrícola, más cercano si cabe a lo romance que a lo latino, nos permite aportar algunos apuntes para la historia de algunas voces y ayudar a su delimitación en el tiempo, a su distribución y a su difusión en el espacio del continuo norteño peninsular.

En estos documentos de Oña figuran voces generales en el centro peninsular, desde Aragón hasta León como *majuelo*, con el sentido colectivo de ‘viña nueva que ya produce frutos’ u *ortal*, que aparece junto a *huerto*, con un sufijo similar al que vemos en *ferreñal*, por ejemplo, si bien, en el caso de esta última voz, Oña prefiere la forma sin sufijar, frente a León o a Aragón, donde las formas en *-al* son habituales. Si es verdad que Oña prefiere el vocalismo en *-e*: *ferrén*, más cercano a la mitad occidental peninsular, al vocalismo en *-a*: *ferrán*, de Valpueda, Cantabria o Álava, inclinado hacia la mitad oriental; salvo en el caso esperable del documento de 822 de Tobillas, donde se recoge vocalismo en *-a*, propio del oeste alavés.

Otras son palabras que solo se documentan en un área central más restringida: es el caso de *comuña*, que con el sentido de ‘mezcla de cereales, trigo y cebada, normalmente’, se constata en Navarra, Oña y Las Huelgas o Palencia –aunque Pensado (1961: 313) menciona ejemplos gallegos– frente al más general *civera*. O *tablada*, que en su acepción de ‘medida para granos’ y, por ende, ‘medida de extensión de la tierra’ solo se testifica en Oña, como, tal vez, *atar*, en relación a los sarmientos de las vides.

Frente a ello, otras voces que tradicionalmente se han considerado orientales, como *ordio* ‘cebada’, se recoge abundantemente en Oña, desde el XII, del mismo modo que en textos leoneses, por ejemplo, lo que permite proponer un corrimiento de *ordio* de este a oeste en la península, debiéramos proponer una progresiva sustitución de *ordio* por *cebada* en el oeste a partir del XIII: los textos de Oña confirman además la ampliación semántica de *cebada*, desde ‘pienso’ hasta ‘cereal panificable’.

De algunas formas recogidas en Oña, existen tan pocos testimonios en otros textos, de hecho, a veces la voz oniense ostenta el título de primera datación o de único testimonio, que es imposible determinar si debemos considerarlas voces orientales (*cerro* [de lino], *obreros de cavadura*) u occidentales (*mostela*). En voces como *centeno*, queda claro que el límite por el este debió de estar en torno a Navarra, pues se trata de una voz usada solamente en la mitad occidental, frente a los derivados orientales de SECALE.

En ocasiones, el corpus nos ofrece sistemáticamente una voz, hoy general, por ejemplo *vendimia*, cuando en la actualidad, la zona de Oña prefiere *vendema*, como es habitual hoy en Aragón, La Rioja, Álava, Navarra, lo que parece indicar que, o bien la variante sin la *yod* ha ido extendiéndose de este a oeste, o bien –y esto será lo más probable– los textos no nos dan cuenta de toda la variación de la lengua hablada.

La onomástica, tanto los nombres de persona como los de lugar, ayudan –o complican– a menudo la tarea. Como ejemplo, frente a la dificultad de dilucidar un sentido específico para *pomar/pumar*, partiendo del léxico común de los documentos, su toponimia nos deja un único *Maello* ‘manzano silvestre’, constancia de nuevo de una voz central de la Península, que se abre en abanico hacia el sur. Del mismo modo, los textos señalan como voz común *nogal* en los contextos romances y *nogar* en los latinos –con una alternancia de los sufijos que denota un uso propio de la escritura y no variación diatópica–; sin embargo, la toponimia de los documentos nos da cuenta de su convivencia con *Nozeda/Nozeda*, advirtiéndonos de que existió una alternancia antigua entre varias formas, que se han perdido en el área de Oña, donde hoy se prefiere *nogal* y *nogalera*, como el resto del centro peninsular.

Con todo y a pesar de estos pocos datos, la tendencia que estos primeros textos romances, y muy especialmente estos de Oña, heredaron de la escritura latina a servirse de términos genéricos –salvo excepciones como las de las pesquisas– y los todavía escasos estudios exhaustivos a partir de corpus documentales castellanos de época temprana nos impiden captar un espectro mayor de la riqueza y variación del romance hablado para así trazar isoglosas más precisas y fiables en la historia del léxico castellano norteño de la alta y plena Edad Media.



## Referencias bibliográficas

- AEBISCHER, Paul (1948) «Las denominaciones de la “manzana”, del “manzano” y del “manzandar” en las lenguas romances, según los documentos latinos de la Edad Media», *Estudios de toponimia y lexicografía románica*, CSIC: Barcelona, 97-129.
- ÁLAMO, Juan del (1950) *Colección diplomática de San Salvador de Oña (822-1284)*, I (822-1214) y II (1215-1283), Madrid: CSIC.
- ALVAR, Manuel (1999) *Atlas lingüístico de Castilla y León*, Junta de Castilla y León: Valladolid [ALCYL].
- ARRIBAS, M<sup>a</sup>. del Carmen Sonsoles (2012) *Las Merindades de Burgos. Un análisis jurisdiccional y socioeconómico desde la Antigüedad hasta la Edad Media*, Tesis doctoral inédita, Universidad de Burgos [<http://hdl.handle.net/10259/201>].
- Becerro Galicano Digital* [[www.ehu.es/galicano/id505](http://www.ehu.es/galicano/id505)] (Consulta 2/11/2015).
- COROMINAS, Joan y José Antonio PASCUAL (2012) *Diccionario crítico etimológico castellano e hispánico*, Gredos: Madrid, CD-Rom [DCECH].
- Corpus Histórico del Español Norteño* [en línea], dirigido por María Jesús Torrens [<http://corhen.es/>] [Consulta 2/11/2015] [CORHEN].
- Dicionario de dicionarios. Corpus lexicográfico da lingua galega* (2006-2013) [<http://sli.uvigo.es/ddd/index.html>], Instituto da Lingua Galega. Universidade de Santiago de Compostela [Ddb]
- DU CANGE, et al. (1883-1887) *Glossarium mediae et infimae latinitatis*, éd. augm., L. Favre: Niort [<http://ducange.enc.sorbonne.fr>] [DU CANGE].
- ENGUITA, José María (1989) «Un vocabulario agrícola turiasonense de 1382», *Turiaso*, 8, 139-178.
- FERNÁNDEZ CATÓN, José María (1999) *Index verborum de la documentación medieval leonesa. I. Monasterio de Sahagún (857-1300)*, Centro de Estudios e Investigación «San Isidoro»: León.
- FERNÁNDEZ CATÓN, José María (2002) *Index verborum de la documentación medieval leonesa. III. Archivo Catedral de León (775-1300)*, Centro de Estudios e Investigación «San Isidoro»: León.
- FERNÁNDEZ FLÓREZ, José Antonio (1994) *Colección diplomática del monasterio de Sahagún (857-1300. V (1200-1300)*, Centro de Estudios e Investigación «San Isidoro»: León.
- FERNÁNDEZ FLÓREZ, José Antonio y Marta HERRERO DE LA FUENTE (1999) *Colección documental del monasterio de Santa María de Otero de las Dueñas, I (854-1108)*, Centro de Estudios e Investigación «San Isidoro»: León.

- FERNÁNDEZ-ORDÓÑEZ, Inés (2011a) *La lengua de Castilla y la formación del español*, Real Academia de la Lengua: Madrid.
- FERNÁNDEZ-ORDÓÑEZ, Inés (2011b) «El norte peninsular y su papel en la historia de la lengua española», en Sara Gómez Seibane y Carsten Siner (eds.), *Estudios sobre tiempo y espacio en el español norteño*, Cilengua: San Millán de la Cogolla, 23-68.
- FORT, Rosa (1994) *Léxico romance en documentos medievales aragoneses (siglos XI y XII)*, Diputación General de Aragón: Zaragoza.
- GARCÍA ANDREVA, Fernando (2010) «Estudio léxico del Becerro Galicano emilianense», *Aemilianense*, II, 37-126.
- GARCÍA ARIAS, Xose Lluís (2015) *Diccionario general de la lengua asturiana*, Editorial Prensa asturiana: Oviedo [<http://mas.lne.es/diccionario/>] [DGLA].
- GARCÍA DE CORTÁZAR, José Ángel (2004) *Sociedad y organización del espacio en la España medieval*, Universidad de Granada: Granada.
- GARCÍA DE DIEGO, Vicente (1950) «El castellano como complejo dialectal y sus dialectos internos», *Revista de Filología Española*, XXXIV, 107-124.
- GOICOECHEA, Cesáreo (1961) *Vocabulario riojano*, anejo VI del *Boletín de la Real Academia Española*: Madrid.
- GONZÁLEZ BACHILLER, Fabián (2002) *El léxico de las colecciones calceatenses de los siglos XII y XIII*, Universidad de La Rioja: Logroño.
- GONZÁLEZ BACHILLER, Fabián (2004) «Voces relativas a la flora y la agricultura en documentos riojanos medievales», *Berceo*, 146, 67-87.
- GONZÁLEZ OLLÉ, Fernando (1964) *El habla de la Bureba. Introducción al castellano actual de Burgos*, Anejo LXXVIII de la *Revista de Filología Española*: Madrid.
- KULA, Witold (1980) *Las medidas y los hombres*, Siglo XXI: Madrid.
- LAPESA, Rafael (2003) *Léxico hispánico primitivo (siglos VIII al XII)*, Espasa: Madrid [LHP].
- LÍBANO, M. Ángeles y Consuelo VILLACORTA (2013) *Paisaje rural y explotación agropecuaria: léxico de los recursos naturales y de la vida cotidiana en el aragonés, navarro y romance vasco (siglos XIII-XVI)*, Universidad de Zaragoza: Zaragoza.
- LÓPEZ DE GUEREÑU, Gerardo (1958) «Voces alavesas», *Euskera*, III, 163-361.
- MARTÍNEZ EZQUERRO, Aurora (1997) «Notas de metrología calagurritana del siglo XIII», *Kalakorikos*, 2, 69-84.
- MENÉNDEZ PIDAL, Ramón (1999 [1926]) *Orígenes del español. Estado lingüístico de la Península Ibérica hasta el siglo XI*, Espasa-Calpe: Madrid.

- MÍNGUEZ, José María (1976) *Colección Diplomática del Monasterio de Sahagún (Siglos IX y X)*, Centro de Estudios e Investigación «San Isidoro»/ Archivo Histórico Diocesano/Caja de Ahorros y Monte de Piedad de León: León.
- MORALA, José Ramón (2007) «Léxico de la vida cotidiana. El trabajo en el campo», en J. M. Fernández Catón, *Monarquía y sociedad en el Reino de León. De Alfonso III a Alfonso VII*, Centro de Estudios e Investigación «San Isidoro»: León, 377-444.
- MORALA, José Ramón (2008) «Isoglosas en el Romance Primitivo», en B. Díez (ed.) *El Primitivo Romance Hispánico*, Instituto castellano y leonés de la Lengua: León, 209-221.
- NORTES, Oliverio (1979) «Estudio del léxico latino medieval en diplomas aragoneses anteriores a 1157 (Términos referentes a la composición de la sociedad y a la vida rural)», *Archivo de Filología Aragonesa*, XXIV-XXV, 15-255.
- PASCUAL, José Antonio (1976) «Sobre la etimología del castellano *mostela* ‘gavilla’», *Archivum*, XXVI, 193-196.
- PASTOR, José María (2008) «Caracteres lingüísticos de La Rioja (y II): claves léxicas», *Kalakorikos*, 13, 297-391.
- PATO, Enrique (2011) «Sobre geografía lingüística española: distribución y áreas léxicas de la MUSTELA», *Dialectología*, 6, 45-53.  
<<http://www.publicacions.ub.edu/revistes/dialectologia6/>>.
- PENSADO, José Luis (1960) «Estudios de lexicografía asturiana», *Archivum*, X, 53-120.
- PENSADO, José Luis (1961) «Notas lexicográficas a la *Colección diplomática de San Salvador de Oña*», *Vox Romanica*, 20, 307-326.
- PERDIGUERO, Hermógenes, M<sup>a</sup> Nieves SÁNCHEZ y Ángeles LÍBANO (2012) *Aspectos léxicos en los Beceros de Valpuesta*, Universidade da Coruña: A Coruña.
- PHARIES, David (2002) *Diccionario etimológico de los sufijos españoles*, Gredos: Madrid.
- PHARIES, David (2004) «Tipología de los orígenes de los sufijos españoles», *Revista de Filología Española*, LXXXIV, 1, 153-167.
- PIEDRAFITA, Elena (2005-2006) «La alimentación en Aragón en el siglo XIII: el modelo clerical y el nobiliario», *Revista Zurita*, 80-81, 99-132.  
<[ifc.dpz.es/recursos/publicaciones/26/29/piedrafita.pdf](http://ifc.dpz.es/recursos/publicaciones/26/29/piedrafita.pdf)>.
- RAMOS, Emiliana (1999) *Vitoria y su jurisdicción a fines del XV. El Apeo de 1481-1486 (según un traslado de 1526)*, vol. II, Ayuntamiento de Vitoria-Gasteiz.
- RAMOS, Emiliana (2013) «Consideraciones al proceso de fijación del romance en la escritura (siglos X al XIII)», *Cuadernos del Instituto Historia de la Lengua*, 8, 241-267.

- REAL ACADEMIA DE LA LENGUA (1726-1739) *Diccionario de Autoridades* [<http://web.frl.es/DA.html>] [Autor].
- REAL ACADEMIA DE LA LENGUA *Nuevo Tesoro Lexicográfico de la Lengua española* [<http://ntlle.rae.es/ntlle/SrvltGUISalirNtlle>] (Consulta 26/10/2015) [NTLLE].
- REAL ACADEMIA ESPAÑOLA (2001) *Diccionario de la lengua española*, 22ª edición [[http://buscon.rae.es/drae/?type=3&val=TABLA%20RASA&val\\_aux=&origen=REDRAE](http://buscon.rae.es/drae/?type=3&val=TABLA%20RASA&val_aux=&origen=REDRAE)] [DRAE].
- REAL ACADEMIA ESPAÑOLA *Corpus diacrónico del español* [<http://www.rae.es>] (Consulta 30/10/2015) [CORDE].
- RUIZ ASENCIO, José Manuel, Irene RUIZ ALBI y Mauricio HERRERO JIMÉNEZ (2010) *Los Becerros Gótico y Galicano de Valpuesta*, volumen I: Estudio, edición e índices, Real Academia Española/Instituto Castellano y Leonés de la Lengua: Madrid.
- RUIZ DE LOÍZAGA, Saturnino (1988) *La viña en el occidente de Álava en la Alta Edad Media (850-1150)*. Cuenca Omecillo-Ebro, Burgos.
- SÁIZ BARRIO, Miguel Ángel (1991) *Léxico cántabro*, Ediciones Tantin: Santander.
- SALAS, Diana (1986) «Terminología de la producción y organización de los espacios de cereal en La Rioja en los siglos X a XIII», *Segundo Coloquio sobre Historia de La Rioja (Logroño, 2-4 de octubre de 1985)*, tomo I, Logroño, 289-302.  
<[dialnet.unirioja.es/descarga/articulo/555524.pdf](http://dialnet.unirioja.es/descarga/articulo/555524.pdf)>.
- SÁNCHEZ GONZÁLEZ DE HERRERO, M. Nieves (2005) «El léxico de los documentos alfonsíes», *Estudis romànics*, 27, 41-72.
- SÁNCHEZ MARTÍN, Javier (2007) «Aproximación al léxico de los pesos y medidas de capacidad en la época renacentista», *Interlingüística*, 17, 951-960.
- TABERNERO, Cristina (1996) *La configuración del vocabulario en el romance navarro. Estudio sobre documentos reales de los siglos XIII y XIV*, Eunsa: Pamplona.
- VÁZQUEZ, Jesús (1989) «Pervivencia de apelativos de la flora y de la fauna en la toponimia de Sobremonte (Huesca)», *Archivo de Filología Aragonesa*, XLII-XLIII, 149-172.
- VICENTE, Irene (2007) «El léxico de la agricultura en documentos medievales de la catedral de Toledo (siglos XI al XIII)», *Interlingüística*, 17, 1058-1967.
- WARTBURG, Walter von (1940-1969) *Französisches Etymologisches Wörterbuch*, B. G. Teubner: Leipzig [<https://apps.atilf.fr/lecteurFEW/index.php/page/view>] [FEW].

Received 11 November 2015.

Accepted 8 February 2015.

## **AT THE DIALECTAL CROSSROADS: THE SPANISH OF ALBUQUERQUE, NEW MEXICO**

Mark WALTERMIRE

New Mexico State University\*

markw@nmsu.edu

### **Abstract**

Bills & Vigil (2008) have established two dialects of New Mexican Spanish – Traditional Spanish and Border Spanish. In general, archaisms and Anglicisms predominate in the north of the state (Traditional Spanish) whereas Mexicanisms predominate in the south (Border Spanish). The recent incorporation of Mexicanisms in the Spanish of Albuquerque has placed it at a dialectal crossroads where a Traditional variety is being supplanted by Mexican Spanish, thereby making it more similar to Border Spanish. This study addresses the extent to which the Spanish of Albuquerque has absorbed Mexicanisms and explores how they were introduced into the community. The use of Mexicanisms is compared to that of archaisms and Anglicisms using an original 20-hour corpus of spoken Albuquerque Spanish. Contrary to the claims of Bills & Vigil (2008), the Spanish of Albuquerque is much more similar to Border Spanish than Traditional Spanish and should be reclassified to reflect this difference.

### **Keywords**

dialect change, social variation, lexical variation, Spanish in the U.S., Mexican immigration

### **EN LA ENCRUCIJADA DIALECTAL: EL ESPAÑOL DE ALBURQUERQUE, NUEVO MÉJICO**

### **Resumen**

Bills & Vigil (2008) han establecido la existencia de dos dialectos del español de Nuevo Méjico – el español tradicional y el español de frontera. En general, los arcaísmos y los anglicismos predominan en el norte del estado (español tradicional) mientras que los mexicanismos predominan en el sur (español fronterizo). La reciente incorporación de mexicanismos en el español of Albuquerque lo ha situado en una encrucijada dialectal donde una variedad tradicional está siendo suplantada por el español mejicano,

---

\* New Mexico State University, Department of Languages and Linguistics, MSC 3L, Las Cruces, NM, USA 88003.

haciéndola más similar a la frontera española. Este estudio se refiere a la medida en que el español de Albuquerque ha absorbido Mexicanismos y explora cómo fueron introducidos en la comunidad. El uso de mexicanismos se compara con el de los arcaísmos y anglicismos utilizando un corpus original de 20 horas de habla española de Albuquerque. Contrariamente a lo que indican Bills & Vigil (2008), el español de Albuquerque es mucho más similar al español de la frontera que el español tradicional y debería ser reclasificado para reflejar esta diferencia.

**Palabras clave**

cambio dialectal, variación social, variación léxica, español en EE.UU., emigración mejicana

**1. The Spanish of New Mexico**

Spanish expeditions leaving from New Spain (now Mexico) scoured the southwest of what is now the United States throughout the early part of the sixteenth century. The first permanent settlement in this region was not until 1598, however, near present-day Española, New Mexico, which is located about 25 miles northwest of Santa Fe, as shown in Figure 1.

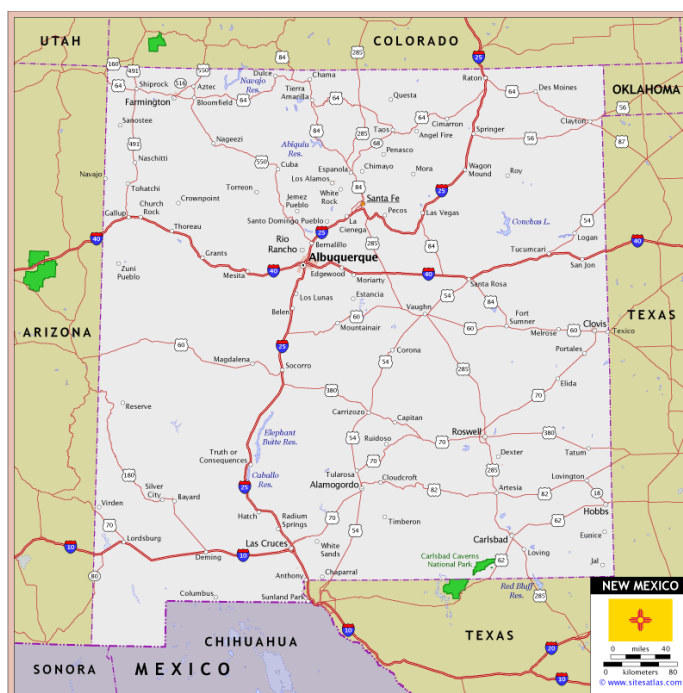


Figure 1. Map of New Mexican Cities and Town

Permanent settlements by Spanish explorers in other areas of what is now the southwest of the United States did not occur until much later (Texas - 1659, Arizona - 1700, California - 1769, Colorado - 1851) (Silva-Corvalán 2001: 298-299). It should be noted that “such ‘Spaniards’ had already become thoroughly ‘Americanized’ – often more specifically ‘Mexicanized’, manifesting a hybrid culture and language enriched by contact with the Native Americans of the Caribbean and Mexico” (Bills & Vigil 1999: 43). Three major historical factors have shaped New Mexican Spanish. These are: 1) relative isolation from the rest of the Spanish-speaking world for several centuries; 2) the gradual settlement of English speakers in New Mexico beginning in the mid-1800s; and 3) massive waves of immigration of Spanish speakers, primarily from Mexico, into the United States in the second half of the twentieth century.

The linguistic consequences of these developments include the common use of archaisms, Anglicisms, and Mexicanisms. Since New Mexico was the first area to be settled by speakers using now-archaic traits of Spanish and it was so distant from other Spanish-speaking areas, many archaic forms have survived among New Mexican Spanish speakers, particularly among older speakers. Many now-archaic traits of Spanish have been preserved due to New Mexico’s relative isolation from other dialects of Spanish for such an extensive period. This isolation resulted from the fact that New Mexico was located “1,500 miles from Mexico City and initially 750 miles from the closest Hispanophone town in Mexico” (Bills & Vigil 1999: 43). Due to such limited contact with other varieties of Spanish for such a long period of time, archaic features have been largely preserved in northern New Mexican Spanish. From the extensive data collected in the New Mexico-Colorado Spanish Survey (hereafter NMCOSS), Bills & Vigil (1999) have revealed the distribution of archaic lexical forms such as *túnico/traje* ‘woman’s dress’ (standard *vestido*), *mesmo* ‘same’ (standard *mismo*), *asina* ‘thus’ (standard *así*) in various parts of New Mexico. Archaic morphological forms (such as *vide* ‘I saw’ (standard *vi*), *vido* ‘he/she saw’ (standard *vio*), *semos* ‘we are’ (standard *somos*), *haiga* ‘there is/are, pres. subj.’ (standard *haya*)) and phonological forms (“such as the retention of the /x/ fricative corresponding to orthographic *h*” (Bills & Vigil 1999: 50) and retention of labiodental [v] corresponding to orthographic *v* (Torres Cacoullós & Ferreira 2000) are also prevalent in this dialect. These archaic features have been passed

on to younger generations, though younger speakers use them to a lesser extent (Bills & Vigil 2008) since many of them do not speak Spanish and largely possess receptive skills in this language (Bills 1997; Bills & Vigil 1999; Bills & Vigil 2008; Hernández-Chávez, Bills & Hudson 1996; Hudson, Hernández-Chávez & Bills 1995). The relative isolation of Spanish in present-day New Mexico persisted with very little external contact until the mid-1800s. With the Treaty of Guadalupe Hidalgo in 1848, “a great swath of northern Mexico was ceded to the United States [...] One surprising consequence of this event was an accelerated opening up of new areas to Hispano settlers” (Bills & Vigil 1999: 48).

Settlement by non-Hispanos was relatively slow. According to Williams (1986: 126), “Anglo-Americans constituted less than 9 percent of the territorial population [of New Mexico] in 1880.” By the 1940s, however, “only half of the population of New Mexico was Hispanic” (Simmons 1977: 163). The settlement of non-Hispanos has had a very strong impact on the Spanish of the area, but its impact has been less linguistic than it has been socio-cultural. Contact with English has contributed many loanwords to the lexical repertory of New Mexican Spanish, some established and some spontaneous, and has also resulted in frequent code-switching among bilinguals in all areas of the state. The historical context just described has directly affected the use of English-origin loanwords in the Spanish of New Mexico. Common loanwords include *troca* (standard *camioneta* ‘truck’), *lonche* (standard *almuerzo* ‘lunch’), and *reporte* (standard *informe* ‘report’). Moreover, English has displaced the use of Spanish to a great degree and has gradually become the dominant language of the state. According to Bills & Vigil (1999: 55), “the children of Hispanic heritage are abandoning the Spanish language entirely and growing up as English monolinguals. More and more, those who speak Spanish in the southwest United States [read New Mexico] tend to be first-generation Mexican immigrants and their children”. Though this situation may be reversing somewhat due to bilingual education and heritage language programs throughout the state, the effect of the dominance of English is undeniable.

The speech of Mexican immigrants has also affected New Mexican Spanish, primarily at the lexical and morphological levels. The use of Mexicanisms, which Company Company (2010: xvi), in her introduction to the *Diccionario de Mexicanismos*, defines as “a collection of words, phrases, expressions, and meanings characteristic of



the speech of Mexico, which distinguish the Mexican variety from Peninsular Spanish” is common throughout the state (translation mine). These are not simply forms that scarcely exist in other varieties (such as *chamaco/a* ‘child’ or *zacate* ‘grass’) but also forms whose meaning is different in other varieties (such as *padre* ‘cool’ and *carnal* ‘brother’). The presence of Mexican speakers is now felt by many in the community due to several decades of increased immigration.<sup>1</sup> Several participants for the current study indicated this during interviews, as shown in the following comments:

- (1) E: *Y, y, y pues, uh, uh, el, el, ¿el español se habla mucho aquí en Barelás o ...?*  
‘And, and, and, so, uh, uh, is Spanish spoken a lot here in Barelás or ...?’  
C: *No. No como más antes. No más los que hablan español, uh. Es muy, uh, es más los paisas, los mexicanos, los que hablan.*  
‘No. Not like before. The only ones who speak Spanish, uh. It’s really, uh, it’s more the Mexicans, uh, who speak.’  
E: *Y los viejos también, ¿no?*  
‘And old people too, right?’  
C: *Y los chicanos, nosotros, que son de Barelás, hablan como, uh, el slang.*  
‘And the Chicanos, us, that are from Barelás, speak like, uh, slang.’ (15/10: 59-11: 21)
- (2) *Hay familias aquí que vivieron hasta el ... ¿Quién sabe? Pero, uh, ahora estos días son muchos mexicanos ahora.*  
‘There are families that lived here until the ... Who knows? But, uh, now these days there are a lot of Mexicans now.’ (15/24: 55-25: 10)

The use of Mexican Spanish was stigmatized by many New Mexicans in the past, who claimed a Spanish heritage and rejected connections to Mexico. This discrimination is recounted by several participants for the current study who witnessed it first-hand, as exemplified in the following comments:

---

<sup>1</sup> According to U.S. Census data obtained from American Fact Finder ([factfinder.census.gov](http://factfinder.census.gov)), Mexican immigration to Albuquerque tripled from 1980 to 2010. In the 1980s, a total 4,882 immigrants of Mexican origin migrated to Albuquerque. By the 1990s, this number nearly doubled to 9,067. A decade later, that number grew even more, with 12,516 Mexican immigrants settling in Albuquerque during the 2000s.

- (3) *Había restaurantes aquí y lugares que tenían unas, uh, cartas en las puertas, uh, cuando entras que decían “No se permiten perros ni mexicanos”.*

‘There were restaurants here and places that had some, uh, signs on the doors, uh, when you enter that said “No dogs or Mexicans allowed”.’ (05/7: 20-7: 31)

This type of discrimination, of course, is a consequence of the myth that northern New Mexicans are descended from Spain. The same participant, however, rejects this idea in the following comment.

- (4) *Yo soy, aquí como dicen en inglés, Spanish American. No puedes ser Spanish American, o Spanish. Lo que ese quiere decir para mí es que nacistes en España y te hicistes ciudadano aquí. Y luego dicen, “No no no, yo soy puro.” O como me dicen también, “Nosotros no tenemos sangre mezclada. Semos puros españoles.” ¿Cómo puro?*

‘I’m, like they say here in English, Spanish American. You can’t be Spanish American, or Spanish. What that means to me is that you were born in Spain and you became a citizen here. And then they say, “No no no, I’m pure.” Or like they also tell me, “We don’t have mixed blood. We’re pure Spanish.” How pure? (05/2: 42-3: 07)

Somewhat ironically, given the lack of prestige of Mexican Spanish in the past, the presence of recent immigrants has resulted in the exposure of more standard linguistic forms of Spanish to New Mexicans. As a result, less standard lexical and morphological forms are gradually being replaced by more standard Mexican forms, particularly by younger generations (Bills & Vigil 2008). The standardization of New Mexican Spanish is also resulting from younger New Mexicans’ exposure to the standard form of Spanish taught in universities. Since Spanish is a heritage language for many younger New Mexicans, they enroll in university classes in order to learn the grammar of a language they have heard used in the home throughout their lives. Bills & Vigil give the examples of *vestido* ‘woman’s dress’, which is replacing *túnico* and *traje*; *falda* ‘skirt’, which is replacing *naguas*; standard *blusa* ‘blouse’, which is replacing *cuerpo*; and *lata* ‘tin can’, which is replacing *bote* and *jarro* (Bills & Vigil 1999: 56). More standard morphological forms are also being learned at the expense of non-standard forms. For example, non-

standard *vide* 'I saw' is being replaced by standard *vi* in large part due to more exposure to formal Spanish in schools (Bills & Vigil 2008: 227).

## 2. Barelas, Albuquerque

The Albuquerque neighborhood of Barelas, which is located just south of downtown, was one of the first permanent settlements in New Mexico. As such, the Spanish of this community is the longest-standing spoken variety in Albuquerque.<sup>2</sup> Spanish colonizers heading north along the Camino Real used the landing at Barelas (originally spelled "Varelas") in the early 1500s as a crossover point on the Río Grande, which it still borders to the east. Due to its importance, it was used during expeditions led by Coronado in 1540 and Oñate in 1598. Barelas was formally established in the late 1600s as a ranching settlement by Don Pedro Varela. Currently, the neighborhood covers about ten square blocks south of downtown Albuquerque, running north and south from Coal Avenue to Bridge Boulevard and east and west from 2<sup>nd</sup> Street to 12<sup>th</sup> Street. This location is shown in Figure 2.

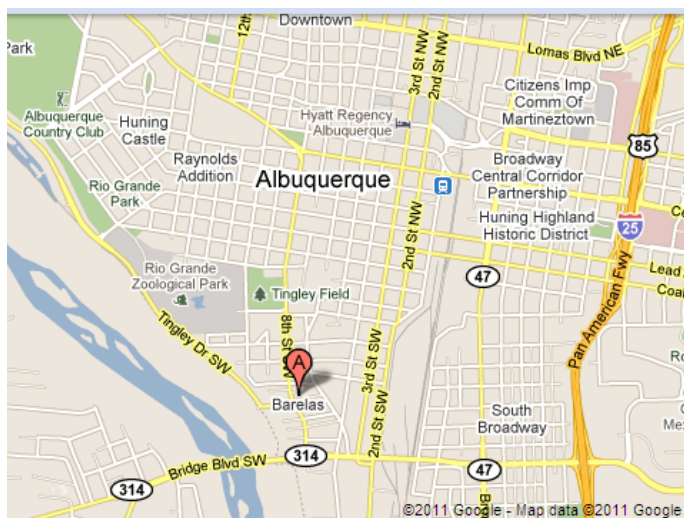


Figure 2. Map of Barelas [Source: Google Maps]

<sup>2</sup> The Spanish of Barelas will represent that of Albuquerque for the current study. Other, more recently established areas of Albuquerque have been populated by Mexican immigrants rather than native *buquerqueños*. It is important that a non-immigrant community represent Albuquerque since the Spanish spoken in the city is not Mexican Spanish. The variety of Spanish analyzed in the current study is distinctly New Mexican.

Due to its long history, the variety of Spanish spoken in Barelás is actually one of the oldest in the entire state of New Mexico and represents features that have long been associated with traditional New Mexican Spanish, which is linguistically quite distinct from its neighbor to the south (Bills & Vigil 2008).<sup>3</sup> According to Bills & Vigil (2008: 5), speakers of this dialect “represent early settlement prior to the twentieth century and today reside primarily in the upper Río Grande drainage area of central and northern New Mexico”. The variety of Spanish spoken in Barelás represents one of the southernmost dialects of traditional New Mexican Spanish. Historically, Barelás was not considered an immigrant community. Unlike other areas of Albuquerque, the Spanish of Barelás has not yet assimilated many of the characteristics of border Mexican Spanish, which has greatly influenced the Spanish spoken in the southern part of the state. The incorporation of traits from border Mexican Spanish in Barelás began more recently as a natural result of growing numbers of immigrants moving into the community. In this sense, Barelás represents a dialectal crossroads where a traditional, somewhat archaic variety is being supplanted by border Mexican Spanish.

Traditional New Mexican Spanish is most distinct from Border New Mexican Spanish in terms of its lexical inventory, which has been documented extensively in the past century (Bills & Vigil 2008; Cobos 1983; Espinosa 1909; Hills 1906; Ornstein 1975). In general terms, archaisms and Anglicisms are more prominently used in the northern two-thirds of the state (Bills & Vigil 2008: 51-64, 173) whereas modern Mexican terms are more commonly used in the southern third (Bills & Vigil 2008: 39). According to Bills & Vigil (2008: 39), “certain features characteristic of the popular speech of modern Mexico prevail mostly in the southern part of New Mexico [...] and in other areas where immigrants have been most likely to find employment. The spatial constraints on this most recent Mexican influence are the basis for our distinguishing the two major dialects we label Border Spanish and Traditional Spanish”.

The purpose of the current study is determine the extent to which the Spanish of Barelás has incorporated Mexicanisms; if it has incorporated a significant number of them, to determine whether the Spanish of Albuquerque is in need of reclassification as

---

<sup>3</sup> The use of the term traditional here, though somewhat contentious even to the authors themselves, is used only insofar as a distinction with southern (or border) New Mexican Spanish. It by no means implies that this variety is more traditional, in the cultural sense, than any other variety.

a dialect. To accomplish this, words of the three lexical types just discussed (archaisms, Anglicisms, and Mexicanisms) included in the dialect maps in Bills & Vigil (2008) was investigated using a 20-hour original corpus of spoken *Bareleño* Spanish recorded by the author in 2010.<sup>4</sup> Since the number of these terms is hardly vast enough to make any definitive claims as to whether the Spanish of Albuquerque more closely approximates Traditional Spanish or Border Spanish, the entire Barelas corpus will be examined, with each of the terms encountered in the corpus being classified as archaisms, Anglicisms, and Mexicanisms. Via the quantitative analysis of patterns of usage of these lexical types, the possibility of reclassifying the Spanish of Albuquerque will be explored.

### 3. Methodology

The two major dialects of New Mexican Spanish (the Traditional Spanish of the north and the Border Spanish of the south) as proposed by Bills & Vigil (2008) are based on the geographical distribution of the three lexical types just described (archaisms, Anglicisms and Mexicanisms) as reported by 357 Spanish speakers. As part of the NMCOS, various terms fitting these types were solicited from participants “by means of pictures and real objects, which were then grouped into semantic categories (colors, birds, domesticated animals, foods, clothing, etc.)” (Bills & Vigil 2008: 27). The distribution of these terms are displayed in a series of dialect maps in this work. A corpus of spoken Albuquerque Spanish from the neighborhood of Barelas will be utilized to identify terms belonging to these three lexical types given that the number of forms presented in Bills & Vigil (2008) is limited and does not include the range of forms present in the Barelas corpus. The advantage of using interview data is that these forms were never solicited from participants and, as such, reflect their actual usage within the

---

<sup>4</sup> None of the Mexicanisms from Bills & Vigil (2008) were encountered in the Barelas corpus. This is largely due to the infrequency with which these words are actually used in day-to-day discourse (e.g. *partidura/partido/parte* ‘part (in hair)’; *chuparrosa/colibrí* ‘hummingbird’; etc.). Of the archaisms and Anglicisms given in Bills & Vigil (2008), three of each type were encountered in the Barelas corpus. These are: *túnico* ‘dress’, *albercoque* ‘apricot’, *vide* ‘I saw’, *troca* ‘pickup truck’, *suera* ‘sweater’, and *queque* ‘cake’.

community. The use of interview data will ensure an accurate determination of the extent to which each lexical type is used in the Spanish of Albuquerque.

### *3.1 Fieldwork*

Sociolinguistic interviews with 15 fluid bilinguals of Spanish and English were conducted by the author in the summer of 2010 in the neighborhood of Barelás. Though this number seems small, there are only 21 speakers from Albuquerque represented in the NMCOS, which is far and wide the most comprehensive documentation of New Mexican Spanish. Furthermore, Barelás is a small, tightknit community of fewer than 3,500 residents and, as such, fewer speakers are needed to achieve an accurate representation of the neighborhood. All participants have lived in Barelás for at least 20 years. In an attempt to approximate the vernacular, which is representative of unmonitored, informal speech (Labov 1972, 1984), various protocols were followed. Initial interviews were set up with close friends and family members of a main contact in the community who is a close friend of the author. More importantly, however, is that almost every interview was recorded only after having made initial contact and engaged in social interactions with participants. Only a couple of interviews were conducted “on the spot” (i.e. immediately after meeting a participant). Participants chose the locations and times of the interviews. They were encouraged to speak about topics that were of particular interest to them and were not discouraged from switching between Spanish and English, though an attempt was made to steer speech toward Spanish whenever possible. No pre-written questions were used in any of the interviews, allowing for spontaneous interaction and fluid discourse. Each speaker was recorded in Spanish for approximately one hour, yielding a total of more than 20 hours of recorded spontaneous conversation. All interviews were recorded as WAV files using a lapel-style omnidirectional microphone as this is one of the least unobtrusive means by which to capture spontaneous language data. After the interviews were conducted, they were transcribed and organized by interview session in order to facilitate the process of token extraction.

### 3.2 Selection of participants

In order to achieve as representative a sample as possible, the selection of participants was based on 2000 census data (factfinder.census.gov), which includes the neighborhood of Barelás as well as adjoining areas spanning downtown Albuquerque to the north and the South Broadway section of the city to the south.<sup>5</sup> According to census data, there are 22,349 residents living in this part of the city. Unfortunately, detailed demographic data for the neighborhood of Barelás is not available through the 2000 census. For more detailed sociodemographic data, the website [www.city-data.com](http://www.city-data.com) was used. More specific population estimates show that the population of Barelás is very small, with approximately 3,500 residents.<sup>6</sup>

Language use is shaped largely by the social and professional changes one experiences throughout one's lifespan. Younger adult speakers, for example, are actively engaged in raising children and pursuing professional pursuits while older speakers may be retired or have children who have already moved away from home. For this reason, participants were selected based on generational groupings that correspond roughly with these changes (Group 1 = 25-50; Group 2 =  $\geq$  51). Unfortunately, census data only show the number of residents who are under the age of five, between the ages of 18 and 65, and older than 65. Given the lack of more specific figures with regards to the 18 to 65 year-old group, a roughly equal number of participants was chosen for each generational grouping.

With respect to the population's distribution by sex, census data are much more specific. According to these data, there are 10,712 women and 11,637 men living in this area of the city. Population estimates by sex for Barelás ([www.city-data.com](http://www.city-data.com)) parallel these figures (1,703 females and 1,796 males). With this in mind, a roughly equal number of females and males were interviewed. To ensure the accurate representation of general social characteristics among the set of participants, an approximately equal

---

<sup>5</sup> Census data for 2010 according to zip codes were not available at the time of authorship.

<sup>6</sup> For detailed information see <http://www.city-data.com/neighborhood/South-Valley-Albuquerque-NM.html>. The specific nature of these data are due to the use of this website for real estate purposes.

number of male and female participants was chosen for each generational grouping (G1/F = 4; G1/M = 4; G2/F = 3; G2/M = 4).

Barelas is a low-income neighborhood which has faced economic hardships since the decline of the railway boom that led to relative prosperity in the area at the turn of the twentieth century. The average single-family income in 2009, according to data from [www.city-data.com](http://www.city-data.com), was only \$24,020, just over half of the average single-family income for greater Albuquerque in the same year (\$44,594). The occupations held by residents are also a reflection of its middle to middle-lower class history. Occupations requiring advanced degrees and formal training are scarce among the neighborhood's residents while occupations in service, construction, maintenance, production, transportation, and sales are extremely common. According to [www.city-data.com](http://www.city-data.com), 72.2% of all employed women in the community hold these jobs while a total of 87.1% of men have similar jobs.<sup>7</sup> Participants were selected based on this occupational distribution in the community. Of a total of 15 participants, only two participants (one male and one female) have jobs that require advanced degrees and formal training.

Though participants were not selected based on language preference or their parents' nationalities, these two variables are important in the current study in that they may be conditioning the use of Mexicanisms in Barelas. There is a tacit assumption in the community that Spanish is being lost at the expense of English, especially among younger speakers, and that the only way to maintain Spanish is by speaking it in the home.<sup>8</sup> According to several participants,

(4) *Se les va olvidando en eso que van aprendiendo el inglés. Se les va olvidando, como siguen hablando tanto el inglés en la escuela, con los amigos.*

'They're forgetting it as they're learning English. They're forgetting it, as they keep speaking so much English in school, with their friends.' (14/14: 56-15: 04)

---

<sup>7</sup> It should be noted that these percentages would be almost the same if women in the community were employed in transportation, which represents 10% of the male workforce.

<sup>8</sup> This is true for the 15 participants of the current study. Of the youngest generation, not a single participant prefers Spanish. Of the eight younger speakers, three prefer English and five claim to have no preference at all. The older generation exhibits a wider range of variation regarding language preference, with only two speakers who prefer English and two who claim no preference at all; and, not surprisingly three speakers actually prefer Spanish.



- (5) *Se me hace que cuando tienes padres mexicanos, nunca se pierde porque estás hable, hable con ellos.*

'It seems to me that when you have Mexican parents, you never lose it because you're always talking with them.' (13/19: 05-19: 11)

It should come as no surprise, then, that all of the participants in the current study who prefer English were raised in a household in which both parents were born in the United States. Participants with at least one Mexican-born parent either prefer Spanish or have no preference. This does not mean that a participant raised by two U.S.-born parents, however, automatically prefers English. This is also an assumption made by many and one which is particularly untrue of older speakers, many of whom grew up speaking Spanish, as expressed in the following comment:

- (6) *Parece que las gentes ahora, que vienen de México, están aquí y me dicen, "Pues, ¿dónde aprendiste tú todo el español tan bonito?" Les digo, "¡pues, yo nunca lo perdí!"*

'It seems that the people now, that are coming from Mexico, they're here and they say to me, "So, where did you learn all of that really beautiful Spanish?" I tell them, "Well, I never lost it." (04/22: 43-22: 54)

Of the participants for the current study, about an even number were raised in a household of two U.S.-born parents (N=8) as were raised in a household with at least one Mexican-born parent (N=7). There is a greater preference for English in the community overall with five participants who prefer this language, three participants who prefer Spanish, and seven participants who have no strong preference for either language.

#### **4. Results**

All examples of the three lexical types for the current study – archaisms, Anglicisms and Mexicanisms – were extracted from transcriptions of the Barelás

interviews. These types appear in the appendix. The raw numbers for each of the three lexical types in the Spanish of Albuquerque are shown in Figure 3.

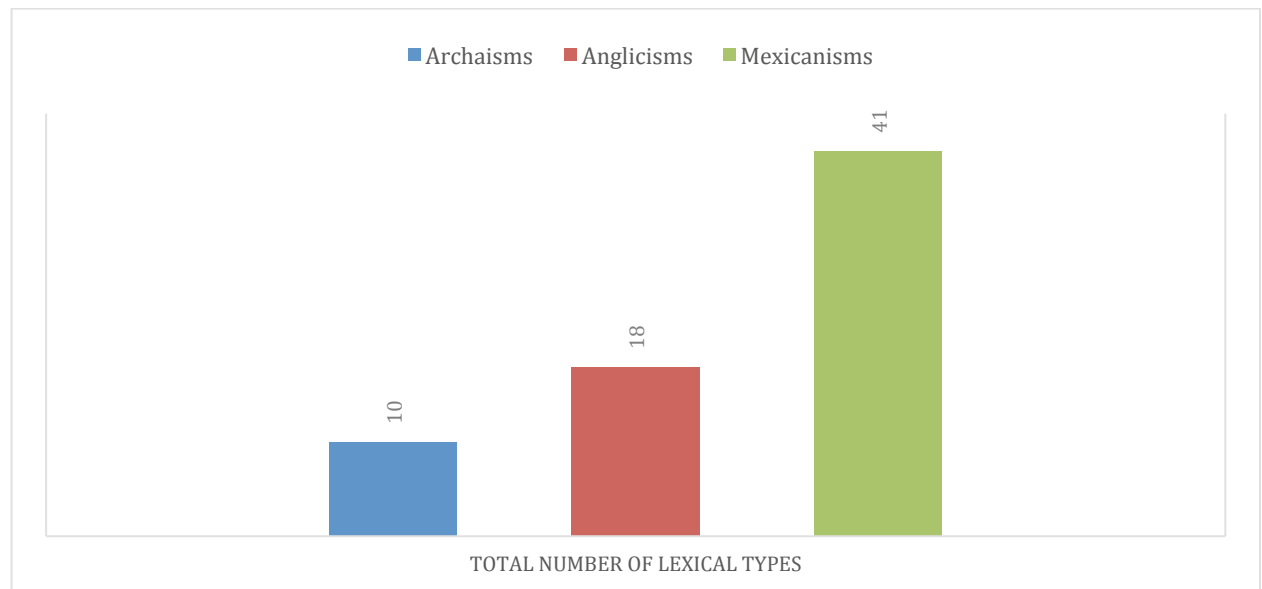


Figure 3. Raw Numbers of Each Lexical Type in Albuquerque Spanish

The number of Mexicanisms (N=41) in the Spanish of Albuquerque is over four times that of archaisms. A mere 10 archaic lexical types were encountered in the Barelas corpus. This is important given that archaisms have long characterized Traditional New Mexican Spanish. The relative lack of use of archaic forms in the Spanish of Albuquerque demonstrates that the Spanish of the city should no longer be considered Traditional Spanish. The use of Anglicisms is also characteristic of Traditional Spanish, but the number of these forms (N=18) is less than half that of Mexicanisms in the Spanish of Albuquerque. The extensive use of Mexicanisms, some of which are very recent (even in varieties of Mexican Spanish), is somewhat surprising given that the Spanish of Albuquerque has always been classified as Traditional New Mexican Spanish, which is characterized by a lack of Mexicanisms. The use of Mexicanisms is supposed to only be common in Border Spanish; but these results clearly show that this is not the case. The use of Mexicanisms is quite robust in the Spanish of Albuquerque, which seems to be the result of an influx of Mexican-born immigrants to the city over the past several decades. It would seem that the use of Mexicanisms comes from those immigrants and their children. Since only native-born *bareleños* were interviewed for the current study,

the use of Mexicanisms by foreign-born Mexican residents will not be explored here. It is plausible, however, that their children have been instrumental in incorporating Mexicanisms in the Spanish of the city. To further explore the leaders of lexical change in the community, we will not turn to the rates of use of Mexicanisms according to the social characteristics of participants. Frequencies of use of Mexicanisms according to these characteristics are displayed in Table 1.

<b>Factor group</b>	<b>Factor</b>	<b>N</b>	<b>%</b>
<i>Sex</i>			
	Female	30/59	50.8
	Male	114/189	60.3
<i>Generation</i>			
	First (25-50)	77/103	74.8
	Second ( $\geq$ 51)	67/145	46.2
<i>Language Preference</i>			
	None	38/57	66.7
	English	90/146	61.6
	Spanish	16/45	35.6
<i>Parents' Birthplace</i>			
	Mexico	31/48	64.6
	United States	113/200	56.5

---

Sex:  $\chi^2 = 1.66$ ;  $df = 1$ ;  $p = 0.198$ ; Generation:  $\chi^2 = 20.2$ ;  $df = 1$ ;  $p = 0.000$ ; Language Preference:  $\chi^2 = 11.9$ ;  $df = 2$ ;  $p = 0.003$ ; Parents' Birthplace:  $\chi^2 = 1.04$ ;  $df = 1$ ;  $p = 0.308$

Table 1. Use of Mexicanisms according to participants' social characteristics

First, it should be noted that not only is there a greater number of types of Mexicanisms in the Spanish of Albuquerque (N = 41; see Figure 3) but also a greater number of tokens of Mexicanisms (N = 144) than archaisms and Anglicisms combined (N = 104). It can fairly be said, then, that the use of Mexicanisms is much more characteristic of the Spanish of Albuquerque than is the use of either archaisms or Anglicisms. That said, the only factors that seem to be of any significance in the

conditioning of use of Mexicanisms within the community are generation and language preference. Frequencies of use of Mexicanisms are very similar by sex (with males using Mexicanisms slightly more frequently than females, at a rate of 60.3% compared to 50.8%) and, somewhat surprisingly, by parents' birthplace (with those of Mexican-born parents using them only about 8% more often than those with U.S.-born parents). It was expected that speakers with at least one Mexican-born parent would use Mexicanisms at a considerably higher rate than other members of the community. Alas, this is not the case. Members of the youngest generation, rather, regardless of where their parents were born, are those who use Mexicanisms most often. They use Mexicanisms at a rate far higher than that of members of the oldest generation (74.8% compared to 46.2%). This difference is similar according to language preference. Very surprisingly, the use of Mexicanisms is most common among participants who have no clear preference for either Spanish or English and those who prefer English. On the surface, this is counterintuitive, but this result is likely due to the fact that younger speakers categorically prefer English or have no preference. In other words, the rates of use are high among these participants not due to their language preferences but due to their age. To see which of these factor groups is statistically significant in the conditioning of use of Mexicanisms in the Spanish of the community, we will now turn to a multivariate analysis of the social factors described in the previous section and their possible conditioning effect on the use of Mexicanisms in Albuquerque Spanish. This analysis was conducted using GoldVarb X, which is a statistical analysis program that generates probability weights corresponding to observed frequencies in a corpus (Lawrence, Robinson & Tagliamonte 2001). This type of analysis is well suited for the current study in that it was designed to handle non-continuous dependent variables with two possible applied variants or groups of variants (in this case, Mexicanisms vis-à-vis archaisms and Anglicisms). In this way, statistically relevant probabilities for the social conditioning of the use of Mexicanisms, which are only characteristic of Border Spanish, can be assessed for a variety in which their use has been claimed to be minimal. The results of this analysis are shown in Table 2.

<b>Factor</b>	<b>N</b>	<b>%</b>	<b>Factor Weight</b>
First Generation	77	74.8	.67
Second Generation	67	46.2	.37
			<i>Range</i> 30

---

Other factor groups included in analysis: 1) sex, 2) language preference, and 3) parents' birthplace.

*Table 2.* Multivariate analysis of the probabilities of co-occurrence of the use of Mexicanisms and participants' social characteristics ( $p = 0.000$ ,  $N = 248$ , Input = 0.589, Log likelihood = -158.282)

As suspected, it is not language preference that conditions the use of Mexicanisms in the community, but rather generation. It is statistically probable that younger speakers will use Mexicanisms in their Spanish (with a probability weight of .67) while it is statistically improbable that older speakers will do the same (with a probability weight of only .37). None of the other factor groups are statistically significant. This means that younger speakers are incorporating Mexicanisms into their Spanish regardless of their language preference, sex, or whether they grew up in a household with at least one Mexican-born parent. These results make it clear that the idea that having Mexican parents leads to maintenance of Spanish is not true. The presence of Mexican Spanish in the community has already led to maintenance among many younger speakers, who have incorporated Mexicanisms into their Spanish since this is their model of Spanish in the community. The use of Traditional Spanish is only common among older speakers of the community who prefer Spanish. These speakers ( $N=3$ ) grew up speaking Traditional Spanish and, as such, use very few Mexicanisms relative to the rest of the community (at a rate of only 35.6%; see Table1). If this change continues to advance, the Spanish of Albuquerque will more closely resemble Border Spanish with the passing of this generation. This should not be altogether surprising since younger speakers are in constant contact with some variety of Mexican Spanish. It is the variety taught in Albuquerque schools, many of which have bilingual education programs. Younger members of the community have consistent interactions with other younger speakers

who were raised in households with at least one Mexican-born parent. Given the strength of peers on the language development of adolescents (Schieffelin & Ochs 1986) and teenagers (Eckert 1989), it is no surprise that Mexicanisms are being spread into the community by the youngest generation. Their use seems to start with children of Mexican immigrants, who began settling in the city in great numbers starting in the 1980s. Now these children are grown, as are the children of U.S.-born parents, who have now assimilated these characteristics. This pattern will likely not reverse with continuing immigration from Mexico, which will ultimately lead to the incorporation of a greater number of Mexicanisms into the Spanish of Albuquerque over time.

## **5. Conclusions**

The Spanish of Albuquerque does not fit neatly into a dialectal classification as either Traditional Spanish or Border Spanish given that, while it does include archaisms (N = 10) and Anglicisms (N = 18), which are characteristic of Traditional Spanish, it includes many more Mexicanisms (N = 41). As it would be absurd to label Albuquerque Spanish “Border Spanish” since it is located some 270 miles from the border and it cannot be labeled “Traditional Spanish” since the use of Mexicanisms in this variety far outnumbers that of archaisms and Anglicisms (by type and token frequency), I propose that Albuquerque Spanish be considered a third major dialect of New Mexican Spanish. The use of Mexicanisms in Albuquerque Spanish is coming from younger speakers who have grown up with other speakers of Mexican-born parents. Their use is not due to the children of Mexican immigrants themselves, however, as clearly shown in Tables 1 and 2. Increased Mexican immigration starting in the 1980s has resulted in the greater use of Mexicanisms, which have spread throughout the city and have begun to supplant archaisms slowly, partly due to the greater prestige afforded to Mexican Spanish and lesser prestige afforded to archaisms and partly due to the fact that Mexican Spanish is the model for Spanish for younger speakers in the community. In this sense, the entire fabric of Albuquerque Spanish can be seen as changing, which has been happening for quite some time. It seems that Traditional Spanish is alive and well in northern New Mexico among the residents who actually speak the language. It could be true, however,

that Albuquerque is representative of a larger change that may affect even smaller communities of the north. It remains to be seen whether “the Traditional Spanish of New Mexico will undergo the dialect extinction that has already befallen the Traditional Spanish of other southwestern states [...] and will have been reabsorbed by its “mother tongue”, Mexican Spanish” (Bills & Vigil 2008: 345). Regardless of the lexical characteristics of the Spanish spoken in northern New Mexico, it seems that the preservation of Spanish is not as imperiled as it was several decades ago. Somewhat ironically, Mexican Spanish, which was once denigrated among native New Mexicans, has led to the preservation of this language in a place where it has been spoken for over 400 years.

## References

- ACADEMIA MEXICANA DE LA LENGUA (2010) *Diccionario de Mexicanismos*, Mexico City: Siglo XXI.
- BILLS, Garland (1997) “New Mexican Spanish: Demise of the earliest European variety in the United States”, *American Speech*, 72, 154-171.
- BILLS, Garland & Neddy VIGIL (1999) “Ashes to ashes: The historical basis for dialect variation in New Mexican Spanish”, *Romance Philology*, 53, 43-60.
- BILLS, Garland & Neddy VIGIL (2008) *The Spanish Language of New Mexico and Southern Colorado: A Linguistic Atlas*, Albuquerque: University of New Mexico Press.
- COBOS, Rubén (1983) *A Dictionary of New Mexico and Southern Colorado Spanish*, Santa Fe: Museum of New Mexico Press.
- COMPANY COMPANY, Concepción (2010) “Introducción”, in Academia Mexicana de la Lengua (ed.), *Diccionario de Mexicanismos*, México, DF: Siglo XXI, xv-xxiii.
- ECKERT, Penelope (1989) *Jocks and Burnouts: Social Categories and Identity in the High School*, New York: Teachers College Press.
- ESPINOSA, Aurelio (1909) *The Spanish Language in New Mexico and Southern Colorado*, Santa Fe: New Mexican Publishing Company.
- HERNÁNDEZ-CHÁVEZ, Eduardo, Garland BILLS & Alan HUDSON (1996) “El desplazamiento de español en el suroeste de EEUU según el censo de 1990”, in M. Arjona Iglesias, J. López hávez, A. Enríquez Ovando, G. López Lara & M. A. Novella Gómez (eds.), *Actas del X Congreso*

- Internacional de la Asociación de Lingüística y Filología de la América Latina*, México, DF: Universidad Nacional Autónoma de México, 664-672.
- HILLS, Elijah Clarence (1906) "New Mexican Spanish", *Publications of the Modern Language Association of America*, 21, 706-753.
- HUDSON, Alan, Eduardo HERNÁNDEZ-CHÁVEZ & Garland BILLS (1995) "The many faces of language maintenance: Spanish language claiming in five southwestern states", in C. Silva-Corvalán (ed.), *Spanish in Four Continents: Studies in Language Contact and Bilingualism*, Washington, DC: Georgetown University Press, 165-183.
- LABOV, William (1972) *Sociolinguistic Patterns*, Philadelphia: University of Philadelphia Press.
- LABOV, William (1984) "Field methods of the Project on Linguistic Change and Variation", *Language in use: readings in sociolinguistics*, ed. by J. Baugh and J. Scherzer, Englewood Cliffs, NJ: Prentice Hall, 8-53.
- LAWRENCE, Helen, John ROBINSON & Sali TAGLIAMONTE (2001) *GoldVarb: A multivariate analysis application for Windows*, York: University of York Department of Language and Linguistic Science and Computer Services.
- ORNSTEIN, Jacob (1975) "The archaic and the modern in the Spanish of New Mexico", in E. Hernández Chávez, A. Cohen & A. Beltramo (eds.), *El Lenguaje de los Chicanos*, Arlington, VA: Center for Applied Linguistics, 6-12.
- REAL ACADEMIA ESPAÑOLA (2014)<sup>23</sup> *Diccionario de la lengua española (DRAE)*, Madrid: Real Academia Española.
- SCHIEFFELIN, Bambi B. & Elinor OCHS (1986) *Language socialization across cultures*, No. 3., Cambridge: Cambridge University Press.
- SILVA-CORVALÁN, Carmen (2001) *Sociolingüística y pragmática del español*, Washington, DC: Georgetown University Press.
- SIMMONS, Marc (1977) *New Mexico: An Interpretive History*, Albuquerque: University of New Mexico Press.
- TORRES CACOULOS, Rena & Fernanda FERREIRA (2000) "Lexical frequency and voiced labiodental-bilabial variation in New Mexican Spanish", *Southwest Journal of Linguistics*, 19, 2: 1-17.
- WILLIAMS, Jerry (1986)<sup>2</sup> *New Mexico in Maps*, Albuquerque: University of New Mexico Press.

[www.city-data.com](http://www.city-data.com)

<http://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml>



APPENDIX

EXAMPLES OF LEXICAL TYPES IN THE SPANISH OF ALBUQUERQUE

**Archaisms**

albercoques	semos
a(n)sina	traiba
haiga	truje/trujo
muncho/a(s)	túnicos
naiden/nadien	vide/vido

**Anglicisms**

apenar (pagar, from 'peni')	rentar
biles	reporte
cachear	rofe
chequear/chechar	rostear
lonche	suera
marqueta	tracas
mopear	troca
presentes	troquero
queque	yarda

**Mexicanisms**

atole	gacho	plebe
carnal (hermano)	güero/a(s)	quiubo
cerecillos	jale (trabajo)	ruco/a
chamaco/a	jefito/a(s)	simón (sí)
chaparral	loquera (delincuencia)	sobadora
chaparrito/a(s)	mande (¿cómo?)	suave
chavalo	manitos	trapear
chido	nomás (sin razón ni finalidad)	valevergas
chingones	onda (ambiente)	vato
cholo	órale	verga
cirqueros	pachuco	zacate
corretear	padre (chido)	zafado/a (chiflado/a)
desarmador	paisa	zoquete
enjarrar	pinta (prisión)	

## **FIRST DIALECTOLOGISTS**



### **MAHMUD BIN HUSEYIN BIN MUHAMMED AL KASHGARI (apx. 1008-1090)**

“For long years, I have wandered the towns and lands of Turks, Turkmen, Oghuzes, Chigils, Yaghmas and Kirghiz, collecting their words, learning and scraping to my mind the various features of these words. I did this work not because of lack of my language knowledge, quite contrary, I did it to show all the things and differences between Turkic dialects” (Esker (2008: 55).

## 1. The manuscript

It was a day filled with glamour and menace contrasting each other in İstanbul. The first year of First World War. Ottoman Empire was snuggled in many fronts. Despite the hard time, Ali Emiri, a retired civil officer, known for his love of manuscripts and books were still in the trace of new pieces to add on his collection. He visited Burhan's antiquarian bookseller. He had heard about an old and enigmatic manuscript which was written in Arabic. Burhan told Ali Emiri that, this manuscript was handed to him by an old woman, and it was nothing more than just a trivial artifact. But when Ali Emiri examined the manuscript carefully a strange feeling of joy entered his body. The book regarded as a trivial artifact was the most magnificent work that was written on Turkic languages. Ali Emiri could die from happiness but he tried not to show his feeling. Ali Emiri bought this priceless manuscript for just 33 liras. The manuscript in question was *Dīwān Lughāt al-Turk* (Compendium of Turkic Dialects) by Mahmud Kashgari. This is the first dictionary of Turkic languages. This work is not only just a lexicon, but it is also a general observer of Turkic tribes.

The first one who mentioned about this work was Bedreddin Mahmud, mostly known under the name Aynî Antepi. He benefited from Mahmud Kasgharî's work in the first volume of his book *Ikdü-l-Cuman fi Tarih-i Ehli'z Zaman*. Aynî used some material from this work also in the *Tarikh-ul-Shihabi*, which he wrote with his brother. Then, Katib Chelebi, too, mentioned the *Compendium* in his work *Keshf-ul-Zünûn*. However, until 1914, there wasn't any solid proof of the existence of the *Compendium*. The manuscript passed a long voyage through the history and geography until it found its place to the dusty bookshelves of an antique shop in İstanbul.

The discovery of Ali Emiri caused a lot of changes in the development of Turkology. As S. Frederick Starr (2013: 315) stated, it is "a ground-breaking text... and a pioneering work of cultural anthropology". Ali Emiri told about the value of the manuscript to his friends with the following words:

This is not only a book; it is the whole Turkestan and even not only Turkestan, but the whole world. Turkness and the language of Turks will gain

another brightness thanks to this book. This is the Turkic equivalent of the Sibawavh's book. There is no this kind of book on Turkic languages. Even the whole universe will not be sufficient for the true value of this book. There's a similarity between this book and Joseph. He was sold for only a few amount of coins, however, later in Egypt, his value was measured in jewels which is equal to his weight. Now Burhan asked for this book only thirty three liras, yet, I won't sell the book for folds of jewels and emeralds.

Under the protection of the Ottoman Prime Minister Talat Pasha, the work published by the well known scholar Rifat Kilisli between the years of 1915-1917. Later, this brilliant piece of Turkic culture translated into Modern Turkish by Besim Atalay (1939-1943), into Uzbek by Salih Mutallibov (1960-1963), into Modern Uigur by a group of researchers (Abbas *et al.* 1981-1984), into English by Robert Dankoff and James Kelly (1982-1985), into Kazakh by Esker Equeubay (1997-1998), into Persian by Hüseyin Düzgün (2004), into Azerbaijani by Ramiz Esker (2006), into Russian by Alibey Rustamov (2010), and into other languages.

The manuscript is a unicum, but unfortunately, it is not the author's original. It was copied by Muhammad ibn Abubekr after two hundred years from the death of Mahmud Kashgarî. The manuscript is preserved in the National Library (*Millet Kütüphanesi*) in the Fatih quarter in İstanbul.

## 2. Biography

*Dīwān Lughāt al-Turk* has been written by Mahmud Kashgari. This man is the first dialectologist, the first lexicographer, and generally the first known linguist of Turkic languages. Besides, we can regard him also as a folklorist, an ethnographer and a philosopher. It is important to note also that he used a comparative method in linguistics seven centuries before than Rasmus Rask, Franz Bopp and Jacob Grimm.

The full name of Mahmud Kashgari is Mahmud bin Huseyin bin Muhammed al Kashgari. There is not enough data about his life: Our whole information in this regard

consists of some information given in his book and some related passages from various works. So, we cannot show year by year his biographical data. It is known that he was born approximately in 1008 in the village Opal near Kashgar (Esler 2008: 6). According to the Arabic historian Ibn al-Athir, he belonged to a noble family, a monarch lineage from the Karakhanid dynasty who reigned in Central Asia. Mahmud Kashgari had received a good education at *medreses*, where he learned mathematics, philosophy and other fields as well as Arabic and Persian.

Mahmud Kashgari dedicated his life to the examination of Turkic languages and culture. He tried to bring cultural pieces of Turkish tribes together.

Mahmud's life as a wanderer begins with a coup. The ruler of Karahanid Khanate Yaghan Tekin decided to left the throne to his son, Muhammed bin Huseyn, the father of Mahmud. However, during the handover ceremony, the youngest wife of Yaghan Tekin poisoned other princes, including Muhammed bin Huseyn, to appoint his own son Ibrahim to the throne. After that, Mahmud Kashgari had fled for his life and to avoid the attention of soldiers of Ibrahim Khan, he introduced himself usually as a wanderer or scholar (Almas 1988: 333-335).

Mahmud Kashgari travelled among Turks, "through their cities and their steppes, learning their dialects and their verses" (Esler 2008: 55). Eventually, starting from the western shores of Caspian Sea, Mahmud Kashgari reached the lake of Lobnor and from the river basin of Ili to the Kashkaria's southern borders. He completely learned this area, identified the borders and monuments of Turkic people.

Kashgari had worked on the *Compendium of Turkic Dialects* for nearly 15 years. While Seljuks were penetrating Byzantine forces at the Manzikert, Mahmud Kashgari were probably preparing for finishing his work in Baghdad. He completed it during 1075-1077 in this city. This was a period of rattling pace for the advance of Turks to the West with great successes and an era when Turkic dynasties became politically dominant in the Islamic world. As Turkic advance was fierce and hegemony on the Muslim world established, Kashgari appointed a duty to himself for show Arabs how Turkic language is rich and perfect. After finishing his work, Mahmud Kashgari presented it to the caliph of Baghdad. "Since it is dedicated to the caliph al-Muqtadi

(1075-1094), its purpose was perhaps to explain to the Abbasid court the language and customs of their Seljuk overlords” (Dankoff 1975: 68).

Mahmud Kashgari returned to his home city Kashgar from Baghdad approximately in 1080. It is assumed that he settled in the village Opal, near to Kashgar, and built a school (*Mahmudiyye Medrese*), where he lectured for ten years. Mahmud Kashgari died here in 1090.

### 3. *Dīwān Lughāt al-Turk*

Mahmud Kashgari’s work consisted of an introduction and eight chapters. In the introduction, the author states that his aim is to describe the dialects of Turks, because people must learn Turkic languages to understand “this race chosen by Allah”.

The chapters are designed according the traditions of Arabic lexicography. The first chapter deals with the words that begin with a vowel. The second chapter describes the words that do not contain the letters ‘alif, waaw, yaa’. The third chapter contains the words that have a double consonant in the writing. The fifth chapter presents the words that begin with the consonant *y*. The topic of the sixth and seventh chapters are three and four letters words, correspondingly. The eighth chapter focuses on the words that contain a velar nasal or a nasal consonant cluster (*nc*). The meanings of the all headwords are given in Arabic. In addition to definitions, Kashgari gave also example sentences or phrases (like proverbs, verses, passages from the folk literature) for many entries.

*Compendium* cannot be considered just a simple Turkic-Arabic dictionary; it is also a general Turkic encyclopedia, as it contains many information about Turkic ethnography, culture, folklore, etc. Kashgari examined the Turkic tribes one by one and systematically. For example, he gave the following information about Oghuz tribe, to which Turkish, Azerbaijani and Turkmen people belong: “They are divided into twenty two divisions; every division has a special document and symbol (*tamga*), stamped using hot irons on domesticated animals. They recognize each other by these symbols. The first division and the leader of all them is Kinik tribe. Kiniks are the rulers of our

time... The mentioned divisions are the roots and several clans branched from these roots. I didn't mention them to cut short. The names of these divisions originate from their founding fathers".

In the book, there are also valuable descriptions of the grammar of Turkic languages and dialects. Kashgari divided them into two groups: "pure" and "mixed". "The purest and most accurate language is the language of those who know just one language, who do not mix with Persians and who do not customarily visit other lands. The language of those who know two languages and are in close contacts with townsmen is corrupted. The bilingual tribes are Soghdak, Kenchek and Argu... The language of Chomul clan, a nomadic people are different, but they know Turkic very well. The tribes Kay, Yabaku, Tatar and Basmil are the same. Even though every tribe has its own dialect, they speak Turkic very well. The only Turkic-speaking tribes are Kirghiz, Kipchak, Oghuz, Tukhsi, Yaghma, Chighil, Charuk and Ughrak. The languages of Yemeks and Bashqurts are similar to theirs... The shortest one is the language of Oghuzes, the most accurate one is the language of Tukhsis and Yaghmas".

Kashgari examined the dialects by the means of lexicon, morphology and phonetics. The differences between the dialects are also given, like this: "Yagmas, Tukhsis, Kipchaks, Yabakus, Tatars, Kays, Chomuls and Oghuzes always turn *d* into *y* and they never articulate *ð*. They call the beech tree *kayıŋ*, whereas others call it *kaðıŋ*".

Many linguistic and extra-linguistic details were given by Kashgari. For instance, information about the alphabet that used by Uyghur's craved on epitaphs, which was only understood by them and Chinese.

There is also an interesting world map in the book. In the center of the map, we see Balasagun, which was once a trade capital. Kashgari spent a period of his life there. The ruins of this city are near Bishkek, capital of Kyrgyzstan. The map is in the form of a circle. The north, south, east and west directions are indicated. The points at the end of the West direction, where according to Kashgari, Kipchaks and *Frenks* live, reach Volga river. The southwest border of the map reaches Ethiopia and the south border reaches India. There are China and Japan at the east border of the map. The map is remarkable for some reasons. Firstly, it shows the regions inhabited by Turks and their

distribution. Secondly, we can learn old Turkic names of lands, rivers and seas. Thirdly, it is the first map of Japan, even though it is more like a draft.

Finally, it should be mentioned that beside the *Compendium*, there was also another work of Mahmud Kashgari, *Kitabu Javahiru'n-Nehv fi Lughati't-Turk*, which is a work on Turkic syntax, but it is still missing. By the way, the Euroasian Association Writers in Ankara, established a great price, a big amount of gold, for the one who will find the manuscript of this work.

Toğrul SULTANZADE

Turkish Maarif College, Famagusta,  
Turkish Republic Northern Cyprus  
tugrul\_sultanzade@hotmail.com.tr

## References

- ABBAS, Abdusalam *et al.* (1981-1984) *Mehmut Keshkeri, Türki Tillar Divani*, Urumchi, 1-3.
- ALMAS, Turğun (1988) *Kedimki Uyğur Edebiyiti*, Keşker.
- ATALAY, Besim (1939-1943) *Divanü Lûgat-it Türk Tercümesi*, Ankara, 1-4.
- DANKOFF, Robert (1975) "Kashgari on the Beliefs and Superstitions of the Turks", *Journal of the American Oriental Society*, 95/1, 68-90.
- DANKOFF, Robert & James KELLY (1982-1985) *Mahmūd al-Kāşyarī Compendium of The Turkic Dialects (Dīvānu Luyāt at-Turk)*, Cambridge Mass., 1-3.
- DÜZGÜN, Hüseyin (2004) *Mahmud el-Kaşğari*, Divanü Lûgat-it-Türk, Tebriz.
- EQEUBAY, Esker Kurmashuli (1997-1998) *Mahmut Kashkari*, Türk Sözdigi. Almati.
- ESKER, Ramiz (2006) *Mahmud Kaşğari, Divanü Lûgat-it-Türk*, Bakı, 1-4.
- ESKER, Ramiz (2008) *Mahmud Kaşğarinin Divanü Lûgat-it-Türk Eseri üzre Bibliografik ve Qrammatik Gösterici*, Bakı.
- MUTALLIBOV, Salih (1960-1963) M., *Mahmud Kashgariy, Türkiy Süzler Devani (Devanü Lugati't-Türk)*, Tashkent, 1-3.



RUSTAMOV, Alibey R. (2010) *Maxmud al-Kashgari, Divan Lugat at-Turk (Svod T'urkskix Slov)*, Moskva, 1-3.

STARR, S. Frederick (2013) *Lost Enlightenment: Central Asia's Golden Age*, Princeton: Princeton University Press.