

XML Annotation of Hebrew Elements in Judaeo-Arabic Texts.

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).⁴ 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.⁵ 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).⁶ 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>

⁵ TEI 5 Guide lines, 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.

instance (Bahat 2002; Bar-Asher 2013; Bar-Asher 1992; Berk-Seligson 1968; Chetrit 1991; Hary 1999; Leslau 1945; Szulmajster-Celnikier 2005).

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,⁷ 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.⁸

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 syntactic 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.⁹ 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 For more information about the difference between borrowing and code-switching, see Onysko 2012. Field 2002; Haugen 1950; Lipski, 2005; Myers-Scotton 1992; Poplack et al. 1988.

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-Muḥāḍara wal-Muḍākara*; Yehuda Halevi’s (c. 1075 – 1141) *al-Kitāb al-Kuzari*; and Saadia ben Joseph al-Fayyumi’s (882/892 – 942) *Kitāb al-Mukhtār fī l-Amānāt wa-l-I’tiqādāt*. 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-Muḥāḍara wal-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-Mukhtār fī l-Amānāt wa-l-I’tiqādāt* 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.¹⁰ The annotation was done by a single annotator - the author of the article.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 **@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 **@xml:lang**. The attribute **@ana** is another attribute used widely along with the element <foreign> in the corpus; it is used to indicate the analysis of the foreign

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: <http://www.jewishmanuscripts.org>
The study at hands uses the plain digitalized texts, the linguistic annotation has been done manually to these texts by the author of the paper.

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*

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

¹² For more information see the methodology section in this paper, De Brabanter 2004.

<distinct >	To indicate certain unusual uses of languages, such as abbreviations.
<persName>	To tag personal names.

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.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.

The numerical results

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

TABLE 2 *Numerical Results of Hebrew Elements in the Corpus.*

The texts	Word count	Intra-CS	Inter-CS	Borrowing	Hebrew Quotations	Arabic Quotations
כתאב אלמהאצ'רה ואלמד'אכרה <i>Kitāb al-Muḥāḍara wal-Muḍākara</i>	24316 words	112	4	7	112	4

¹³ This is a universal coding system of the languages codes used with the *xml:lang* attribute, for the complete universal list of the language tags used in TEI, see this link: <http://www.iana.org/assignments/language-subtag-registry/language-subtag-registry>

Moshe ibn Ezra						
אלכתאב אלכזרי <i>al-Kitāb al-Kuzari</i>	21902 words	339	1	129	94	0
Yehuda Halevi						
כתאב אלמכתאר פי אלאמנאנת ואלאעתקאדאת <i>Kitāb al-Mukhtār fī l- Amānāt wal-I'iqādāt</i>	20826 words	37	1	0	236	0
Saadia ben Joseph al- Fayyumi						
Total	67044	488	6	136	442	4

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 Ezra *Kitāb al-Muḥāḍara wal-Muḍākara*.

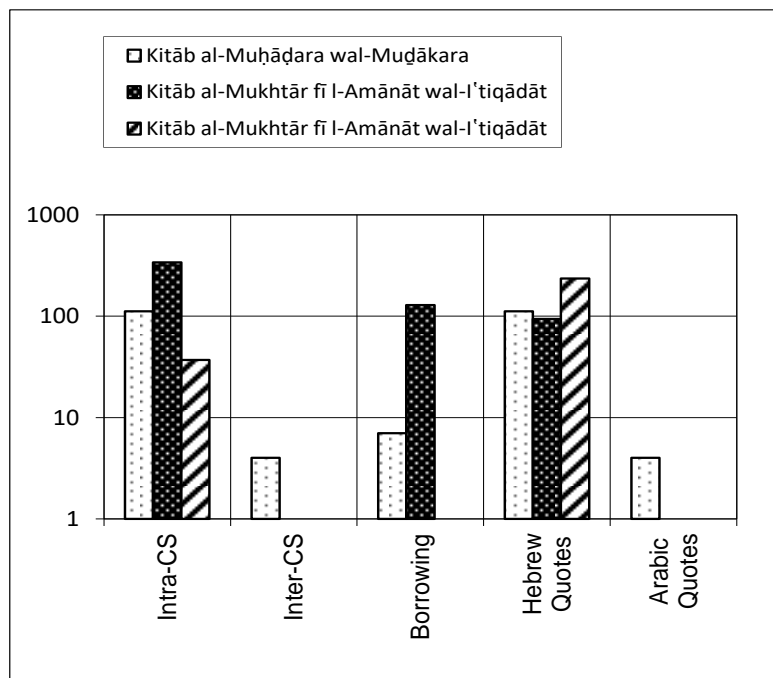


FIGURE 1. *Hebrew elements in the corpus.*

The use of intra-CS varies according to each text; in *Kitāb al-Muḥāḍara wal-Muḍākara*, 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-l'tiqādāt* 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 *Shurūḥ*.

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

```
<p n="20">
  מרתג'לי אלשער ונאת'ריה נת'רא. ונאת'רו אלשער ענד אלערב יסמון אלפארטין, אי אלד'ין ישערין
  ארתג'אלא
  <foreign xml:lang="he" ana="inter-CS">כדויד חשבו להם כלי שיר</foreign>
  ׀
  <foreign xml:lang="he" ana="intra-CS">שיר</foreign>
  אסם אלשער, כקולה
  <quote xml:lang="he">ויהי שירו חמשה ואלף</quote>
```

FIGURE 2. XML annotation of Hebrew codes in *Kitāb al-Muḥādarah wa al-Muzākarah*, p. 20

Arabic transcription

مرتجلي الشعر وناتريه نثرا. وناترو الشعر عند العرب يُسمون الفارطين¹⁶، أي الذين يشعرون
ارتجالاً، كدوיד حشبو להם כלי שיר. ושיר اسم الشعر، כקולה ויהי שירו
חמשה ואלף.¹⁷

(1)

وناترو الشعر عند العرب يُسمون الفارطين، أي الذين يشعرون ارتجالاً، كدوיד حشبو להם כלי שיר.

According to Arabs, poetry extemporizers are called *Fāreṭīn*. i.e. persons who extemporize poetry, “like David, they devise for themselves instruments of music”.¹⁸

The inter-sentential case is obvious in example (1), as the paragraph in page 20 in *Kitāb al-Muḥāḍara wal-Muḍākara* 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) *Murtajel*. 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

```
<p n="50">
  יוג'ד אלא מתי אתפק וחית' אתפק. הכד'א צארט אלנבוה פי נסלה פי אלשאם כת'ר אהלהא טול בקאיהם
  פי אלשאם מע אלקראין אלמענינה מן אלטהאראת ואלעבאדאת ואלקראבין לא סימא בחצ'ור אל
  <foreign xml:lang="he" ana="borrowing">שכינה</foreign>
  לאן [אלאמר] אללהי כאלמרתקב למן יסתחק אן יתצל בה פיציר לה אללהא כאלאנביא ואלאוליא. כמא
  אן אלעקל מרתקב למן כמלת טבאיעה וועתדלת נפסה ואכ'לאקה אן יחל פיה עלי אלכמאל כאלפלספה
  כמא אן אלנפס מרתקבה למן כמלת קואה אלטביעיה כמאלא מסתעדא לפצ'ילה אויד פתחל פיה כאלהיואן
  כמא אן אלטביעיה מרתקבה ללמוזאג' אלמתעאדל פי כיפיאתה לתחלה פיציר נבאתא. קאל אלכ'זרי הד'ה
  ג'מל עלם תחתאג' אלי תפציל לים נחן אלאן וסאסאלך ענה פי מוצ'ע אלעלם פצל כלאמך פי
  פצ'איל
  <foreign xml:lang="he" ana="intra-CS">ארץ ישראל</foreign>
  קאל אלחבר אנהא כאנת מוקופת להדאיה אלמעמור מוקרהה לאסבאט בני אסראיל מגד' תפרקת אלאלסן
  <distinct>כ'ק'</distinct>
  <quote xml:lang="he">בהנחל עליון גוים</quote>
  ולם יצח ל
  <persName>אברהם</persName>
  ליתצל באמר אללהי ואן יתעאהד ויתעאקד מעה אלא בעד חצולה פי הד'ה אלארץ' פי משהד
  <foreign xml:lang="he" ana="intra-CS">בין הבתרים</foreign>
  פמא ט'נך בג'מלה צפוח אסתחקוא אסם
  <foreign xml:lang="he" ana="intra-CS">עם ייי</foreign>
  ופי ארץ' כ'אצה תסמת
  <foreign xml:lang="he" ana="intra-CS">נחלת ייי</foreign>
  פי אוקאת מפרוצ'ה מן ענדה
  <distinct>תע'</distinct>
  לא מצטלה עליהא ולא מאכ'וד'ה מן עלום אלנג'ום ולא גיר ד'לך בל מא יתסמי
  <foreign xml:lang="he" ana="borrowing">מועדי ייי</foreign>
  מע טהאראת ועבאדאת וכלמאת ואפעאל מקדרה מן ענדה תתסמי
  <foreign xml:lang="he" ana="borrowing">מלאכת ייי ועבודת ייי</foreign>
</p>
```

FIGURE 3. Intra-CS cases in *al-Kitāb al-Kuzari*, p. 50

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 القرانين المعينة من الطهارات
 والعبادات والقرابين لا سيما بحضور الـ سכינה. لأن الأمر الالهي
 كالمرتقب لمن يستحق أن يتصل به فيصير له الهأ كالأنبياء والأولياء

كما أن العقل مرتقب لمن كملت طبائعه واعتدلت نفسه واخلقه أن يحل
فيه على الكمال كالفلاسفة كما أنّ النفس مرتقبة لمن كملت قُواه
الطبيعية كمالاً مستعداً لفضيلة أزيد فتحل فيه كالحَيوان كما أنّ
الطبيغة مرتقبة للمجاز المتعادل في كفيّتها لتحلّه فيصير نباتاً
قال الكزري هذه جُمْل علم تحتاج إلى تفصيل ليس نحن الآن بسبيله
وسأسألك عنه في موضع العلم فَصِلْ كلامك في فضائل اَرْمِز ישראל.
قال الحبر أنها كانت موقوفة لهداية المعمور مقررة لأسباط

בני ישראל منذ تفرقت الألسان כִּדְּכֹהֵנְהוּלָּא עֲלִיּוֹן גִּזִּים. ولم يصح لـ אברהם
ليُتصل بأمر إلهي وأن يتعاهد ويتعاقد معه إلا بعد حصوله في هذه الأرض
في مشهد בין הבתרים. فما ظنك بجملة صفة استحقوا اسم יי وفي أرض
خاصة تسمت נחלת יי في أوقات مفروضة من عنده تع لا مصطلحة عليها ولا
مؤاخذة (مأخوذة) | من علوم النجوم ولا غير ذلك بل ما يتسمى מזוּעָדֵי יי مع طهارات
وعبادات وكلمات وأفعال مقترّة من عنده تتسمى מְלֻאָכֶת יי וְעִבּוּדֹת יי

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 *shekhinah* (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 **מזעדי** יי مع طهارات

و عبادات وكلمات وأفعال مقدرة من عنده تتسمى **מלאכת** יי **ועבודת** יי

... 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* **يتسمى** (is being called) to indicate the Hebrew terms (**מזעדי** יי) and (**מלאכת** יי **ועבודת** יי). 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 **אברהם**

It was not allowed to Abraham...

(2-4)

وسأسألك عنه في موضع العلم **فَصِلْ** كلامك في فضائل **ארץ** **ישראל**.

I will ask for an explanation about this (subject) in the section for science, so continue your talk about the virtues of *ēretz yisrā’el* (Israel).

In the above example, the Hebrew lexical items constitute a part of the Arabic verbal sentence. It begins with the imperative verb /*ṣil*/ (continue), and the Hebrew term comes in the position of the genitive with the Arabic object *fadā’il* (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).

Quotations

```
<p n="16">
ד'א תעלם אלכ'פה ואלת'קל, פנקול בדאים אלהרכה פתוג'ד סהלה או צעבה. ואלאכ'רי אלנטק אד' קאל
<quote xml:lang="he">לא יתגו בגרונם</quote>
והו ג'מלה אלכלאם ד'ו אלסמא ואלתאליף ואלמקדמאת ואלבראהין עלי מא שרחנא. ת'ם חקק עלם
אלעקל פאמר באלצדק לא אלכד'ב כקולה
<quote xml:lang="he">כי אמת יהגה חכי ותועבת שפתי רשע, בצדק כל אמרי פי אין בהם נפתל</quote>
ועקש
ת'ם חקק לנא אלעלם אלצ'רורי באנה כל מא יודי אלי דפע שי מן אלמחסוס או מן אלמעקול פהו באטל,
פאמא בטלאן מא דפע אלמחסוס, פכמא קאל
<quote xml:lang="he">טורף נפשו באפו הלמענך תעזב ארץ ויעתק צור ממקומו</quote>
ואמא בטלאן מא דפע אלמעקול מן אלכד'ב ואלצדק, פקאל פיה
<quote xml:lang="he">ואם לא אפו מי יכזיבני וישם לאל מלתי</quote>
ת'ם ערפנא אן ג'מיע אלעלום מבניה עלי מא אדרכנאה בחואסנא אלתי ד'כרנאהא, ומתפרעה מנאה
ומסתכ'רג'ה ענאהא, אד' קאל
<quote xml:lang="he">שמעו חכמים מלי ויודעים האזינו לי כי און מלין תבחן וחידך יטעם לאכל</quote>
ת'ם חקק לנא אלכ'באר אלצאדקה אנהא חק בקולה
<quote xml:lang="he">אחוך שמע לי וזה חזיתי ואספרה אשר חכמים יגידו ולא כחדו מאבותם</quote>
להם לבדם נתנה הארץ ולא עבר זר בתוכם
ולאה שראיט שרחנאהא פי תפסיר הד'ה אלפואסיק פי מוצ'עהא
</p>
```

FIGURE 4. XML annotation of Hebrew quotations in *Kitāb al-Mukhtār fi al-Amānāt wa al-i'tiqādāt*, p. 16

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 fi al-Amānāt wa al-i'tiqādāt*:

אד' קאל إذ قال

כקולה כقوله

פכמא קאל فکما قال

פקאל פיה فقال فيه

בקולה بقوله

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:

```

</persName>ירמיהו</persName>
</distinct>'ע' אלס'</distinct>
</quote>'ה' ואתפלל אל ה'<quote xml:lang="he">
וצלואת
</persName>דניאל</persName>
ר
</persName>עזרא</persName>
</distinct>'ע' ג'מיע אלס' ואלרה'</distinct>
. מן אללה, ומא מאת'ל הד'א ושאבהה .
</p>

```

FIGURE 5. Distinct features in the book of *Kitāb al-Muḥāḍara wal-Muḍākara*, p. 20

(4-1)

'ע' אלס' = 'ע' אלס' (عليه السلام).

(4-2)

'ע' ג'מיע אלס' ואלרה' = 'ע' ג'מיע אלס' ואלרה' (عليهم جميع السلام والرحمه)

(4-3)

ואתפלל אל ה' = ואתפלל אל יהוה

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-s-salām* (peace upon him), and *alayhim gamī‘a-l-salāmu wa-l-r-raḥma* (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 יעקב אבן עזרא), "ספר העיונים והדיונים": ערך, הגיה ותרגום בצרוף הערות, אברהם שלמה הלמן, ירושלים, מקיצי נרדמים, תשל"ה

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

סעדיה בן יוסף פיומי גאון, ספר הנבחר באמונות ובדעות, כתאב אלמכ'תאר פי אלמנאנאת ואלאעתקאדאת, לרבנו סעדיה בן יוסף פיומי, מהד' י' קאפח, ירושלים, תשל"ל.

- Ahmed, Mohamed A. H. 2016. "Arabic Codes in Hebrew Texts: On the Typology of Literary Code-switching." *Journal of Jewish Languages* 4: 203-230.
- Bahat, Ya'aqov. 2002. *The Hebrew Component in the Written Arabic of the Jews of Morocco*. Jerusalem: Bialiq.
- Bar-Asher, Moshe. 2013. "Judeo-Arabic, North Africa, Hebrew Component." In *Encyclopedia of Hebrew Language and Linguistics* (Vol 2), ed. Geoffrey Khan. Leiden: Brill, 397.
- Bar-Asher, Moshe. 2010. "The Maghribī Sharḥ to tractate Avot", *Jerusalem Studies in Arabic and Islam* 37: 283-295.
- Bar-Asher, Moshe. 1992. *La composante hébraïque du judéo-arabe algérien*. Jerusalem: Magnes Press.
- Bassiouny, Reem. 2009. *Arabic sociolinguistics: topics in diglossia, gender, identity, and politics*. Washington, DC: Georgetown University Press.
- Berk-Seligson, Susan. 1986. "Linguistic Constraints on Intrasentential Code-Switching: A Study of Spanish/Hebrew Bilingualism." *Language in Society* 15.3: 313-348.
- Blau, Joshua. 1980. *A Grammar of Medieval Judaeo-Arabic*. Jerusalem: Magnes Press (in Hebrew).
- Blau, Joshua. 1981. *The Emergence and Linguistic Background of Judaeo-Arabic*. Jerusalem: Ben-Zvi Institute.
- Braunmüller, Kurt & Gisella Ferraresi, eds. 2003. *Aspects of Multilingualism in European Language History*. Amsterdam: John Benjamins.
- Chetrit, Joseph. 1991. "A socio-pragmatic study of the Hebrew component in the Judeo-Arabic poetry of North Africa: Textual aspects." *Massorot* 5–6.251–311. (In Hebrew).
- Clyne, Michael. 2003. *Dynamics of Language Contact: English and Immigrant Languages*. Cambridge: Cambridge University Press.
- De Brabanter, Philippe. 2004. "Foreign-Language Quotations and Code-Switching: the Grammar Behind." *ESSE Conference* (European Society for the Study of English. Saragosse, 2004.)
- DeRose, Steven. 1999. "XML and the TEI". *Computers and the Humanities* 33.1/2: 11 - 30.
- Field, Fredric. 2002. *Linguistic Borrowing in Bilingual Contexts*. Amsterdam, Philadelphia: John Benjamins.
- Dozy, Reinhart. 1881. *Supplement aux dictionnaires Arabes*. Leiden: Brill.
- Hary, Benjamin. 1992. *Multiglossia in Judeo-Arabic, with an Edition, Translation and Grammatical Study of the Cairene Purim Scroll*. Leiden: Brill.

- Hary, Benjamin. 1995. "Judaean-Arabic in its Sociolinguistic setting." In *Language and Culture in the Near East: Diglossia, Bilingualism, Registers*. eds, Izre'el Shlomo & Drory Rina. Leiden: Brill, 74–99.
- Hary, Benjamin. 1996. "Adaptations of Hebrew script." In *The world's writing systems*. eds, Peter Daniels & William Bright. New York: Oxford University Press.
- Hary, Benjamin. 1999. "Hebrew elements in Egyptian Judeo- Arabic texts." In *Vena Hebraica in*
- Hary, Benjamin. 2009. *Translating religion: Linguistic analysis of Judaean-Arabic sacred texts from Egypt*. Leiden, Boston: Brill.
- Hary, Benjamin. 2012. "Judaean-Arabic as a mixed language." In *Middle Arabic and mixed Arabic: Diachrony and synchrony*, eds. Liesbeth Zack & Arie Schippers. Boston: Brill, 125–143.
- Haugen, Einar. 1950. "The Analysis of Linguistic Borrowing." *Language* 26: 210- 31.
- Henkin-Roitfarb, Roni. 2011. "Hebrew and Arabic in Asymmetric Contact in Israel." *Lodz Papers in Pragmatics* 7:1.
- Isleem, Martin. 2014. "Hebrew codeswitching: the case of the Druze community in Israel." *International Journal of Applied Linguistics*. doi: 10.1111/ijal.12091
- Jastrow, Otto.1990. *Der arabische Dialekt der Juden von 'Aqra und Arbīl*. Wiesbaden: O. Harrassowitz.
- Jefferson, Judith & Ad Putter, eds. 2012. *Multilingualism in medieval Britain (c. 1066-1520): sources and analysis*. Turnhout: Brepols.
- Khan, Geoffrey. 2007. "Judaean-Arabic." In *Encyclopedia of Arabic Language and Linguistics*, eds. Lutz Edzard & Rudolf de Jong. Leiden: Brill, 526–536.
- Leslau, Wolf. 1945. "Hebrew Elements in the Judaean-Arabic Dialect of Fez." *The Jewish Quarterly Review, New Series* 36.1: 61-78.
- Lipski, John. 1977. "Code-switching and the problem of bilingual competence." In *The fourth LACUS forum*, M. Paradise (ed.). Columbia, SC: Hornbeam Press.
- Lipski, John. 2005. "Code-switching or Borrowing? No sé so no puedo decir, you know." In *Selected Proceedings of the Second Workshop on Spanish Sociolinguistics*, eds. Lotfi Sayahi & Maurice Westmoreland. Somerville, MA: Cascadilla Proceedings Project: 1-15.
- Mansour, Jacob. 1974. *The Judaean-Arabic dialect of Baghdad*. Haifa: University of Haifa.
- Meyers-Scotton, Carol. 1992. "Comparing codeswitching and borrowing". In *Codeswitching*. Carol Eastman. Clevedon: Multilingual Matters, 19-39.

- Miller, Elaine. 2000. *Jewish Multiglossia: Hebrew, Arabic, and Castilian in Medieval Spain*. Newark and Del: Juan de la Cuesta.
- Mitaib Murad, Tareq. 2013. "Arab Practicum Guides Code-switch to Hebrew: Attitudes, Factors and Reasons". *Journal of Language Teaching and Research* 4.6: 1160-1166.
- Morag, Shelomo, Moshe Bar-Asher & Maria Luisa Mayer-Modena, eds. 1999. *Judaeorum Linguis: Proceedings of the 2nd International Conference on the Hebrew and Aramaic Elements in Jewish Languages*. Milan: Università degli studi di Milano, 67–91.
- Muysken, Pieter. 2000. *Bilingual Speech: A Typology of Code-Mixing*. Cambridge: Cambridge University Press.
- Myers-Scotton, Carol. 1998. *Codes and Consequences: Choosing Linguistic Varieties*. New York: Oxford University Press.
- Nurmi, Arja & Päivi Pahta. 2004. "Social stratification and patterns of code-switching in early English letters." *Multilingua* 23: 417–456.
- Onysko, Alexander. 2012. *Anglicisms in German. Borrowing, Lexical Productivity, and Written Codeswitching*. Berlin: De Gruyter.
- Pahta, Paivi & Arja Nurmi. 2006. "Code-switching in the Helsinki Corpus: A thousand years of multilingual practices." In *Medieval English and its Heritage: Structure, Meaning and Mechanisms of Change*, eds. Nikolaus Ritt, Herbert Schendl, Christiane Dalton-Puffer & Dieter Kastovsky. Frankfurt: Peter Lang, 203–220.
- Poplack, Shana, David Sankoff & Christopher Miller. 1988. "The Social Correlates and Linguistic Process of lexical borrowing and Assimilation." *Linguistics* 26: 47-104.
- Poplack, Shana. 1980. "Sometimes I'll Start a Sentence in Spanish Y TERMINO EN ESPAÑOL": Toward a Typology of Code-switching." *Linguistics* 18(7/8): 581–618.
- Poplack, Shana. 1993. "Variation Theory and Language Contact." In *American Dialect Research*, ed. Dennis Preston. Amsterdam, Philadelphia: John Benjamins, 251-86.
- Poplack, Shana, David Sankoff & Christopher Miller. 2009. "The Social Correlates and Linguistic Processes of Lexical Borrowing And Assimilation." *Linguistics*, 26(1), pp. 47-104, online at doi:10.1515/ling.1988.26.1.47, retrieved Nov. 28, 2017,
- Schendl, Herbert. 1996. "Text types and code-switching in medieval and Early Modern English." *Vienna English Working Papers (VIEWS)* 5: 50-62.
- Schendl, Herbert. 1997. "To London fro Kent / Sunt preidia depopulantes: Code-switching and Medieval English Macaronic Poems." *Vienna English Working Papers (VIEWS)* 6: 52-66.

- Schendl, Herbert. 2001. "Code-switching in Medieval English Poetry." In *Language Contact in the History of English*, eds. Dieter Kastovsky & Arthur Mettinger. Frankfurt/M.: Lang, 305-335.
- Schendl, Herbert. 2002. "Code-choice and Code-switching in some early 15th Century Letters." In *Middle English from Tongue to Text*, eds. Peter Lucas & Angela Lucas. Frankfurt/M.: Lang, 247-262.
- Schendl, Herbert. 2005. "English Historical Code-switching in a European Perspective." In *Language and Language Practices* (Copenhagen Studies in Bilingualism, 36), eds. Christine Dabelsteen & Jens Normann Jrgensen. Copenhagen: University of Copenhagen, 194-208.
- Schendl, Herbert & Laura Wright, eds. 2011. *Code-switching in Early English*. Berlin, Boston: De Gruyter Mouton.
- Stillman, Norman. 2005. "The Judaeo-Arabic Heritage". In *Sephardic and Mizrahi Jewry: From the Golden Age of Spain to Modern Times*, ed. Zion Zohar. New York: New York University Press, 40-54.
- Szulmajster-Celnikier, Anne. 2005. "Code-switching in Yiddish: A Typology." *La Linguistique* 41. 2: 87-106.
- Tirosh-Becker, Ofra. 2012. "Mixed Linguistic Features in a Judeo-Arabic Text from Algeria: The Šarḥ to the Haftarot from Constantine." In *Language and Nature*, eds. Rebecca Hasselbach & Na'ama Pat-El, Chicago: Oriental Institute of the University of Chicago, 391-406.
- Wagner, Esther-Miriam & Magdalen Connolly. 2017. "Code-switching in Judaeo-Arabic Documents from the Cairo Geniza." *Multilingua*, 0.0 (2017), doi:10.1515/multi-2017-0050, retrieved Dec. 4, 2017,.
- Wagner, Esther-Miriam. 2010. *Linguistic Variety of Judaeo-Arabic in Letters from the Cairo genizah*. Leiden: Brill.
- Wittern, Christian, Arianna Ciula, & Conal Tuohy. 2009. "The making of TEI P5." *Lit Linguist Computing* 24.3: 281-296.