Adversarial Examples in Arabic - Supplementary Material

Basemah Alshemali
College of Computer Science and Engineering
Taibah University
Almadinah, KSA
College of Engineering and Applied Science
University of Colorado at Colorado Springs
Colorado Springs, USA
balshema@uccs.edu

Jugal Kalita
College of Engineering and Applied Science
University of Colorado at Colorado Springs
Colorado Springs, USA
jkalita@uccs.edu

I. Background: Arabic Language

Arabic is a Semitic language spoken by more than 420 million people in 22 countries[1]. It is a highly structured language where morphology plays a very important role. Arabic shares some special features along with other Semitic languages such as[2] : (1) It is written from right to left; (2) There is no capitalization; and (3) Letters change their shape according to their position in the word.

A. Adjectives in Arabic

Unlike in English, adjectives in Arabic follow the nouns they modify and agree with them in gender, number, and definiteness. For example:
(1) بنت جميلة, (βɪnt ʤaːmiːlh, girl pretty-singular-feminine-indefinite).
(2) البنت الجميلة, (ɑlbɪnt alʤaːmiːlh, the girl pretty-singular-feminine-definite).
The adjective جميلة (ʤaːmiːlh, pretty-singular-feminine-indefinite) in (1) and الجميلة (alʤaːmiːlh, pretty-singular-feminine-definite) in (2) both agree with the preceding nouns بنت (βɪnt, girl-indefinite) in (1) and البنت (ɑlbɪnt, girl-definite) in (2) respectively in number (singular), gender (feminine), and definiteness.

1) Definite Adjectives in Arabic: To form a definite adjective from a singular, dual, or plural masculine or feminine adjective, one can attach ال (al) to the *beginning* of the indefinite adjective. For example:
(1) كبير, (kaːbiːr, Big-singular-masculine-indefinite).
(2) الكبير, (alkaːbiːr, Big-singular-masculine-definite).
(3) كبيرات, (kaːbiːraːt, Big-plural-feminine-indefinite).
(4) الكبيرات, (alkaːbiːraːt, Big-plural-feminine-definite).

2) Feminine Adjectives in Arabic: To form a singular feminine adjective from a singular masculine adjective, one can add ة, sounds (t) or (h), to the *end* of the adjective. For example:
(1) كبير, (kaːbiːr, Big-singular-masculine-indefinite).
(2) كبيرة, (kaːbiːrh, Big-singular-feminine-indefinite).

3) Dual Adjectives in Arabic: Unlike English, which has only two numbers, singular and plural, Arabic has a third number called dual. It is applied to situations where there are two individuals being referred to, as a group. To form a dual masculine adjective from a singular masculine, one can add ان (aːn) or ين (ajn) to the end of the adjective. For instance:
(1) كبير, (kaːbiːr, Big-singular-masculine-indefinite).
(2) كبيران, (kaːbiːraːn, Big-dual-masculine-indefinite).

On the other hand, a dual feminine adjective can be formed by adding تان (taːn) or تين (tajn) to the end of a singular masculine adjective:
(1) كبير, (kaːbiːr, Big-singular-masculine-indefinite).
(2) كبيرتان, (kaːbiːrtaːn, Big-dual-feminine-indefinite).

4) Plural Adjectives in Arabic: To form a plural masculine adjective, one can add ون (oːn) or ين (iːn) to the end of the singular masculine adjective. For instance:
(1) كبير, (kaːbiːr, Big-singular-masculine-indefinite).
(2) كبيرون, (kaːbiːroːn, Big-plural-masculine-indefinite).

On the other hand, a plural feminine adjective can be formed by adding ات (aːt) to the end of the singular masculine adjective:
(1) كبير, (kaːbiːr, Big-singular-masculine-indefinite).
(2) كبيرات, (kaːbiːraːt, Big-plural-feminine-indefinite).

Note that the rules mentioned in Sections I-A1, I-A2, I-A3, and I-A4 are not always applicable, as there are some exceptions[3]. Due to space constraints, we leave investigating these exceptions to the reader.

[1] http://www.worldpopdata.org/

[2] https://dl.acm.org/citation.cfm?id=1644881

[3] https://www.iasj.net/iasj?func=fulltext&aId=52881

II. MADAMIRA Arabic Morphological Analyzer

MADAMIRA[4] is a morphological analyzer for Arabic. Short reviews from the HARD and BRAD corpora were analyzed by MADAMIRA and relevant results are presented in Table II and Table III.

TABLE I
The morphological features provided by MADAMIRA and their descriptions.

Feature	Description
diac	word with diacritic markers
lex	lexicon
asp	aspect
cas	case
enc	enclitic values
gen	gender
mod	mood
num	number
per	person
pos	part-of-speech
prc	proclitic values
stt	state
vox	voice
gloss	English gloss
stem	stem
source	source of the word analysis

TABLE II
The morphological features provided by MADAMIRA for the sentence فندق ممتاز (IPA: funduq mumtaːz, English: excellent hotel, word by word: hotel excellent-singular-masculine-indefinite). Descriptions of the features provided in Table I.

Feature	فندق (Hotel)	ممتاز (Excellent)
diac	فُندُق	مُمتاز
lex	فندق	ممتاز
asp	na	na
cas	n	n
enc	0	0
gen	m	m
mod	na	na
num	s	s
per	na	na
pos	noun	adj
prc	0	0
stt	i	i
vox	na	na
gloss	hotel	excellent; superior
stem	فندق	ممتاز
source	lex	lex

TABLE III
The morphological features provided by MADAMIRA for the sentence رواية سيئة (IPA: riwaːjaːh saːjah, English: Bad novel, word by word: novel bad-singular-feminine-indefinite). Descriptions of the features provided in Table I.

Feature	رواية (Novel)	سيئة (Bad)
diac	روايَةٌ	سَيِئَةٌ
lex	رَواية	سيئ
asp	na	na
cas	n	n
enc	0	0
gen	f	f
mod	na	na
num	s	s
per	na	na
pos	noun	adj
prc	0	0
stt	i	i
vox	na	na
gloss	novel; story	bad
stem	رواي	سيئ
source	lex	lex

III. Samples of Adversarial Examples

Samples of the adversarial examples produced by the Definite attack and the Gender attack are shown in Tables IV and V.

TABLE IV
Example of Definite attack result from BRAD corpus. One word has been modified from an indefinite adjective (**سيئة**, **saːjah**) to a definite adjective (**السيئة**, **alsaːjah**). The modified word is highlighted in red in the original and adversarial text.

Prediction	Text
2 (negative)	**Original Text:** رواية سيئة جدا وفقيرة في معلوماتها. الأجزاء مملة جدا وحتى الأجزاء العميقة فيها أفكار سطحية ومعلومات مغلوطة. باختصار الرواية سطحية بائسة.
	ɾiwaːjaːh saːjah ʤɪdaːn wːfaqiːrah fj malumatha. alajzaː mumilah ʤɪdaːn wːħtaː alajzaː alamiːqah fiːha afkaːr satħiaːh wːmaːlumat maɣlutah. bixtisaːr alriwaːjaːh satħiaːh baːjisah .
	Bad and poor-quality content novel. The parts are very boring even the deep parts have shallow thoughts and misinformation. In short, the novel is shallow and miserable.
4 (positive)	**Adversarial Text:** رواية السيئة جدا وفقيرة في معلوماتها. الأجزاء مملة جدا وحتى الأجزاء العميقة فيها أفكار سطحية ومعلومات مغلوطة. باختصار الرواية سطحية بائسة.
	ɾiwaːjaːh alsaːjah ʤɪdaːn wːfaqiːrah fj malumatha. alajzaː mumilah ʤɪdaːn wːħtaː alajzaː alamiːqah fiːha afkaːr satħiaːh wːmaːlumat maɣlutah. bixtisaːr alriwaːjaːh satħiaːh baːjisah .

[4] https://camel.abudhabi.nyu.edu/madamira/

TABLE V

Example of Gender attack result from the HARD corpus. One word has been modified from a singular masculine adjective (**ممتاز**, **mumtaːz**) to a singular feminine adjective (**ممتازة**, **mumtaːzh**). The modified word is highlighted in red in the original and adversarial text.

Prediction	**Text**
4 (positive)	**Original Text:** فندق ممتاز. موقع الفندق ونظافته مرضية، والاطلالة على البحر وجزيرة النور التي تحتوي على الفراشات ساحرة. قريب من المسجد والممشى وتتوفر المواقف وخدمة صف السيارات بجانب الفندق لكن المسبح كان مغلق للصيانة والمواقف تحتاج الى مظلات.
	funduq mumtaːz. mawqi alfunduq wːnðafatuh murdiah, wːalatlaːlh ala albħr wːʤazirat alnuːr saːhirah. qariːb mn almasʤiːd wːalmamʃa wːtatawafar almawaqif wːxidmat saːf alsajaraːt biʤanib alfunduq lakiːn almasbaħ kaːn muɣlaq lilsjaːnh wːalmawaqif taħtaʤ elaː maðlaːt.
	Excellent hotel. The hotel's location and its cleanliness are satisfactory, the view and the island of light are charming. Close to the mosque and walkway, parking and valet parking are available next to the hotel, but the swimming pool was closed for maintenance and parking spots needed cover.
2 (negative)	**Adversarial Text:** فندق ممتازة. موقع الفندق ونظافته مرضية، والاطلالة على البحر وجزيرة النور التي تحتوي على الفراشات ساحرة. قريب من المسجد والممشى وتتوفر المواقف وخدمة صف السيارات بجانب الفندق لكن المسبح كان مغلق للصيانة والمواقف تحتاج الى مظلات.
	funduq mumtaːzh. mawqi alfunduq wːnðafatuh murdiah, wːalatlaːlh ala albħr wːʤazirat alnuːr saːhirah. qariːb mn almasʤiːd wːalmamʃa wːtatawafar almawaqif wːxidmat saːf alsajaraːt biʤanib alfunduq lakiːn almasbaħ kaːn muɣlaq lilsjaːnh wːalmawaqif taħtaʤ elaː maðlaːt.