
Abstract

The main aim of this study is to introduce a model of TEI (Text Encoding Initiative) annotation of Hebrew elements in Judaeo-Arabic texts, i.e. code-switching, borrowing and Hebrew quotations. This paper will provide an introduction to using XML (Extensible Markup Language) to investigate sociolinguistic aspects in medieval Judaeo-Arabic texts. Accordingly, it will suggest to what extent using XML is useful for investigating linguistic and sociolinguistic features in the Judaeo-Arabic paradigm. To provide an example for how XML annotation could be applied to Judaeo-Arabic texts, a corpus of 300 pages selected from three Judaeo-Arabic books has been manually annotated using the TEI P5. The annotation covers all instances of code-switching, borrowing, and Hebrew quotations in that corpus.

Keywords:
Judaeo-Arabic; Hebrew; Arabic; TEI; XML; code-switching; borrowing

Judaeo-Arabic Texts and Code-switching

Introduction
Code-switching (CS) is considered one of the most common phenomena generated by languages in contact deriving from a bilingual/multilingual context where people communicate using more than one language or dialect due to their linguistic background. Although Judaeo-Arabic texts contain the code-switching phenomenon, very few studies have analyzed the linguistic, literary and typological aspects of code-switching in these texts.
The Judaeo-Arabic variety is considered one of the most important parts of Jewish history, culture and intellectuality.¹ Hary (1992: 73) asserts that “Judaeo-Arabic, Yiddish and Judaeo-Spanish have had the largest impact on the Jewish culture and civilization since the dispersion of the Jews. Among the three, Judaeo-Arabic holds a significant position: it has had the longest recorded history of the three, from pre-Islamic time to the present; additionally, it spans the widest geographical area, from Spain to Yemen and Iraq”. Hary (1992) divides the history of Judaeo-Arabic into five stages: Pre-Islamic Judaeo Arabic, Early-Judaeo Arabic, Classical Judaeo-Arabic, Later Judaeo-Arabic and Modern Judaeo-Arabic.

According to Hary (2009), the ‘religiolect’ varieties² of the Jews generally share some distinct features, the most obvious of which is the use of Hebrew scripts. According to Hary, another important and common trait of Jewish religiolects is the utilization of Aramaic and Hebrew elements. In addition, texts written in Jewish religiolects were written for Jewish readers, mainly about Jewish topics.

The topics found in Judaeo-Arabic texts from the Middle Ages are associated with various genres, including religious texts as well as literary texts, medical, philosophical, magical and scientific texts in addition to Belles lettres which have no religious context. One of the familiar texts in the history of written Judaeo-Arabic is the so-called sharḥ (pl. shurūḥ) or interpretation, the translation of sacred and liturgical texts into Judaeo-Arabic.³

This study investigates the phenomena related to mixed-language texts in general and code-switching in particular regarding data from Judaeo-Arabic texts. The methodology and approach to the corpora are derived from the modern approaches developed to annotate these linguistic phenomena in digital resources.

¹ This study follows Khan terminology. According to Khan 2007:526, the term Judaeo-Arabic refers to any written form of Arabic texts with Hebrew script.
² These language varieties include Judaeo-Arabic, Judaeo-Spanish (Ladino), Judaeo-Persian and Yiddish. Early in his career, Hary considered Judaeo-Arabic an ‘ethnolect variety,’ see Hary 1996.
³ sharḥ texts only come about from the 15th century when people could no longer understand classical Judaeo-Arabic anymore. For more information about sharḥ, see: Bar-Asher 2010; Bar-Asher 1999; Hary 2009; Tirosh-Becker 2012.
Making use of modern technological approaches to tag and annotate sociolinguistic aspects in many Judaeo-Arabic texts, can produce a comprehensive analysis. The digital texts annotated by linguistic tags can simplify and expedite the linguistic analysis of large data and lead to some good precise results.

In turn, the research into historical code-switching, in general, can benefit from the results derived from Judaeo-Arabic materials. As mainstream research on historical code-switching has been focused on English and other European languages from the Middle Ages, the study of the same historical phenomena, yet of non-European languages. i.e., Hebrew and Arabic, is definitely long overdue and could provide interesting comparative data and conclusions.

This study works, however, as an example to show the benefit of using modern linguistic tools to study code-switching in Judaeo-Arabic. It does not pretend to be exhaustive and will conclude with general remarks about code-switching in Judaeo-Arabic texts implicitly and explicitly, as further and more comprehensive studies should follow.

The study uses an encoding structure known as Text Encoding Initiative, or (TEI). The TEI is a number of guidelines that have been established by the Text Encoding Initiative Consortium. These guidelines are based on the Extensible Markup Language (XML) and are currently the most updated and utilized rules for marking up digital resources. The study uses TEI P5, which is the current version of the Text Encoding Initiative.

**Review of the Literature**

Code-switching phenomena between Hebrew and Arabic in modern times have attracted the attention of some scholars, such as (Ahmed 2016; Henkin-Roitfarb 2011; Isleem 2014; Mitaib Murad 2013). Code-switching in Judaeo-x varieties has also been investigated by some studies, for

---

4 The TEI P5 Guidelines are available for free on this website: [http://www.tei-c.org/index.xml](http://www.tei-c.org/index.xml)
5 TEI 5 Guidelines, p. xvi. For more information about the XML and TEI relationship see, DeRose 1999. For the history of the development of the TEI P5 used in this article, see Wittern et al. 2009.
6 For more information about Arabic sociolinguistics, see Bassiouney 2009.

Although Judaeo-Arabic texts contain the phenomenon of code-switching, very few studies have analyzed the linguistic, literary and typological aspects of code-switching in them. Hary (1992, 1995, 2012) emphasises the importance of studying Judaeo-Arabic texts in the context of sociolinguistic settings while Blau (1980) stressed the importance of studying this phenomenon in his early study of Judaeo-Arabic written in the Hebrew script.

Studies of Judaeo-Arabic have mainly concentrated on the content of these texts, religious studies, and the linguistic description of Judaeo-Arabic typology and linguistic features (Blau 1981, 1964; Hary 1992; Jastrow 1990; Mansour 1974). The study by Wagner (2010) is closely related to the texts examined for this paper. Wagner investigated the linguistic variety of the Judaeo-Arabic in letters from the Cairo Genizah covering eight centuries (from the eleventh to the nineteenth centuries). Her study describes the register used in Judaeo-Arabic texts by means of a linguistic analysis of a corpus of Judaeo-Arabic letters from different linguistic levels, including phonology, morphology, syntax and orthography.

Wagner and Connolly (2017) present another essential study of code-switching in Judaeo-Arabic documents from the Cairo Genizah. Their paper demonstrates code-switching practice in various Judaeo-Arabic text genres, such as legal documents, religious correspondences and merchant letters from the Cairo Genizah. They skillfully showed many examples of code-switching, which were varied and influenced by different factors, like the educational status of the author and the theme of the text. Among the most important feature of Judaeo-Arabic texts, as pointed out by the paper, is the so-called “script-switching”. This coined term by Wagner and Connolly (2017) describes a unique linguistic practice in some of Judaeo-Arabic texts. The paper demonstrated some cases of inserted Arabic phrases and words written Arabic script, and not in Hebrew as the dominant script of Judaeo-Arabic texts.
**Code-switching, Borrowing and Foreign-Language Quotations**

Code-switching (CS) is one of the most common phenomena resulting from languages in contact, deriving from a bilingual/multilingual context where people communicate using more than one language or dialect due to their linguistic background. This paper will follow Poplack’s definition of CS as: “the alternation of two languages within a single discourse, sentence or constituent” (Poplack 1980: 583). There are two main types of code-switching: inter and intra-sentential code-switching. While Intersentential CS is the act of switching codes between sentences and clauses, Intrasentential CS occurs only within the sentence boundary. This latter type of CS, or code-mixing, is likely to have more linguistic interference between the languages involved in comparison to the intersentential code-switching. This is due to the fact that changing codes inside one sentence requires a certain style to absorb the inherited elements in this sentence. This act may result in applying, changing or sometimes inventing syntactic and/or morphological forms when integrating such intrasentential codes.

Historical code-switching is a suggested sub-category in the investigation of the CS phenomenon in older texts (Schendl & Wright 2011: 23), early studies on historical CS go back to the late 1990s, when Schendl (1996, 1997) introduced the phenomenon of code-switching in old English texts. Older manuscripts have particular characteristics due to the paucity of information about the audience, events and other sociolinguistic questions related to text, reader and author. The study of code-switching from a historical perspective is well established and considerable studies have been made (Braunmüller et al. 2003; Jefferson et al. 2013; Nurmi & Pahta 2004; Pahta & Nurmi 2006; Schendl 2001, 2002, 2005; Schendl & Wright 2011). Judaeo-Arabic texts constitute a good resource for studying historical code-switching. However, CS in Judaeo-Arabic texts has been inadequately

---

7 For more information about the code-mixing term, see for instance: Muysken, 2000.

8 There is a debate about the way the syntactential boundaries of the two languages involved in code switching are used. Some people claim that in the case of intrasentential code-switching, the syntactic structures are likely to be shared by both languages involved. Accordingly, the relationship between the two languages is symmetric Poplack 1980; Lipski, 1977. On the other hand, some argue the contrary: the syntactic structure of switch between the two languages involved in code-switching is rather based on an asymmetric relation. This means that alternation can happen in one direction, this claim is mainly suggested by Scotton, in which the notions of dominant language vs embedded language were introduced in her Matrix Language Frame model by Myers-Scotton 1998.
investigated to date, and this study closes that gap in the research. This study investigates the phenomena related to mixed-language texts in general and code-switching in particular in Judaeo-Arabic texts in the vein of historical code-switching.

Since Judaeo-Arabic texts are full of Hebrew quotations, it is very important to determine whether these quotations should be considered as instances of code-switching or not. The debate about code-switching and foreign language quotation raised by De Brabanter (2004) deserves attention. His discussion about the issue of foreign quotations is very relevant to the study of code-switching in Judaeo-Arabic texts in general. De Brabanter tried to distinguish between interpreting foreign language sentences incorporated in a text as quotations or rather considering them as instances of CS, he claims:

As it turns out, only the question whether the ‘foreign-language’ sequence we are looking at is mentioned or not (on top of being used) might well be a discriminating factor. If there is mention, we are dealing with non-recruited quotation; if there is not, we are dealing with code-switching.’”

Based on the above quotation, this study considers every instance of Hebrew quotations as code-switching, unless there is an Arabic reference mentioning and/or orthographic quotation marks that signify the Hebrew quotations.

Although the debate about the differentiation between borrowing and code-switching is out of the scope of this paper, it is important to shed light on the two terms. First of all, the two terms here belong to bilingual practices, and highlight languages in contact. While code-switching constitutes mostly switches of sentences and lexical items, borrowing is a sort of switching or mixing between language systems involved in conversational or written practices. For instance, Pfaff (1979: 295-9)

(298) distinguishes between code-switching and borrowing by means of surface syntax, lexical inventory and functional load.

Poplack (1993:256) concludes with this statement to define code-switching: “code-switching is the juxtaposition of sentences or sentence fragments, each of which is internally consistent with the morphological and syntactic (and optionally, phonological) rules of the language of its provenance”.

Concerning borrowing, some scholar argues that the term is more likely associated with single lexical usages, in which it is ruled by morphological and phonological integration in line with the predominant language, in Clyne’s words: "codeswitching is employed for both single-word and multi-word elements, borrowing is limited to the former" (2003: 71). On the other hand, other argue that borrowing can occur in both single and multi-word combination: “Lexical borrowing involves the incorporation of individual L2 words (or compounds functioning as single words) into discourse of L1” (Poplack et al 2009: 52).

The Corpus

The paper investigates code-switching phenomenon in three Judaeo-Arabic works by three authors: Moshe ben Jacob ibn Ezra’s (1055/1060 - 1138) Kitāb al-Muhādara wa-l-Muḍākara; Yehuda Halevi’s (c. 1075 – 1141) al-Kitāb al-Kuzari; and Saadia ben Joseph al-Fayyumi’s (882/892 – 942) Kitāb al-Mukhtar fī l-Amānāt wa-l-Iʿtiqādāt. The texts were chosen to represent various geographic, historical and literary milieu. Geographically, the first two texts were written in Spain, while the latter was written in Iraq. Historically, they span several centuries, from the 10th to the 12th centuries. Literarily, their themes vary, although the three texts have indirect religious purposes. Kitāb al-Muhādara wa-l-Muḍākara tries to prove that the biblical text has poetic value as part of a polemical stance with Arabic models. Al-Kitāb al-Kuzari is written as an imagined dialogue between the king of the Khazars and a Jewish rabbi discussing the teachings of Judaism. In Kitāb al-Mukhtar fī l-Amānāt wa-l-Iʿtiqādāt al-Fayyumi tries to defend Rabbinic Judaism against the views of the who rejected the oral law. Having selected such varied texts, the phenomenon of code-switching of Hebrew elements in Judaeo-Arabic texts can be tested and appropriately analyzed.
A linguistic annotation of 300 pages was conducted, that is, the first 100 pages of each book. Digital versions of the texts available online on the Friedberg Judaeo-Arabic Project website were used as bases for the analysis.\textsuperscript{10} The annotation was done by a single annotator - the author of the article.\textsuperscript{11}

**Annotation Methods**

Code-switching, borrowing and quotation instances in the corpus, were manually annotated by the author of the paper.

To mark the digital resources, the TEI uses basic ‘elements,’ which can be described by ‘attributes.’ In the following is an explanation of the tags used for marking the Hebrew elements in the corpus. The predominant language in the Judaeo-Arabic texts used for the analysis is Arabic (written in Hebrew script), the embedded language is Hebrew. The project uses the TEI universal attribute \texttt{@xml:lang} when detecting the language codes in the text: the value “ar” is used to refer to Arabic sentences and phrases in the text, whereas the value “he” is devoted to tag Hebrew language.

<foreign> is another basic element used in this study to detect, in general, the foreign language(s) lexical items or sentences employed in the texts (other than Arabic). The element <foreign> can be attributed by the universal attribute \texttt{@xml:lang}. The attribute \texttt{@ana} is another attribute used widely along with the element <foreign> in the corpus; it is used to indicate the analysis of the foreign

\textsuperscript{10} The Friedberg Jewish Manuscripts Society maintain a website that contain many resources including some important materials of Judaeo-Arabic texts, please see the website: \texttt{http://www.jewishmanuscripts.org}

\textsuperscript{11} In annotation tasks of this sort, it is better to have more than one annotator to conduct an inter-annotator agreement study in order to validate annotation decisions and resolve any inconsistencies. This is to avoid the risk of subjective decisions that are not anchored in the guidelines. This paper, however, works as a seed to a more prominent project that would include larger corpora. The main contribution of this article is to give an example of the importance of using XML to investigate some sociolinguistic aspects in medieval Judaeo-Arabic texts. The work on a larger project, which includes various medieval annotated corpora, seeks both funding and a professional teamwork of linguists and computational experts. Such teamwork can carry out many tasks that a single annotator cannot do. One of these tasks is to conduct inter-annotator agreement testing.
codes. Basically, three linguistic aspects are marked under the @ana attribute; these are intersentential code-switching (inter-CS), intrasentential code-switching (intra-CS), and borrowing.

Due to the religious characteristics that many Judaeo-Arabic texts possess, one expects to find a great deal of quotations from Jewish religious resources, such as Pentateuch and Talmud. In such cases, when quotation marks or references are used in the texts, the study uses the TEI element <quote> to refer to the Hebrew quotations. The study considers all foreign language usages in the texts as code-switching instances, only those mentioning, or with quotation marks are treated as quotations.12 Again, the attribute @xml:lang is useful to identify the language of the analyzed quotation.

<distinct > is another element one can use to indicate certain unusual uses of languages.

The text is marked as numbered pages using the element <p> together with the attribute @n to refer to the number.

**Table 1 XML elements, attributes and their usage**

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
<th>Attribute</th>
<th>Value</th>
</tr>
</thead>
</table>
| <foreign> | To tag foreign language codes in the text. | @xml:lang | - he = Hebrew.  
|           |                              |           |   - ar = Arabic.  
|           |                              |           |   - grc = Greek.  
|           |                              | @ana      |   - intra-CS = intrasentential code-switching.  
|           |                              |           |   - inter-CS = intersentential code-switching.  
|           |                              |           |   - borrowing.  |
| <quote>   | To tag quotes.               | @xml:lang | - he = Hebrew.  
|           |                              |           |   - ar = Arabic.  |
| <p>       | To tag page numbers.         | @n        | numbers.  |

12 For more information see the methodology section in this paper, De Brabanter 2004.
<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>&lt;distinct&gt;</code></td>
<td>To indicate certain unusual uses of languages, such as abbreviations.</td>
</tr>
<tr>
<td><code>&lt;persName&gt;</code></td>
<td>To tag personal names.</td>
</tr>
</tbody>
</table>

The Analysis

**XPath Queries**

Using the XPath queries can ease the process by counting the results within seconds. The XPath query is used to navigate through elements and attributes in an XML document. For the data annotated in the corpus at hand, the following queries were used to extract the number of the attributes and elements of the data:

1. `count(//foreign[@ana='intra-CS'])`
2. `count(//foreign[@ana='inter-CS'])`
3. `count(//foreign[@ana='borrowing'])`
4. `count(//quote[@xml:lang='he'])`
5. `count(//quote[@xml:lang='ar'])`

XPath query in n.1 is to count all `<foreign>` elements that are connected by the attribute `@ana` with the value 'intra-CS,' which refers to the intrasentential code-switching cases in the text. Moreover, the XPath selects all the cases of intra-CS in the text, therefore one can easily trace every instance of the selected elements and attributes one by one, and read them as they appear in the text and not as if isolated.
The same holds true for the XPath query n.2, yet the search this time was for the value ‘inter-CS.’ The same can be also applied – as in XPath query 3 – when detecting ‘borrowing’ cases in the texts, by replacing the value to search and count the borrowing instances that have been manually annotated.

With the help of XPath query n. 4 the paper was able to count and trace all Hebrew quotations used in the text, as the query searches for the element <quote>, with the attribute @xml:lang to determine the language of the <quote> element, which is determined by ‘he’ for Hebrew language. The same case applies for XPath query 5 with ‘ar’ for Arabic.\textsuperscript{13}

This method is very efficient especially for large corpora that contain multiple texts, as the XPath queries can be used for single XML documents as well as for large XML projects with multiple texts. Additionally, the numerical results should be accurate, provided that the manual XML annotation of plain original text was done accurately.

\textit{The numerical results}

The following table shows the numerical results of the Hebrew use in the selected Judaeo-Arabic texts:

\begin{table}
\caption{Numerical Results of Hebrew Elements in the Corpus.}
\begin{tabular}{|l|c|c|c|c|c|}
\hline
The texts & Word count & Intra-CS & Inter-CS & Borrowing & Hebrew Quotations Arabic Quotations \\
\hline
Kitāb al-Muḥāḍara wal-Muḍākara & 24316 words & 112 & 4 & 7 & 112 & 4 \\
\hline
\end{tabular}
\end{table}

\textsuperscript{13} This is a universal coding system of the languages codes used with the \texttt{xml:lang} attribute, for the complete universal list of the language tags used in TEI, see this link: \url{http://www.iana.org/assignments/language-subtag-registry/language-subtag-registry}
According to Fig.1, the employment of Hebrew uses in the all three texts selected for the corpus is divided between the intra-CS and Hebrew Quotations followed by Borrowing. All of the texts have very few cases of inter-CS. This result is close to the findings of an investigation of Arabic code-switching in some modern Hebrew texts (Ahmed: 2016), in which the use of intra-CS codes is much more common than the use of inter-CS.14

---

14 One should consider some Hebrew lexical items in Judaeo-Arabic texts as loanwords, like rabenu. This study considers this lexical item as a loan word. This item has been used only three times in the book of Ibn Izra Kitāb al-Muhādara wal-Muğākara.
The use of intra-CS varies according to each text; in Kitāb al-Muḥāḍara wal-Muḏākara, the use of both intra-CS and Hebrew quotations is equal. The use of intra-CS in al-Kitāb al-Kuzari is almost four times greater than the insertion of Hebrew quotations. On the other hand, the analysis of the Kitāb al-Mukhtār fī l-Amānāt wal-Iՙtiqādāt shows that the intra-CS aspect was lower than the use of Hebrew quotations. Borrowing is not used very much in the samples taken for this study; only in al-Kitāb al-Kuzari did the author use borrowing extensively, in fact even more than his use of Hebrew codes.

It is predictable in religious texts to find a great deal of religious quotations, especially in the Shurūḥ texts of ritual and interpretation of the holy books, like Mishna and Talmud. This also holds true for the corpus at hand. In the al-Fayyumi sample, the table above shows that using Hebrew quotations is six times more regular than intra-CS, with the same sample showing only one case of
inter-CS. Indeed, the code-switching as a bilingual practice seems to have been used extensively in religious texts compared to other genres. In line with this argument, there are some English religious texts from the medieval and early modern eras showed more code-switching than texts from other genres did (Pahta and Nurmi 2006: 223): “The research also indicates that religion is one of the social domains in medieval and early modern England where code-switching is more frequent than in most other domains, with the exception of science.”

The sample gathered from the book *Kitāb al-Muḥāḍara wal-Muḍākara* has greater percentage of Hebrew quotations than examples of CS. Also, this book has some non-Hebrew instances of CS, the analysis of the sample shows three cases of Greek terms employed by the author in the text.

**XML Annotation of Hebrew Elements in the Corpus**

In the following, XML annotated texts from the corpus will be introduced. The examples were chosen to represent the three main categories of the analysis, i.e. intra/inter-sentential code-switching, borrowing, and Hebrew quotations). An analysis of selected examples is also provided to represent the qualitative interpretation of the corpus.

The cases of CS were marked with the elements <foreign>, which indicates the use of a foreign language in the Judaeo-Arabic texts in the corpus. The Arabic works here as the predominant language of the texts.\(^{15}\) The attributes that the study used to tag the different cases of code-switching were basically: @xml:lang and @ana. The former attribute indicates the foreign language value, which are Hebrew (he), and Greek/old Latin (grc). The attribute @ana is used to identify the type of CS instances.

\(^{15}\) Please note that in some Judaeo-Arabic texts, there is much use of Hebrew quotations from the religious resources, such as the *Shurāh.*
In Fig. 2, the paper shows the three categories of analysis of the embedded foreign codes in the corpus: inter/intra-CS and quotations. The first example here is the inter-CS case on page 20 of the text:

**Inter-sentential Code-switching**

![Figure 2. XML annotation of Hebrew codes in Kitāb al-Muḥādarah wa al-Muzākara, p. 20](image)

Arabic transcription

مرتجلي الشعر وناثريه نثرا. وناثرو الشعر عند العرب يُسمون الفارطين، أي الذين يشعرون ارجالاً، כדויד חשבו להם כלי שיר.

وهم الشهاب. 

(1)

According to Arabs, poetry extemporizers are called Fāreṭi. i.e. persons who extemporize poetry, “like David, they devise for themselves instruments of music”.

The inter-sentential case is obvious in example (1), as the paragraph in page 20 in *Kitāb al-Muḥādarah wa al-Muṣākara* begins with a complete Arabic sentence and a clause. The last sentence was rather in Hebrew. In other words, the Hebrew sentence here, which is a verse from the book of

---

16 Singular *Fāreṭ*: The word refers to the poets who extemporize Arabic poems. According to Arabic lexicon, the common term that gives the same meaning is (مرتجل) *Murtagel*. To my knowledge, the term *Fāreṭ* is probably only found in the Arabic-French dictionary *Supplément aux dictionnaires arabes* (Dozy 1881: 255).

17 1 Kings 4:32

18 Amos 6:5
Amos, has not been integrated within the main Arabic text, as we will see later in example (2). The most important benefit of using the XML tagging is that one can easily trace all instances of certain segments throughout the corpus, or even the corpora. As such, it is simpler to revise, confirm and edit the instances of CS cases one by one.

**Intra-sentential Code-switching and Borrowing**

![Figure 3: Intra-CS cases in al-Kitāb al-Kuzari, p. 50](image)

Arabic transcription of the text:

> يوجد إلى متى اتفق وحيث اتفق هكذا صارت النبوة في نسله في الشام كثر

> أهلها طول بقائهم في الشام | مع القرائن المعينة من الظهورات والعبادات والقرابين لا سيما بحضور الـ كفلات. لأن الأمر الإلهي كالمتقب لم يستحقي أن يتصلى به فيصير له الهَا كالأنبياء والأولياء.
كما أن العقل مرتقب لمن كملت طبائعه واعتدلت نفسه واخلاقه أن يحل فيه على الكمال كال فلاسفة كما أن النفس مرتقبة لمن كملت قواعده الطبيعية كما أن النفس مرتقبة للفضيلة أزيد فتحل فيه كالحيوان كما أن الطبيعة مرتقبة للمجاز المتعادل في كيفية تحليتها فيصير نباتاً كالحيوان. قال الكزري هذه جملة علم تحتاج إلى تفصيل ليس نحن الآن في سبيله وسأسألك عنه في موضع العلم ففصل كلامك في فضائل lưuز إسرائيل.

قال الحبر أنها كانت موقوفة لهدية المعمور مقررة لأسباط بنى إسرائيل منذ تفرق الألسن كـ قـ بـهـانـل عـلـيـوـن ـتـمـلـلـقـيـمـ. ولم يمس لـ سـمـرـمـ ليتصل بأمر إلهي وأن يتعاقد معه إلا بعد حصوله في هذه الأرض في مشهد إـبـرـاهـم. فما ذلك بجملة صفوية استحقوا اسم لـ إـبـراـهـم وفـي أـرض خاصية تسمت زـمـلـاـبـ ـبـ في أوقات مفروضة من عهد تعاون لا مصطلحة عليها ولا مؤاخذة ( مأخوذة ) من علوم النجوم ولا غير ذلك بل ما يسمى زـمـلـاـبـ ـبـ مع طهارات وعبادات وكلمات وأفعال مقدّرة من عهد يسمى زـمـلـاـبـ ـبـ.

Intra-CS constitutes the dominant element found in the corpus of this study. In the following, the paper shows some cases of intra-CS annotated in page 50 from the book al-Kitāb al-Kuzari (see figure 3).

The instances of intra-CS and borrowing cases are well represented in this page. The first example shows a common, yet a significant linguistic phenomenon found in many Judaeo-Arabic texts; this is the use of an Arabic utterance morpheme together with Hebrew lexical units. In (2-1), the Arabic definite article /al/ has been incorporated into the Hebrew lexical item shekhinah:

(2-1)
Especially in the presence of shekhīnah (the dwelling the divine presence of God)

Borrowing also appears when using the so-called ‘asides’ reference before the switched lexical items (e.g. it is called, the so-called, etc…). The example (2-2) illustrates this phenomenon:

(2-2)

... بل ما يتسمى معلومات مع طهارات
وعبادات وكلمات وأفعال مقدّرة من عنده تتسمى عملت وPEAR
... However, what is called God Times, together with cleanliness
cult, words, and other destined actions from God, are called God’s Works and God’s Creation.

In this example, the author uses the Arabic verb yatasammā (is being called) to indicate the Hebrew terms (Moshe YIY and (Mala'h YIY). This also holds true for the Arabic preposition /l/ that has been used with the Hebrew personal name Avraham as shown in (2-3):

(2-3)

 ولم يصح لـ أبراهيم
It was not allowed to Abraham…

(2-4)

وسأسألك عنه في موضوع العلم فصل كلمك في قضايا إسرائيل.
I will ask for an explanation about this (subject) in the section for science, so continue your talk about the virtues of ēretz yisrāʾēl (Israel).

In the above example, the Hebrew lexical items constitute a part of the Arabic verbal sentence. It begins with the imperative verb /ṣill/ (continue), and the Hebrew term comes in the position of the genitive with the Arabic object fadāʾil (virtues). In this way, the intra-CS Hebrew lexical items here are more integrated into the Arabic sentence than the case of inter-CS in (1).
The Hebrew quotations are widely incorporated in almost all Judaeo-Arabic texts. The author has two main options to make such a reference; by means of using lexical items that indicate a quotation, and/or the use of proper orthographic marks, such as quotation marks. The above (fig. 4) is given as an example of the different utterance uses to refer to a quotation in Kitāb al-Mukhtār fī al-Amānāt wa al-i’tiqādāt:

Special Uses of Language in the Texts
Looking at page 20 in Fig. 5, the element <distinct> is used to identify Judaeo-Arabic shortcuts that hold Hebrew style, for instance:
The three examples above show an important feature in Judaeo-Arabic texts; the use of Hebrew style abbreviations is also applied to common Arabic expressions in the texts as if they were Hebrew terms. The first two examples (4-1 and 4-2) show that the Arabic terms alayhe l-salām (peace upon him), and alayhim gamīʿa-l-salāmu wa-l-rāḥma (peace and mercy upon all of them) are written in the same way the Hebrew term יהוה was deployed in example (4-3). Tagging such distinct language uses in Judaeo-Arabic texts can also draw a general picture about the way Arabic and Hebrew were merged in one text. This was not only by means of inserting Hebrew quotations and the frequent insertion of code-switching instances, but also by employing some linguistic usages from both languages in the Judaeo-Arabic texts.
Conclusion

The main aim of this paper was to introduce an initial model of using the XML for annotating and detecting Hebrew elements in Judaeo-Arabic texts. This model is mainly concerned with the linguistic features associated with sociolinguistics. The XML annotation of code-switching (CS), borrowing and foreign language quotations are of the main concern of this study.

The intention of introducing this model is to propose a new methodology for investigating CS in medieval bilingual practices of two important Semitic languages, i.e. Arabic and Hebrew. Studying such linguistic phenomena in the Judaeo-Arabic paradigm, will definitely add to our understanding of historical code-switching by confirming and/or providing new insights into the current scholarship on historical code-switching. This way, I believe that CS in Judaeo-Arabic texts constitutes a vital resource for studying historical CS, and should be compared and contrasted with other resources and corpora, mainly of European languages, under investigation.

References

**Resources:**

רמב"ע (משה בן עקבי טרא), "ספר העיונים והדרים": עריך, תרגום ועריכה, ירושלים, מתקפת חלון

יורשלאע, מוקדש ונרחב, תש"א

ריה"ל (יהודה בן שמואל הלוי), כתאב אלרד ואלדליל פי אלדליל (ספר הכוזרים), תאליף ר' יהודה הלוי, הוציאו א"ל דוד צבי בנעט; התקין את הטקסט חגי בן שמאי, ירושלים: י"א מאגנס, האקדמיה הלאומית הישראלית, תשל"ז

סעדיה בן יוסף פיומי גאון, ספר הנבחר באמונות ובדעות, כתאב אלמכ'א תאר פי אלאמנאנאת ואלאעתקאדאת, לרבנו סעדיה בן יוסף, ירושלים: י"א מונט, תשל"ו.


Hary, Benjamin. 1999. “Hebrew elements in Egyptian Judeo- Arabic texts.” In Vena Hebraica in


