



Nâbra

Syrian Arabic Dialects with Morphological Annotations

Amal Nayouf

Virtual University
Syria

Tymaa Hammouda

Birzeit University
Palestine

Mustafa Jarrar

Birzeit University
Palestine

Fadi Zaraket

AUB
Lebanon

Bassam Kurdy

Virtual University
Syria



We build
tools and
resources
for NLU

Lexical Resources at SinaLab - Birzeit University

Lexicographic Database



150 lexicons
Largest Arabic lexicographic database

Arabic Ontology/ Wordnet



Formal Arabic Wordnet
with ontologically clean content

Annotated Corpora



Dialects,
NER, WSD, synonyms
Intents, hate
....

NLP library



APIs
Linguistic Data,
synonyms, Nested
NER, intents, ...

Synonyms 90s%

WSD 84%

NER 90s%

Intent 88.4%

Offensive 88.4%

Big Linguistic Data Graph

<https://ontology.birzeit.edu>



Natural Language Understanding Tools and Datasets



Open Source

<https://sina.birzeit.edu/resources>



SinaLab

News Team Resources

Resources

Download and try NLP/NLU datasets, corpora, tools and services

+ Lexicographic Database (150 lexicons)

حوسبة المعاجم

+ Arabic Ontology

الأنطولوجيا العربية

+ Dialect Corpora (Currasat)

كراسات مدونة العاميات

+ Arabic Synonyms

استخراج مترادفات

+ Named Entity Recognition (Wojood)

وجود - لاستخراج أسماء الاعلام

+ Word Sense Disambiguation (Salma)

سلمى - محلل دلالي

+ ArBanking77 Intent Detection

تحديد المقصود في المساعدات الآلية

+ Offensive Language Detection

خطاب الكراهية بالعبرية

+ Lemmatizer

+ NLP Tools



Copyright © 2023 Birzeit University

Nabra: Syrian Arabic Dialects with Morphological Annotations

Amal Nayouf

Syrian Virtual University, Syria
amal_124724@svuonline.org

Tymaa Hasanain Hammouda

Birzeit University, Palestine
thammouda@birzeit.edu

Mustafa Jarrar

Birzeit University, Palestine
mjarrar@birzeit.edu

Fadi A. Zaraket, zfadi@utexas.edu

Doha Institute for Graduate Studies, Doha
American University of Beirut, Beirut

Mohamad-Bassam Kurdy

Syrian Virtual University, Syria
t_bkurdy@svuonline.org

Abstract

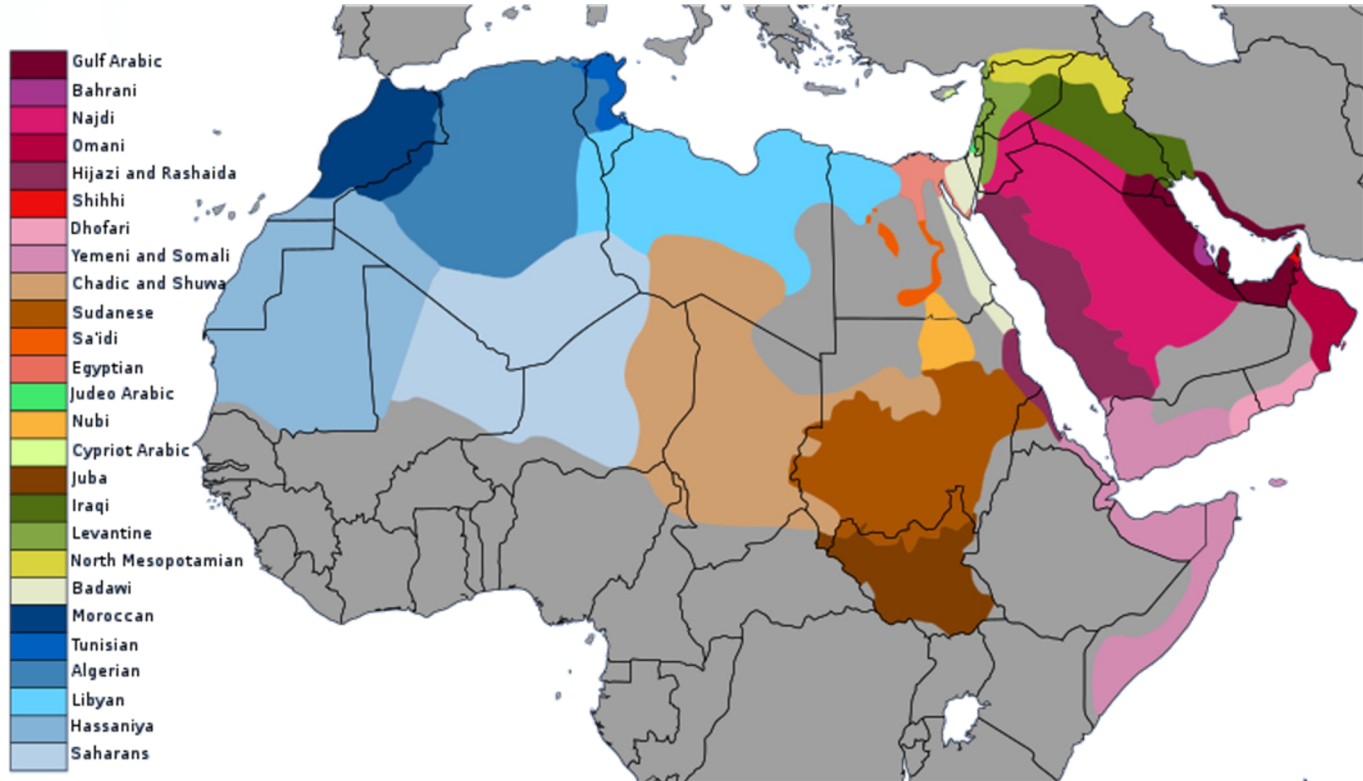
This paper presents Nābrā (نَبْرَة), a corpora of Syrian Arabic dialects with morphological annotations. A team of Syrian natives collected more than 6K sentences containing about 60K words from several sources includ-

ع/أداة مضارعة + ب/أداة مضارعة Particles for continuous present tense	عجكي I am speaking
ع/أداة مضارعة + ب/أداة مضارعة + ت/المضارع المخاطب المؤنث المفرد Feminine present second	عجتغاي You are being proud
موأداة نفى Negation particle	مومشان Not because

Nayouf, A., Hammouda, T., Jarrar, M., Zaraket, F., Kurdy, M.: [Nābrā: Syrian Arabic Dialects with Morphological Annotations](#). In Proceedings of the 1st Arabic Natural Language Processing Conference (ArabicNLP), Part of the EMNLP 2023. ACL.

Arabic is a low resources language

- Classical Arabic
- Modern Standard Arabic (more resources)
- **Arabic Dialects**



Seven morphologically-annotated Arabic dialect corpora (1.35 million tokens)

Curras2 كُرَاس : **Palestinian** dialect corpus (56K tokens)

Baladi بَلَدِي : **Lebanese** dialect corpus (10K tokens)

Nabra نَبْرَة : **Syrian** dialect corpus (60K tokens)

Lisan لِسَان : **Yemeni, Iraqi, Libyan, and Sudanese** dialects corpora

Yemeni (1.2 million), Iraqi (46K tokens), Libyan(52K tokens), Sudanese(53K tokens)



- El Haff, K., Jarrar, M., Hammouda, T., Zaraket, F., (2022). **Curras + Baladi: Towards a Levantine Corpus**. In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2022), Marseille, France. **Best Paper Award**
- Amal Nayouf, Tymaa Hammouda, Mustafa Jarrar, Fadi zaraket, Mohamad-Bassam Kurdy: **Nâbra: Syrian Arabic Dialects with Morphological Annotations**. In Proceedings of the 1st Arabic Natural Language Processing Conference (ArabicNLP), Part of the EMNLP 2023. ACL.
- Mustafa Jarrar, Fadi Zaraket, Tymaa Hammouda, Daanish Masood, Martin Waehlich: **Lisan: Yemeni, Iraqi, Libyan, and Sudanese Arabic Dialect Corpora with Morphological Annotations**. The 20th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA). Pages(-). IEEE. Egypt. 2023



هيك

Search

ع | EN

- Word Stem Lemma Gloss
 Whole Word Substring
 Palestinian Lebanese Iraqi Libyan Sudanese Yemeni

About Publications

313 results (4.6 secs)

Gloss	Lemma	POS	Suffix	Stem	Prefix	Word	Context
this way, this, thus	هيك / هيكذا	ضمير اشارة		هيك	و/عطف	وهيك	... عشرين سنين سنينسي ووهيك سي ينسبر مجموع ناسو ها
this way, this, thus	هيك / هيكذا	ضمير اشارة		هيك	و/عطف	وهيك	اللي اكلناه , عزمونا الاسبوع الماضي , عشان نردلهم العزيمة , وهيك بك تحطيلهم ترترمي
this way, this, thus	هيك / هيكذا	ضمير اشارة		هيك	و/عطف	وهيك	وهي قاطعا بحقلة قمح , صارت عجيبة خلت القمحات يطول حشيشن وهيك قدرت تتحبي بيناتن , ومن هون صرنا نعمل القمحية بليامنا .
this way, this, thus	هيك / هيكذا	ضمير اشارة		هيك	و/عطف	وهيك	بيفتوا : " بلاطة فوق بلاطة صاحبة البيت ضراطة ! " وهيك متكون وصلنا لآخر جولتنا بعالم عيد البربارة .
inon+like_that;thus	هيك	تعجب		هيك	ب/حرف جر	بهيك	يمشون وياه الناس ولاكن بيهم شي اشلون انذر هم يمتحنون بهيك مكان ياوزير ياعر
inon+like_that;thus	هيك	تعجب		هيك	ب/حرف جر	بهيك	يجوز ياخذج يريد يخرع بهيك جهاله
this way, like this, this, thus	هيك	ضمير اشارة		هيك		هيك	وين يعرف يسولف ويدبر هذا الحجي اكيد واحد كايلاه احجي بهيك
inon+like_that;thus	هيك	تعجب		هيك	ب/حرف جر	بهيك	ولله حيرتو العالم ولعبتو بيها طوبه كل بهيك اينزل كلام اليعجه وينشر

313 results (4.6 secs)

Gloss	Lemma	POS	Suffix	Stem	Prefix	Word	Context
be written/be fated/be destined write	كُتِبَ	فعل مضارع مفرد، مذكر، متكلم		كتب	و/عطف+/المضارع المتكلم المفرد	واكتب	وراح اكون هدية واكتب وصية
be written/be fated/be destined write	كُتِبَ	فعل مضارع مفرد، مذكر، مخاطب		كتب	و/عطف+ب/إداة مضارع+ت/المضارع المخاطب المذكر المفرد	ويكتبك	وتقد على القيس ويكتبك
be written/be fated/be destined write	كُتِبَ	فعل ماضي مفرد، مذكر، مخاطب	ت/الماضي: فاعله مخاطب مذكر مفرد	كتب	و/عطف	وكتبت	وكتبت بوصيتك هو السبب
be written/be fated/be destined write	كُتِبَ	فعل مضارع مفرد، مذكر، غائب	ل/حرف جر+ها/ضمير متصل للغائب المؤنث المفرد	كتب	و/عطف+ي/المضارع الغائب المذكر المفرد	ويكتبها	بس شوفت انو قبل كم يوم كان يحطها اغانى ويحكي عنها ويكتبها . . . يا الله كيف
be written/be fated/be destined write	كُتِبَ	فعل مضارع مفرد، مذكر، غائب		كتب	ي/المضارع الغائب المذكر المفرد	يكتب	الابن : (يكتب ستاتوس ع القيس) لا تتحدثوا عن الوجة اكثر الاشياء وجعا
be written/be fated/be destined write	كُتِبَ	فعل مضارع مجهول مفرد، مؤنث، غائب		نكتب	ت/المضارع الغائب المذكر المفرد	تكتب	، وفرجيت كيف هاللغه تطورت من أنجق لهجه ما كان نيسنوا تكتب للغه عالميه ما نيسنوا ما تكتب .
be written/be fated/be destined write	كُتِبَ	فعل مضارع مجهول مفرد، مؤنث، غائب		نكتب	ت/المضارع الغائب المذكر المفرد	تكتب	أنجق لهجه ما كان نيسنوا تكتب للغه عالميه ما نيسنوا ما تكتب .
scatter/sprinkle/write in prose	نُثِرَ	فعل مضارع مفرد، مذكر، غائب		نثر	ب/إداة مضارعة	ينثر	وينثر رمان
they (people) + write + (masc pl)	كُتِبُوا	فعل مضارع	'و/المضارع: فاعله مذكر جمع المذكر السالم	كُتِبُوا	ي/المضارع الغائب المذكر المفرد	يُكْتُبُوا	و الكل يتكلم عن السفن الجرف الأجنبية و كل الصيادين بيغوا يتكثروا



Lemmas	Annotations	Gloss	MSA Lemma	DA Lemma	Person	Gender	Number	POS	Suffix	Stem	Prefix	Token	Context	Dialect
	بصير <input type="checkbox"/> اسم <input checked="" type="checkbox"/> فعل <input checked="" type="checkbox"/> وظيفية <input type="checkbox"/> Count: 2	how	كَيْفَ 1	شَلُون				أداة استفهام		شلون		شلون	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
	بصير <input type="checkbox"/> اسم <input checked="" type="checkbox"/> فعل <input checked="" type="checkbox"/> وظيفية <input type="checkbox"/> Count: 2	will/shall	سَوِّفَ 1	رَخ رَاخ 3				أداة استقبال		دا		دا	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
	بصير <input type="checkbox"/> اسم <input checked="" type="checkbox"/> فعل <input checked="" type="checkbox"/> وظيفية <input type="checkbox"/> Count: 2	enter	دَخَلَ 1	دَخَلَ 1	مخاطب	مؤنث	مفرد	فعل مضارع	ي/للمضارع: فاعله مخاطب مؤنث مفرد	دخل	ت/للمضارع المخاطب المؤنث المفرد	تدخلي	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
	بصير <input type="checkbox"/> اسم <input checked="" type="checkbox"/> فعل <input checked="" type="checkbox"/> وظيفية <input type="checkbox"/> Count: 2	hand over/surren	سَلَّمَ 1	سَلَّمَ 1	مخاطب	مؤنث	مفرد	فعل مضارع	ي/للمضارع: فاعله مخاطب مؤنث مفرد	سلم	ت/للمضارع المخاطب المؤنث المفرد	تسلمي	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
	بصير <input type="checkbox"/> اسم <input checked="" type="checkbox"/> فعل <input checked="" type="checkbox"/> وظيفية <input type="checkbox"/> Count: 2	in childbed	نَفَسَاءَ 1	نَفَسَاءَ 1		مؤنث	مفرد	اسم		نفسا	ع/حرف جر+ال/أداة تعريف	عالنفسا	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
	بصير <input type="checkbox"/> اسم <input checked="" type="checkbox"/> فعل <input checked="" type="checkbox"/> وظيفية <input type="checkbox"/> Count: 2				علامة ترقيم		شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
	بصير <input type="checkbox"/> اسم <input checked="" type="checkbox"/> فعل <input checked="" type="checkbox"/> وظيفية <input type="checkbox"/> Count: 2	not	مَا 2	مَا 2				أداة نفي		ما		ما	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
	بصير <input type="checkbox"/> اسم <input checked="" type="checkbox"/> فعل <input checked="" type="checkbox"/> وظيفية <input type="checkbox"/> Count: 2	become/begin to	صَارَ	صَارَ	غائب	مذكر	مفرد	فعل مضارع		صير	ب/أداة مضارعة	بصير	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
	بصير <input type="checkbox"/> اسم <input checked="" type="checkbox"/> فعل <input checked="" type="checkbox"/> وظيفية <input type="checkbox"/> Count: 2	to/for + Allah/God +	الله	الله		مذكر	مفرد	اسم		الله	ي/أداة نداء	ياالله	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
	بصير <input type="checkbox"/> اسم <input checked="" type="checkbox"/> فعل <input checked="" type="checkbox"/> وظيفية <input type="checkbox"/> Count: 2	sisters sister/count	أَخْتٌ 1	أَخْتٌ 1	-	مؤنث		اسم		خيتو		خيتو	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية

Nabra Syrian Corpora

Corpus Collection:

From: Social media posts, scripts of movies and series, lyrics of songs and local proverbs. Size: 60K tokens, 6K sentences.

Local Syrian Dialects Covered:

Aleppo, Damascus, Deir-ezzur, Hama, Homs, Huran, Latakia, Mardin, Raqqah, and Suwayda

Dialect لهجة	Damascus (Shami) شامية	Aleppo حلبية	Latakia ساحلية	Raqqa رقاوية	Deir-Ezzur ديرية	Homs حمصية	Huran حوران	Suwayda سويداء	Hama حموية	Mardin ماردلية	Total المجموع
Tokens	17,274	9,255	7,893	6,284	4,322	4,139	3,807	3,150	2,322	1,575	60,021
Unique Tokens	7,123	4,452	3,829	3,389	2,453	2,047	2,094	1,681	1,355	949	29,372
Sentences	1,181	787	829	679	519	518	457	381	340	243	5,934
Unique MSA Lemma	4,230	2,825	2,548	2,367	1,909	1,543	1,580	1,312	1,051	686	20,051
Unique DA lemma	4,351	2,969	2,681	2,490	1,954	1,591	1,646	1,354	1,095	710	20,841
Nouns	7,700	4,251	3,771	3,316	2,384	2,064	2,090	1,527	1,135	694	28,932
Verbs	3,524	1,897	1,557	985	714	709	518	554	369	339	11,166
Functional Words	6,027	3,090	2,560	1,960	1,213	1,359	1,194	1,069	815	534	19,821

Annotation Guidelines

DA Sentence: شلون دا تدخلي تسلمي ع النفسا ... ما بصير يالله خيتو

Gloss	MSA Lemma	DA Lemma	Person	Gender	Number	POS	Suffix	Stem	Prefix	Token	Context	Dialect
how	كَيْفَ 1	شَلُون				أداة استفهام		شلون		شلون	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
will/shall	سَوْفَ 1	رَخ رَاخ 3				أداة استقبال		دا		دا	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
enter	دَخَلَ 1	تَدَخَّلْ 1	مخاطب	مؤنث	مفرد	فعل مضارع	ي/للمضارع: فاعله مخاطب مؤنث مفرد	دخل	ت/للمضارع المخاطب المؤنث المفرد	تدخلي	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
hand over/surren	سَلَّمَ 1	سَلَّمْ 1	مخاطب	مؤنث	مفرد	فعل مضارع	ي/للمضارع: فاعله مخاطب مؤنث مفرد	سلم	ت/للمضارع المخاطب المؤنث المفرد	تسلمي	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
in childbed	نَفَسَاءَ 1	نَفَسَاءَ 1		مؤنث	مفرد	اسم		نفسا	ع/حرف جر + أداة تعريف	عالنفسا	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
..				علامة ترقيم		شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
not	مَا 2	مَا 2				أداة نفي		ما		ما	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
become/begin to	صَارَ	صَارَ	غائب	مذكر	مفرد	فعل مضارع		صير	ب/أداة مضارعة	بصير	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
to/for + Allah/God +	الله	الله		مذكر	مفرد	اسم		الله	ي/أداة نداء	ياالله	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
sisters sister/count	أَخْتٌ 1	أَخْتٌ 1	-	مؤنث		اسم		خيتو		خيتو	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية

Annotation Guidelines

DA Sentence: شلون دا تدخلي تسلمي ع النفسا ... ما بيصير يالله خيتو

Gloss	MSA Lemma	DA Lemma	Person	Gender	Number	POS	Suffix	Stem	Prefix	Token	Context	Dialect
how	كَيْفَ 1	شَلُون				أداة استفهام				شَلُون	شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
										دا	شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
										تدخلي	شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
										تسلمي	شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
										عالنفسا	شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
										..	شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
										ما	شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
										بصير	شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
										يالله	شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
										خيتو	شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية

Token:

The “correct”/unified spelling of rowToken.

- Unintentional typos and spelling mistakes are corrected.
- Letters repeated more than two times are removed.
- Odd and rare spellings are corrected, if necessary, to follow more common spellings.

to/for + Allah/God +

الله

الله

مذكر

مفرد

اسم

الله

ي/أداة نداء

يالله

شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو

حلبية

sisters | sister/count

أخت 1

أخت 1

-

مؤنث

اسم

خيتو

خيتو

شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو

حلبية

Annotation Guidelines

DA Sentence: شلون دا تدخلي تسلمي ع النفسا ... ما بصير يالله خيتو

Gloss	MSA Lemma	DA Lemma	Person	Gender	Number	POS	Suffix	Stem	Prefix	Token	Context	Dialect
how	كَيْفَ 1	شَلُون								شَلُون	شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
will/shall	سَوِّفَ 1	رَخْ رَاخْ 3								دا	شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
enter	تَخَلَّ 1	تَخَلَّ 1	مخاطب	مؤنث			مؤنث مفرد	ت/للمضارع المخاطب المؤنث المفرد		تدخلي	شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
hand over/surren	سَلَّمَ 1	سَلَّمَ 1	مخاطب	مؤنث	مفرد	فعل مضارع	ي/للمضارع: فاعله مخاطب مؤنث مفرد	سلم	ت/للمضارع المخاطب المؤنث المفرد	تسلمي	شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
in childbed	نَقَسَاءَ 1	نَقَسَاءَ 1		مؤنث	مفرد	اسم		نفسا	ع/حرف جر+ال/أداة تعريف	عالنفسا	شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
..				علامة ترقيم		شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
not	ما 2	ما 2				أداة نفي		ما		ما	شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
become/begin to	صَارَ	صَارَ	غائب	مذكر	مفرد	فعل مضارع		صير	ب/أداة مضارعة	بصير	شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
to/for + Allah/God +	الله	الله		مذكر	مفرد	اسم		الله	ي/أداة نداء	ياالله	شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
sisters sister/count	أَخْت 1	أَخْت 1	-	مؤنث		اسم		خيتو		خيتو	شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية

Prefixes (سوابق)

Annotation Guidelines

DA Sentence: شلون دا تدخلي تسلمي ع النفسا ... ما بصير يالله خيتو

Gloss	MSA Lemma	DA Lemma	Person	Gender	Number	POS	Suffix	Stem	Prefix	Token	Context	Dialect
how	كَيْفَ 1	أُون 1						شلون		شلون	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
will/shall	سَوِّفَ 1	أُون 3						دا		دا	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
enter	تَدَخَلَ 1	دَخَلَ 1					ي/للمضارع: فاعله مخاطب مؤنث مفرد	دخل	ت/للمضارع المخاطب المؤنث المفرد	تدخلي	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
hand over/surren	سَلَّمَ 1	سَلَّمَ 1	مخاطب	مؤنث	مفرد	فعل مضارع	ي/للمضارع: فاعله مخاطب مؤنث مفرد	سلم	ت/للمضارع المخاطب المؤنث المفرد	تسلمي	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
in childbed	نَقَسَاءَ 1	نَقَسَاءَ 1			مفرد	اسم		نفسا	ع/حرف جر+ال/أداة تعريف	عالنفسا	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
..				علامة ترقيم		شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
not	مَا 2	مَا 2				أداة نفي		ما		ما	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
become/begin to	صَارَ 1	صَارَ 1	غائب	مذكر	مفرد	فعل مضارع		صير	ب/أداة مضارعة	بصير	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
to/for + Allah/God +	اللَّهُ 1	الله 1			مفرد	اسم		الله	ي/أداة نداء	ياالله	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
sisters sister/count	أَخْتٌ 1	أَخْتٌ 1	-		مؤنث	اسم		خيتو		خيتو	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية

Suffixes (لواحق)

Annotation Guidelines

DA Sentence: شلون دا تدخلي تسلمي ع النفسا ... ما بصير يالله خيتو

Gloss	MSA Lemma	DA Lemma	Person	Gender	Number	POS	Suffix	Stem	Prefix	Token	Context	Dialect
how	كَيْت 1	شَلُون						شلون		شلون	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
will/shall	سَوِّت 1	رَخ رَاخ 3						دا		دا	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
enter	تَخَل 1	تَخَل 1	مخاطب				ي/ فاعل مؤنث مفرد	دخل	ت/للمضارع المخاطب المؤنث المفرد	تدخلي	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
hand over/surren	سَلَم 1	سَلَم 1	مخاطب	مؤنث	مفرد	فعل مضارع	ي/للمضارع: فاعله مخاطب مؤنث مفرد	سلم	ت/للمضارع المخاطب المؤنث المفرد	تسلمي	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
in childbed	نَقْساء 1	نَقْساء 1			مفرد	اسم		نفسا	ع/حرف جر + أداة تعريف	عالنفسا	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
..				علامة ترقيم		شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
not	ما 2	ما 2				أداة نفي		ما		ما	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
become/begin to	صَارَ	صَارَ	غائب	مذكر	مفرد	فعل مضارع		صير	ب/أداة مضارعة	بصير	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
to/for + Allah/God +	الله	الله		مذكر	مفرد	اسم		الله	ي/أداة نداء	ياالله	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
sisters sister/count	أخْت 1	أخْت 1	-	مؤنث		اسم		خيتو		خيتو	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية

Stem (الساق)

Annotation Guidelines

DA Sentence: شلون دا تدخلي تسلمي ع النفسا ... ما بصير يالله خيتو

Gloss	MSA Lemma	DA Lemma	Person	Gender	Number	POS	Suffix	Stem	Prefix	Token	Context	Dialect
how	كَيْت 1	شَلُون				أداة استفهام		شلون		شلون	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
						أداة استقبال		دا		دا	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
			مؤنث	مفرد		فعل مضارع	ي/للمضارع: فاعله مخاطب مؤنث مفرد	دخل	ت/للمضارع المخاطب المؤنث المفرد	تدخلي	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
			مؤنث	مفرد		فعل مضارع	ي/للمضارع: فاعله مخاطب مؤنث مفرد	سلم	ت/للمضارع المخاطب المؤنث المفرد	تسلمي	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
			مؤنث	مفرد		اسم		نفسا	ع/حرف جر+ال/أداة تعريف	عالنفسا	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
..				علامة ترقيم		شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
not	ما 2	ما 2				أداة نفي		ما		ما	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
become/be gin to	صار	صار	غائب	مذكر	مفرد	فعل مضارع		صير	ب/أداة مضارعة	بصير	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
to/for + Allah/God +	الله	الله		مذكر	مفرد	اسم		الله	ي/أداة نداء	ياالله	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
sisters sister/count	أخت 1	أخت 1	-	مؤنث		اسم		خيتو		خيتو	شلون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية

Same SAMA tagsets:
POS, Person, Gender,
Number.

Annotation Guidelines

DA Sentence: شلون دا تدخلي تسلمي ع النفسا ... ما بيصير يالله خيتو

Gloss	MSA Lemma	DA Lemma	Person	Gender	Number	POS	Suffix	Stem	Prefix	Token	Context	Dialect
how	كَيْفَ 1	شَلُون				أداة استفهام				شَلُون	شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
will/shall	سَوْفَ 1	بَاخِ 1 بَاخِ 3									شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
enter	تَخَلَّ 1	تَخَلَّ 1									شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
hand over/surrender	سَلَّمَ 1	سَلَّمَ 1									شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
in childbed	نُقِساءَ 1	نُقِساءَ 1								عَالِيسَا	شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
..				علامة ترقيم			شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
not	مَا 2	مَا 2				أداة نفي		ما		ما	شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
become/begin to	صَارَ	صَارَ	غائب	مذكر	مفرد	فعل مضارع		صير	ب/أداة مضارعة	بصير	شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
to/for + Allah/God +	الله	الله		مذكر	مفرد	اسم		الله	ي/أداة نداء	ياالله	شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
sisters sister/count	أَخْتِ 1	أَخْتِ 1	-	مؤنث		اسم		خيتو		خيتو	شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية

MSA Lemma (مدخلة معجمية فصحي):

- SAMA lemmas
- If no SAMA lemma, and one.

Annotation Guidelines

DA Sentence: شلون دا تدخلي تسلمي ع النفسا ... ما بيصير يالله خيتو

Gloss	MSA Lemma	DA Lemma	Person	Gender	Number	POS	Suffix	Stem	Prefix	Token	Context	Dialect
how	كَيْفَ 1	شَلُون				أداة استفهام		شَلُون		شَلُون دا تدخلي تسلمي ع النفسا ... ما بيصير يالله خيتو	حلبية	
will/shall	سَوِّفَ 1	رَخ رَاخ 3								شَلُون دا عالنفسا ... ما بيصير يالله خيتو	حلبية	
enter	تَخَلَّ 1	تَخَلَّ 1	مخاطب							شَلُون دا عالنفسا ... ما بيصير يالله خيتو	حلبية	
hand over/surren	سَلِّمَ 1	سَلِّمَ 1	مخاطب							شَلُون دا عالنفسا ... ما بيصير يالله خيتو	حلبية	
in childbed	نُقِّسَاءَ 1	نُقِّسَاءَ 1								شَلُون دا عالنفسا ... ما بيصير يالله خيتو	حلبية	
..				علامة ترقيم		..		شَلُون دا تدخلي تسلمي عالنفسا ... ما بيصير يالله خيتو	حلبية	
not	مَا 2	مَا 2				أداة نفي		مَا		شَلُون دا تدخلي تسلمي عالنفسا ... ما بيصير يالله خيتو	حلبية	
become/begin to	صَارَ	صَارَ	غائب	مذكر	مفرد	فعل مضارع		صِير	ب/أداة مضارعة	بصير	شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية
to/for + Allah/God +	اللَّهِ	الله		مذكر	مفرد	اسم		الله	ي/أداة نداء	ياالله	شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير ياالله خيتو	حلبية
sisters sister/count	أَخْتٌ 1	أَخْتٌ 1	-	مؤنث		اسم		خيتو		شَلُون دا تدخلي تسلمي عالنفسا ... ما بصير يالله خيتو	حلبية	

Dialect Lemma (مدخلة معجمية عامية):

- From MSA, then its MSA and DA lemmas are the same.
- If not, introduce a new DA lemma.

- Randomly selected about 7K annotations and reviewed them manually
 - Initially, nine linguists annotated the corpus, consisting of 60K tokens. Subsequently, six linguists reviewed the annotations to approve or correct them.
- ❖ F1-Score, precision, recall metrics and Cohen's Kappa

Feature	P	R	F	κ
Stem	0.813	0.797	0.805	0.796
POS	0.874	0.862	0.868	0.843
Prefix	0.981	0.979	0.980	0.939
Suffix	0.947	0.941	0.944	0.837
DA Lemma	0.763	0.742	0.753	0.739
MSA Lemma	0.789	0.770	0.780	0.767

https://sina.birzeit.edu/currasat/

Try Online

Currasat
Arabic Dialects Corpora
مدونة اللهجات العربية

انبحث

Search

Word Stem Lemma Gloss
 Whole Word Substring

Palestinian Lebanese Syrian Iraqi Libyan Sudanese
 Yemeni

About Publications

2 results (0.72 secs)

Gloss	Lemma	POS	Suffix	Stem	Prefix	Word	Context
become angry	غَضِبَ 1	فعل ماضي مفرد، مؤنث، غالب	ت/للماضي: فاعله غالب مؤنث مفرد	انبحث		انبحث	انبحثت مثل الأبوة
become angry	غَضِبَ 1	فعل ماضي مفرد، متكلم	ت/للماضي: فاعله متكلم مفرد	انبحث		انبحثت	يوم انا ما انبحثت ... بس انتي كل مادا تسرخيلي تقبلي تاعي ابركي والا لهيك



Summary

- **Baladi**: Lebanese morphologically annotated corpus (9.6K)
- **Curras**: Palestinian morphologically annotated corpus (56K)
- **Nabra**: Syrian morphologically annotated corpora (60K)

Baladi + Curras + Nabra = a more **Levantine Corpus** (125.6K)

- **Lisan**: Iraqi, Yemeni, Sudanese, and Libyan morphologically annotated corpora (1.2M)

Total Tokens: **Currasat** contains (1.325.6M tokens)

References

1. Mustafa Jarrar, Sanad Malaysha, Tymaa Hammouda, Mohammad Khalilia: SALMA: Arabic Sense-Annotated Corpus and WSD Benchmarks. In Proceedings of the 1st Arabic Natural Language Processing Conference (ArabicNLP), Part of the EMNLP 2023. ACL.
2. Amal Nayouf, Tymaa Hammouda, Mustafa Jarrar, Fadi zaraket, Mohamad-Bassam Kurdy: Nâbra: Syrian Arabic Dialects with Morphological Annotations. In Proceedings of the 1st Arabic Natural Language Processing Conference (ArabicNLP), Part of the EMNLP 2023. ACL.
3. Mustafa Jarrar, Ahmet Birim, Mohammed Khalilia, Mustafa Erden, and Sana Ghanem: ArBanking77: Intent Detection Neural Model and a New Dataset in Modern and Dialectical Arabic. In Proceedings of the 1st Arabic Natural Language Processing Conference (ArabicNLP), Part of the EMNLP 2023. ACL.
4. Haneen Liqreina, Mustafa Jarrar, Mohammed Khalilia, Ahmed Oumar El-Shangiti, Muhammad AbdulMaged: Arabic Fine-Grained Entity Eecognition. In Proceedings of the 1st Arabic Natural Language Processing Conference (ArabicNLP), Part of the EMNLP 2023. ACL.
5. Mustafa Jarrar, Muhammad Abdul-Mageed, Mohammed Khalilia, Bashar Talafha, AbdelRahim El-madany, Nagham Hamad, Alaa' Omar: WojooodNER 2023: The First Arabic Named Entity Recognition Shared Task. In Proceedings of the 1st Arabic Natural Language Processing Conference (Arabic- NLP), Part of the EMNLP 2023. ACL.
6. Nouran Khallaf, Elin Arfon, Mo El-Haj, Jon Morris, Dawn Knight, Paul Rayson, Tymaa Hammouda, Mustafa Jarrar: Open-source thesaurus development for under-resourced languages: a Welsh case study. The 4th LDK Conference on Language, Data and Knowledge, Vienna, Austria, 12-15 September 2023
7. Nagham Hamad, Mustafa Jarrar, Mohammad Khalilia, Nadim Nashif: Offensive Hebrew Corpus and Detection using BERT. The 20th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA). IEEE. Egypt. 2023
8. Sana Ghanem, Mustafa Jarrar, Radi Jarrar, Ibrahim Bounhas: A Benchmark and Scoring Algorithm for Enriching Arabic Synonyms. The 12th International Global Wordnet Conference (GWC2023), Global Wordnet Association. (pp.). San Sebastian, Spain, 2023
9. Sanad Malaysha, Mustafa Jarrar, Mohammad Khalilia: Context-Gloss Augmentation for Improving Arabic Target Sense Verification. The 12th International Global Wordnet Conference (GWC2023), Global Wordnet Association. (pp.). San Sebastian, Spain, 2023
10. Mustafa Jarrar, Mohammed Khalilia, Sana Ghanem: Wojoood: Nested Arabic Named Entity Corpus and Recognition using BERT. In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2022), Marseille, France. 2022
11. Mustafa Jarrar, Fadi Zaraket, Tymaa Hammouda, Daanish Masood, Martin Waehlich: Lisan: Yemeni, Iraqi, Libyan, and Sudanese Arabic Dialect Corpora with Morphological Annotations. The 20th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA). Pages(-). IEEE. Egypt. 2023 arXiv, DOI 10.48550/ARXIV.2212.06468. 2023
12. Karim El Haff, Mustafa Jarrar, Tymaa Hammouda, Fadi Zaraket: Curras + Baladi: Towards a Levantine Corpus. In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2022), Marseille, France. 2022
13. Mustafa Jarrar: The Arabic Ontology - An Arabic Wordnet with Ontologically Clean Content. Applied Ontology Journal, 16:1, 1-26. IOS Press. 2021
14. Moustafa Al-Hajj, Mustafa Jarrar: ArabGlossBERT: Fine-Tuning BERT on Context-Gloss Pairs for WSD. In Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021). PP 40--48, 2021
15. Moustafa Al-Hajj, Mustafa Jarrar: LU-BZU at SemEval-2021 Task 2: Word2Vec and Lemma2Vec performance in Arabic Word-in-Context disambiguation. In Proceedings of the Fifteenth Workshop on Semantic Evaluation (SemEval2021) Task 2: Multilingual and Cross-lingual Word-in-Context Disambiguation (MCL-WIC). PP 748--755, Association for Computational Linguistics. 2021
16. Eman Naser-Karajah, Nabil Arman, Mustafa Jarrar: Current Trends and Approaches in Synonyms Extraction: Potential Adaptation to Arabic. In Proceedings of the 2021 International Conference on Information Technology (ICIT). PP 748--755, Association for Computational Linguistics. pp. 428-434, IEEE. 2021
17. Mustafa Jarrar, Eman Karajah, Muhammad Khalifa, Khaled Shaalan: Extracting Synonyms from Bilingual Dictionaries. The 11th International Global Wordnet Conference (GWC2021), Global Wordnet Association. (pp. 215-222). Pretoria, South Africa, 2021
18. Kareem Darwish, Nizar Habash, Mourad Abbas, Hend Al-Khalifa, Husein T. Al-Natsheh, Houada Bouamor, Karim Bouzoubaa, Violetta Cavalli-Sforza, Samhaa R. El-Beltagy, Wassim El-Hajj, Mustafa Jarrar, Hamdy Mubarak: A Panoramic Survey of Natural Language Processing in the Arab World. Communications of the ACM, April 2021, Vol. 64 No. 4, Pages 72-81
19. Mustafa Jarrar: Digitization of Arabic Lexicons. Arabic Language Status Report. UAE Ministry of Culture and Youth. Pages 214-2017. Dec 2020
20. Mustafa Jarrar, Hamzeh Amayreh: An Arabic-Multilingual Database with a Lexicographic Search Engine. The 24th International Conference on Applications of Natural Language to Information Systems (NLDB 2019). Pages(234-246). LNCS 11608, Springer. 2019
21. Mustafa Jarrar, Hamzeh Amayreh, John P. McCrae: Representing Arabic Lexicons in Lemon - a Preliminary Study. The 2nd Conference on Language, Data and Knowledge (LDK 2019). Pages(29-33). CEUR, Volume 2402. ISSN:1613-0073. Leipzig, Germany. 2019
22. Diana Alhafi, Anton Deik, Mustafa Jarrar: Usability Evaluation of Lexicographic e-Services. The 16th IEEE/ACS International Conference on Computer Systems and Applications (AICCSA). Pages(1-7). IEEE. Abu Dhabi, UAE. 2019
23. Mustafa Jarrar, Fadi Zaraket, Rami Asia, Hamzeh Amayreh: Diacritic-Based Matching of Arabic Words. ACM Asian and Low-Resource Language Information Processing. Volume 18, No 2, Pages(10:1-10:21), ACM, ISSN:2375-4699. December, 2018
24. Mustafa Jarrar: Search Engine for Arabic Lexicons. The 5th Conference on Translation and the Problematics of Cross-cultural Understanding. The Forum for Arab and International Relations. Doha, Qatar. December, 2018
25. Diab Abuaiadah, Dileep Rajendran, Mustafa Jarrar: Clustering Arabic Tweets for Sentiment Analysis. The 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications. Pages(499-506). IEEE Computer Society. ISBN:9781538635810. (doi.10.1109/AICCSA.2017.162). Hammamet, Tunisia. 2017
26. Mustafa Jarrar, Nizar Habash, Faeg Alrimawi, Divam Akra, Nasser Zalmout: Curras: An Annotated Corpus for the Palestinian Arabic Dialect. Journal Language Resources and Evaluation. Pages(745-775). Volume(51)