TTS Survey Report

Enhancing Arabic Speech Technology

with the world's most comprehensive Arabic lexicon

Abstract

The recent advances in deep learning and neural network technology have dramatically improved speech technology [5]. Although the quality of Arabic speech technology has been steadily improving, it still lags significantly behind the other major languages, such as Chinese and English. CJKI's **Comprehensive Arabic Full Form Lexicon,** or *ArabLEX* (**1.2 billion entries**) [1, 2], is the most comprehensive Arabic computational lexicon ever created. This reports analyzes the significant error rates in he leading TTS systems, and shows how *ArabLEX* can address these shortcomings by dramatically enhancing the quality of both TTS and ASR.

1. Orthographical ambiguity

One reason that Arabic speech technology lags behind is that the Arabic script is highly ambiguous. Words are often written as a string of consonants with no indication of vowels. For example, كاتب can represent as many as *seven* pronunciations: *kāatib*, *kātibun*, *kātibin*, *kātaba*, *kātibi*, *kātiba* and *kātibu*. Many other characteristics of the Arabic script contribute to a high level of orthographic ambiguity, as described in Halpern's paper on Arabic named entities [4].

The morphological complexity of such cliticized forms as ولكاتباتهما *walikaṯibātíhimaṯ*, and the absence of vowel diacritics, makes Arabic TTS especially challenging. That is, determining the morphological composition of such forms, and the correct vowels for such consonants as ت in ولكاتباتهما often requires morphological, semantic and contextual analysis which tax the capabilities of state-of-the-art speech technology.

2. Improving TTS accuracy

The extreme orthographic ambiguity of Arabic has led to unacceptably high error rates, even by the TTS systems offered by major players such as Google, Apple and Microsoft.

Our institute has conducted a survey to determine the scope of this problem, some of the results of which are reported in the Appendix. Surprisingly, we discovered that it is not unusual for over 50%, and even 80%, of the words in a sentence to be mispronounced, and that there is a trend for cliticized words to be incorrectly pronounced. For example, the cliticized word وَلِلْكَاتِبِينَ, correctly pronounced *walilkạtibīna,* is mispronounced as *walilkātibáyna.*

As can be seen from the Appendix, the error rate of Arabic TTS is unacceptably high. Such a high error rate would be unthinkable in the other major world languages. Another issue is **prosody** (stress and intonation) and **vowel neutralization** (e.g. نا *nạ* is written as a long vowel in أنا but is shortened in actual pronunciation to *na*). This is a complex issue, described in detail in Halpern's paper on Arabic stress [3].

Speech synthesis, speech recognition and prosody in current Arabic speech technology are, on the whole, inaccurate, unnatural and often unpleasant to the ear. The time has come for developers to make serious efforts to make dramatic improvements.

3. Improving ASR accuracy

ArabLEX includes features specifically designed to support automatic speech recognition (ASR). For speech synthesis (TTS), it is only necessary to *generate* one accurate pronunciation. For example, كاتبون 'writers' in standard Arabic is pronounced *kạtibūna,* but for ASR it is also necessary to *recognize* the less formal variant pronunciation *kạtibūn* Similarly, the standard pronunciation of أكتب 'I write' is *ʾáktubu,* but the final vowel is often omitted and it is pronounced *ʾáktub.*

The above alternatives are on a *phonemic* level. That is, the phoneme /na/ is being replaced by the phoneme /n/ as a result of vowel omission. There are also variations on the *phonetic* level; that is, certain phonemes have regional allophones. For example, ج in such words as جمل *jamal* is pronounced [gɛ̈mɛl] in Egypt, [ʤɛ̈mɛ̈l] in the Gulf region, and [ʒɛ̈mɛ̈l] in the Levant. It is important top note that this does not refer to the local dialects in those regions, but to a *regional varieties* of Modern Standard Arabic (MSA).

Thus *ArabLEX* not only represents ج in the standard IPA [d͡ʒ] for TTS, it also lists the regional [ʒ] and [g] for ASR training. The goal is to enable the recognition of these allophones, but not to generate them.

4. Benefits to speech technology

One of the key components for training speech technology systems is the *pronunciation dictionary.* A major feature of *ArabLEX* is that it can serve as an extremely comprehensive pronunciation dictionary.

ArabLEX not only maps all unvocalized forms (including all cliticized forms) to their vocalized counterparts and to their lemmas, but also provides precise **phonemic transcriptions** (CARS system) [4] and **phonetic transcriptions** (IPA) that includes precise word stress and vowel neutralization for each entry. For example, in the IPA *wɛ̈likɛ̈ːˈtibikumɛˈ*), the stressed syllable is indicated by (ˈ) (U+0C28), while (ˈ) (U+02D1) indicates that the final ɛ is neutralized vowel of optional half length. These features can help developers significantly enhance the quality of Arabic TTS, and can be used in training ASR systems to achieve higher recognition rates.

To summarize, *ArabLEX* can bring the following benefits to speech technology:

- Covers approximately **1.2 billion entries,** including millions of proper nouns.
- Covers *all combinations* of proclitics and enclitics for inflected wordforms (mostly verbs, nouns, adjectives and proper nouns).
- Tens of millions of orthographic variants for all wordforms.
- Provides an exhaustive list of alternative pronunciations of identical unvocalized strings to enable orthographical disambiguation (e.g. six alternatives for كاتباتك).
- Future versions will provide 'importance flags' to help determine the most likely alternative.
- Highly accurate phonemic transcriptions for all wordforms, including precise stress and vowel neutralization
- Phonetic transcriptions (IPA) indicate the correct allophonic variants in context as well as regional variants for ASR.

References

[1] Halpern, Jack. Arablex: Comprehensive Arabic Full Form Lexicon

[2] Halpern, Jack. Very Large-Scale Lexical Resources to Enhance Chinese and Japanese Machine Translation

[3] Halpern, Jack. Word Stress and Vowel Neutralization in Modern Standard Arabic

[4] Halpern, Jack. CJKI Arabic Romanization System

[5] Wajdan Algihab, Imam Muhammad bin Saud Islamic University. Arabic Speech Recognition with Deep Learning: A Review

APPENDIX: TTS SURVEY RESULTS

Below are the results of a survey conducted by CJKI (our institute) to compare the TTS systems of Google, Apple (iPhone) and Microsoft (Bing), showing high error rates for all three. The **Unvocalized** field is the original Arabic text, the **Vocalized** field indicates the correct pronunciation, and the **CJKI** field shows the correct pronunciation in CARS phonemic transcription [4]. The CARS transcriptions in these **Google, iOS** and **Bing** columns indicate the actual pronunciation by the three TTS engines. Mispronunciations are indicated in red, and the error rate is given in the column headers.

Table 1 and 3 are based on text c**omposed** for this survey, while tables 2 and 4 use a sentence **extracted** from the web. (It is noteworthy that the error rate for the composed text is actually much lower than for the extracted text.) Tables 1 and 3 compare the results on a word-by-word basis, whereas tables 2 and 4 compare them on a sentence-by-sentence basis, showing the context. The fact that the error rate is sometimes over 80% is surprising and unacceptable to users.

Table 1: Mispronounced Words in Composed Text

Unvocalized	Vocalized	Google (13%)	iOS (31%)	Bing (25%)	CJKI (0%)
عدد	عَدَّدَ	ɛádadu	ɛádada	ɛádada	ɛáddada
الكاتب	ٱلْكَاتِبُ	lkātibu	lkātibi	lkātibu	lkātibu
ما	مَا	mạ	mạ	mạ	mạ
قال	قَالَ	qāla	qāla	qāla	qāla
إن	إِنَّ	ʾínna	ʾínna	ʾínna	ʾínna
هؤلاء	هٰؤُلَاءِ	hạʾulāʾi	hạʾulāʾi	hạʾulāʾi	hạʾulāʾi
الحكام	ٱلْحُكَّامَ	lḥukkāmi	lḥukkāmi	lḥukkāmi	lḥukkāma
يفعلونه	يَفْعَلُونَهُ	yafɛalūnahu	yafɛalūnahu	yafɛalūnahu	yafɛalūnahu
في	فِي	fị	fị	fị	fị
الخارج	ٱلْخَارِجِ	lkhāriji	lkhārija	lkhāriji	lkhāriji

مثل	مِثْلَ	míthli	míthli	míthli	míthla
الهجمات	ٱلْهَجَمَاتِ	lhajamāti	lhajamāti	lhajamāti	lhajamāti
الإلكترونية	ٱلْإِلِكْتُرُونِيَّةِ	l’ilikturụníyyati	l’ilikturụníyyati	l’ilikturụníyyati	l’ilikturụníyyati
ومطاردة	وَمُطَارَدَةِ	wamuṭārádati	wamuṭārídati	wamuṭārídati	wamuṭārádati
المعارضين	ٱلْمُعَارِضِينَ	lmuɛạriḍīna	lmuɛạriḍīna	lmuɛạriḍīna	lmuɛạriḍīna
اللاجئين	ٱللَّاجِئِينَ	llạjiʾīna	llạjiʾīna	llạjiʾīna	llạjiʾīna
في	فِي	fị	fị	fị	fị
العواصم	ٱلْعَوَاصِمِ	lɛawāṣimi	lɛawāṣimi	lɛawāṣimi	lɛawāṣimi
الغربية	ٱلْغَرْبِيَّةِ	lgharbíyyati	lgharbíyyati	lgharbíyyati	lgharbíyyati
وللكاتبين	وَلِلْكَاتِبِينَ	walilkạtibīna	walilkātibáyna	walilkātibáyna	walilkạtibīna
من	مِنَ	mína	mína	mína	mína
الصحفيين	ٱلصَّحَفِيِّينَ	ṣṣaḥafịyīna	ṣṣaḥafịyīna	ṣṣaḥafịyīna	ṣṣaḥafịyīna
العرب	ٱلْعَرَبِ	lɛárabi	lɛárabi	lɛárabi	lɛárabi
صرح	صَرَّحَ	ṣárraḥa	ṣáraḥa	ṣáraḥa	ṣárraḥa
بأن	بِأَنَّ	biʾánna	biʾánna	biʾánna	biʾánna
عليهم	عَلَيْهِمْ	ɛaláyhim	ɛaláyhim	ɛaláyhim	ɛaláyhim
أن	أَنْ	ʾan	ʾan	ʾan	ʾan
يكتبوا	يَكْتُبُوا	yaktúbuwu	yaktúbuwu	yaktúbuwu	yaktúbuwu
ما	مَا	mạ	mạ	mạ	mạ
تمليه	تُمْلِيهِ	tumallīhi	tamlīhi	tamlīhi	tumlīhi
عليهم	عَلَيْهِمْ	ɛalayhim	ɛalayhim	ɛalayhim	ɛalayhim
ضمائرهم	ضَمَائِرُهُمْ	ḍamāʾíruhum	ḍamāʾírihim	ḍamāʾírihim	ḍamāʾíruhum

Table 2: Mispronounced Words in Extracted Text

Unvocalized	Vocalized	Google (80%)	iOS (90%)	Bing (70%)	CJKI (0%)
الاخوات	ٱلْأَخَوَاتُ	ʾalikhwātu	ʾalʾakhawāti	ʾalʾakhawātu	ʾalʾakhawātu
المتزوجات	ٱلْمُتَزَوِّجَاتُ	lmutazawwijātu	lmutazawwijāti	lmutazawwijāti	lmutazawwijātu
اللاتي	ٱللَّاتِي	lṭi	llaṭi	llāṭi	llāṭi
رزقن	رُزِقْنَ	rízqin	rúzqin	rúzqin	ruzíqna
بابناء	بِأَبْنَاءَ	bạbināʾun	bibnāʾi	bibnāʾi	biʾabnāʾa
فليكتبن	فَلْيَكْتُبْنَ	falayiktíbna	falktíbna	falktíbna	falyaktúbna
اسمائهم	أَسْمَائَهُمْ	ʾismāʾahum	smāʾihim	smāʾihim	ʾasmāʾahum
وسبب	وَسَبَبَ	wasábaba	wasábaba	wasábaba	wasábaba
التسميه	ٱلتَّسْمِيَةِ	lttasammīhu	lttasammīhi	lttasammīhi	ttasmíyati
رجاءا	رَجَاءًا	rajjāʾan	rajāʾ	rajāʾ	rajāʾan

Table 3: Mispronounced Sentences in Composed Text

TTS	Sentence	Error %
Unvocalized	عـدد الكـاتب مـا قـال إن هـؤلاء الحكـام يفعلـونه فـي الخـارج مثـل الهجمـات الإلكترونيـة ومطـاردة المعارضـين اللاجئيـن فـي العواصـم الغربية. وللكاتبين من الصحفيين العرب صرح بأن عليهم أن يكتبوا ما تمليه عليهم ضمائرهم	-
Vocalized	عَدَّدَ ٱلْكَـاتِبُ مَـا قَـالَ إِنَّ هٰـؤُلَاءِ ٱلْحُكَّـامَ يَفْعَلُـونَهُ فِـي ٱلْخَـارِجِ مِثْـلَ ٱلْهَجَمَاتِ ٱلْإِلِكْتُرُونِيَّـةِ وَمُطَـارَدَةِ ٱلْمُعَارِضِـينَ ٱللَّاجِئِيـنَ فِـي ٱلْعَوَاصِـمِ ٱلْغَرْبِيَّةِ. وَلِلْكَاتِبِينَ مِنَ ٱلصَّحَفِيِّينَ ٱلْعَرَبِ صَرَّحَ بِـأَنَّ عَلَيْهِـمْ أَنْ يَكْتُبُـوُا مَا تُمْلِيهِ عَلَيْهِمْ ضَمَائِرُهُمْ	0%
CJKI	ɛáddada lkātibu mạ qāla ʾínna hạʾulāʾi lḥukkāma yafɛalūnahu fi lkhāriji míthla lhajamāti lʾilikturụníyyati wamuṭārádati lmuɛạriḍīna llạjiʾīna fi lɛawāṣimi lgharbíyyati. walilkạtibīna mína ṣṣaḥafiyīna lɛárabi ṣárraḥa biʾánna	0%

	ɛaláyhim ʾan yaktúbuwu mạ tumlīhi ɛalayhim ḍamāʾíruhum	
Google	ɛádadu lkātibu mạ qāla ʾínna hạ̱ʾulāʾi lḥukkāmi yafɛalūnahu fi̱ lkhāriji míthli lhajamāti l’ilikturụníyyati wamuṭārádati lmuɛạriḍīna llạ̱jiʾīna fi̱ lɛawāṣimi lgharbíyyati. walilkạtibīna mína ṣṣaḥafi̱yīna lɛárabi ṣárraḥa biʾánna ɛaláyhim ʾan yaktúbuwu mạ tumalīhi ɛalayhim ḍamāʾíruhum	13%
iOS	ɛádada lkātibi mạ qāla ʾínna hạ̱ʾulāʾi lḥukkāmi yafɛalūnahu fi̱ lkhārija míthli lhajamāti l’ilikturụníyyati wamuṭārídati lmuɛạriḍīna llạ̱jiʾīna fi̱ lɛawāṣimi lgharbíyyati. walilkātibáyna mína ṣṣaḥafi̱yīna lɛárabi ṣáraḥa biʾánna ɛaláyhim ʾan yaktúbuwu mạ tamlīhi ɛalayhim ḍamāʾírihim	31%
Bing	ɛádada lkātibu mạ qāla ʾínna hạ̱ʾulāʾi lḥukkāmi afɛalūnahu fi̱ lkhāriji míthli lhajamāti l’ilikturụníyyati wamuṭārídati lmuɛạriḍīna llạ̱jiʾīna fi̱ lɛawāṣimi lgharbíyyati. walilkātibáyna mína ṣṣaḥafi̱yīna lɛárabi ṣáraḥa biʾánna ɛaláyhim ʾan yaktúbuwu mạ tamlīhi ɛalayhim ḍamāʾírihim	25%

Table 4: Mispronounced Sentences in Extracted Text

TTS	Sentence	Error %
Unvocalized	الاخوات المتزوجات اللاتي رزقن بابناء فليكتبن اسمائهم وسبب التسميه رجاءا	-
Vocalized	اَلْأَخَوَاتُ ٱلْمُتَزَوِّجَاتُ ٱللَّاتِي رُزِقْنَ بِأَبْنَاءَ فَلْيَكْتُبْنَ أَسْمَائَهُمْ وَسَبَبَ ٱلتَّسْمِيَةِ رَجَاءًا	0%
CJKI	ʾalʾakhawātu lmutazawwijātu llāti̱ ruzíqna biʾabnāʾa falyaktúbna ʾasmāʾahum wasábaba ttasmíyati rajāʾan	0%
Google	ʾalikhwātu lmutazawwijātu lti̱ rízqin bạbināʾun falayiktíbna ʾismāʾahum wasábaba lttasammīhu rajjāʾan	\80%
iOS	ʾalʾakhawāti lmutazawwijāti llati̱ rúzqin bibnāʾi falktíbna smāʾihim wasábaba lttasammīhi rajāʾ	90%
Bing	ʾalʾakhawātu lmutazawwijāti llāti̱ rúzqin bibnāʾi falktíbna	70%