



Kalmasoft

Arabic Morphological Lexicon

Specifications

Aramolex©

Version 1.30



Table of Contents

Introduction.....	3
Aramolex.....	3
Contents.....	4
Versioning.....	4
Vocalization.....	4
Grammatical categories.....	5
Verbs.....	5
Nouns.....	5
Affixes.....	5
Clitics.....	5
Lexicon generation.....	6
Verb conjugation.....	6
Noun declension.....	6
Reference web page.....	6
POS code shortening.....	6
Database structure.....	7
Database field details.....	8
Data redundancy.....	11
Lexicon downsizing.....	12
Quantative downsizing.....	13
Qualitative downsizing.....	13
Lexicon minimal set.....	13
Lexicon optimization.....	14
Orthographic considerations.....	15
Database samples.....	16
Verb paradigm sample.....	16
Noun paradigm sample.....	16
Dataset delivery considerations.....	17
Statistical facts.....	18
Appendixes.....	19
Appendix (A): Arabic verb forms "templates".....	19
Appendix (B): Imperfect verb inflectional patterns.....	19
Appendix (C): Affixed particles.....	20
Appendix (D): Diminutive forms.....	20
Appendix (E): Verbal noun forms.....	21
Appendix (F): Hyperbolic participle.....	23
Appendix (G): Broken Plurals.....	23
Appendix (H): Instrumental noun.....	24
Appendix (I): Pseudoverbal adjective.....	25
Appendix (J): Grammatical terms.....	25



Introduction

Arabic is a non-concatinative language, it can be described as derivational language meaning that the morphotactics depend rather on affixation i.e. adding morphemes onto the word without changing the radical or the "root", that is, preserving the order of the verb core binyanim, this results in the highly regular inflectional/derivational patterns distinguishing the Arabic language.

It also gives the way to the highly prolific vocabulary that characterizes the Arabic morphology to be reproduced using any suitable production tool carefully designed to work on a root-based algorithm; for example a root like [ksr] "to break" may be seeded into the module to yield roughly 30,000 conjugations which is the thorough listing of the verb paradigm plus all inflectional morphemes applied using specific orthographic and heuristic rules, this is theoretically true for any other triconsonantal sound root with minor exceptions.

Current inflection/derivation tools are limited to the third person masculine singular past tense form which serves as the "dictionary form" used to identify any verb instead of the infinitive "verbal noun" the way appears in the English dictionaries for example; this basic form has no but a minor usage in NLP related operations e.g. Tokenization, Stemming or POS tagging not serving as practical. So it is important to generate the "real world" derivational candidates found in ordinary literature texts by having an exhaustive morphological lexicon.

Aramolex

Aramolex is an Arabic morphological lexicon, a dictionary database that is automatically generated to represent a full-form lexicon for the entire regular vocabulary of the Arabic language beside the other non-regular surface forms, the scope of the lexicon is the Modern Standard Arabic (MSA) with all possible morphological contents within the scope, Aramolex is semi-automatically annotated and provided with all disambiguating morphological information in an intuitive and ready-to-use format, the dictionary is also continuously undergoing optimization, maintenance, and extensions accompanied by subsequent filtering of candidates not recognized in the literature.



Contents

Contents of the dictionary are mainly the surface forms with sufficient grammatical information as well as glossaries.

Subcategories that are not included:

1. Proper nouns (personal names, place names, any other animate or inanimate).
Separate onomasticons available here www.kalmasoft.com/KLEX/dbgivnm.htm
2. Loanwords and neologisms.
3. Isolated forms and determiners (particles, prepositions, conjunctions, pronouns, demonstrative pronouns etc.)
4. Colloquial forms.
5. Archaic/classical and Quranic forms.

Versioning

This morphological dictionary will follow simple version control system, 0.5 is a draft, 1.0 is the final document, 1.2 is modification 2 and so on.

Vocalization

The database of the lexicon is structured into 20 fields, one field is "Vocalized" which include Arabic fully vocalized forms that are not necessarily follow the orthography found in the common text since some diacritics are considered extra and usually omitted e.g. The "Fat'ḥa" before "Taa' Marbuta" and before "Alif Maqsura" i.e. all necessary and extra diacritics were explicitly shown to disambiguate the surface form, this will render the whole dictionary an intelligible source of linguistic information that is definitely unambiguous instead of having partially vocalized forms.



Grammatical categories

Verbs

Verbs are conjugated according to the following subcategories

Tenses	Imperfective, Perfective, Future
Voices	Active, Passive
Moods	Indicative, Subjunctive, Jussive, Emphatic, Intense Emphatic, Imperative
Forms	I through X, see appendix A

Nouns

Nouns are declined according to the following subcategories

Nominals	Gerund, Active participle, Passive participle, Adjective, Place/Time and Utilization nouns, Broken Plurals.
Cases	Nominative, Accusative, Genitive

Verb classes, transitivity, patterns, defectiveness, infinitives, diptote etc; all these descriptive will be listed as subcategories.

Affixes

Affixes (both prefixes and suffixes) are mostly inflectional morphemes added the surface form to mark change in grammatical property or function.

Prefixes	ل، ب، ف، س، و، ي، ت، ن، ا	Refer to Appendix (C)
Suffixes	نا، ون، ني، ة	

The affixes above are not part of the dictionary but can be generated on request.

Clitics

Clitics are suffix morphemes mark the direct object and/or indirect object and mostly related to person, number and gender, these are attached to the surface form; clitics are grouped in pairs between two brackets and referred to in field named "Arguments".



Lexicon generation

The morphological lexicon is automatically generated, that is a process for building the whole paradigm for both verbs and nouns, entries undergo continuous refinement by software scripts and manual checking for odds and discrepancies.

Verb conjugation

Verb conjugation paradigm will be generated using Kalmasoft Verb Conjugator

<http://www.kalmasoft.com/KMAPS/mrlconj.htm>

Noun declension

Noun declension paradigm will be generated using Kalmasoft Noun Inflector/Decliner

<http://www.kalmasoft.com/KMAPS/mrlfct.htm>

Reference web page

All information regarding this lexicon is published in the reference web page

<http://www.kalmasoft.com/KLEX/aramolex.htm>

POS code shortening

The part of speech entry in the "POS" field should be 12 characters long if should reflect and fully describe the surface form, this may be impractical for subsequent electronic processing as this encoded POS tag has to be parsed to extract the grammatical information, shortening the POS code is an option open for discussion but we recommend using a subset of Kalmasoft Tagset here <http://www.kalmasoft.com/KMAPS/ktagset.htm>

The POS tag suggested by KTagset is encoding each grammatical class using a single letter, thus a set of four characters are enough for referencing the following grammatical classes {Category+Tense+Mood+Voice} for verbs and four for the nouns {Category+Class+Case+Definiteness}; extra information are placed in the "Subcategory" field e.g. transitivity of the verbs.



Database structure

A flat text file, TAB delimited, encoded in Windows-1256 (Arabic) with the following fields:

No	Field	Description
1	ID	4 digits Root identification number
2	SubID	5 digits unique ID for orthographic inflections
3	Status	Status code for the record, for internal tracking and processing
4	Category	Main category part of speech
5	Subcategory	Extended POS subcategory for sorting and statistical purposes
6	POS	Part of speech code, used for tagging purposes
7	Root	Arabic root, concatenated radical
8	Lemma	Arabic canonical form (headword)
9	Vocalized	Vocalized Arabic, word surface form fully diacritized
10	KATS	KATS version of "Vocalized" field for software coding compatibility purposes, please refer to www.kalmasoft.com/devtool.htm#kats.htm
11	Arguments	Person/Number/Gender combination marking the subject, object for verbs, and possessive pronouns (proclitic/enclitic pronouns) for nouns, 3-6 characters max.
12	Affixes	Prefixes and suffixes, pairs in brackets
13	Variety	Reference to Arabic varieties using the surface form
14	Gloss	POS glossary, plain English description for POS field
15	Production	Reproduction code used by Kalmasoft
16	Statistics	Statistics about the form, use as necessary
17	EXT1	Reserved for Kalmasoft future extension
18	EXT2	Reserved for Kalmasoft future extension
19	Created	Creation date
20	Modified	Modification date



Database field details

	Field name	Code: details	Description
	Status		
		O: entry checked and ready	
		C: entry to be checked	Use only for testing
		X: ignore this entry	Do not use
		D: entry to be deleted	
	Category		
		V: Verb	
		N: Noun	
		P: Particles	
	POS		
		P: Perfective "past" (PRF)	Tenses
		I: Imperfective "present" (IMF)	Tenses
		F: Future (FUT)	Tenses
		A: Active voice (ACT)	Voices
		P: Passive voice (PAS)	Voices
		I: Indicative (IND)	Moods
		S: Subjunctive (SUB)	Moods
		J: Jussive (JUS)	Moods
		N: Intense Emphatic (INT)	Moods
		E: Emphatic (EMP)	Moods
		M: Imperative (IMP)	Moods
		N: Nominative (NOM)	Noun cases
		A: Accusative (ACC)	Noun cases
		G: Genitive (GEN)	Noun cases



	Field name	Code: details	Description
		M: Masdar (MAS)	Nominal derivatives
		E: Relative nominal construct (RNC)	Nominal derivatives
		C: Active participle (APP)	Nominal derivatives
		P: Passive participle (PPP)	Nominal derivatives
		H: Hyperbolic participle (HYP)	Nominal derivatives
		I: Utilization Noun (UTL)	Nominal derivatives
		S: Superlative (SUP)	Nominal derivatives
		F: Definite (DEF)	Definiteness
		I: infefinite (NDF)	Definiteness
		J: Relative adjective (NSB)	Nominal/Sub-derivatives
		B: Pseudoverbal adjective (PSA)	Nominal/Sub-derivatives
		D: Diminutive (DIM)	Nominal/Sub-derivatives
		I: Intensive (INT)	Nominal/Sub-derivatives
		L: Locative noun (LOC)	Nominal/Sub-derivatives
		T: Temporal noun (TMP)	Nominal/Sub-derivatives
		V01: Form I (I)	Verb forms (templates)
		V02: Form II (II)	Verb forms (templates)
		V03: Form III (III)	Verb forms (templates)
		V04: Form IV (IV)	Verb forms (templates)
		V05: Form V (V)	Verb forms (templates)
		V06: Form VI (VI)	Verb forms (templates)
		V07: Form VII (VII)	Verb forms (templates)
		V08: Form VIII (VIII)	Verb forms (templates)
		V09: Form IX (IX)	Verb forms (templates)
		V10 : Form X (X)	Verb forms (templates)



	Field name	Code: details	Description
		S: Singular	Numbers
		D: Dual	Numbers
		P: Plural	Numbers
		1: First person	Persons
		2: Second person	Persons
		3: Third person	Persons
		F: Feminine	Genders
		M: Masculine	Genders
	Lemma		
		Verbs	Indicative-Perfective-Singular-Masculine
		Nouns	Nominative-Singular-Masculine
		Particles	Singular-Masculine

Notes:

- Categories in the POS field are the main ones, many other noun derivatives exist, they generally result in identical surface forms from the morphological perspective.
- Lemma is generated according to common rules regarding the headwords, other options are available.



Data redundancy

For the purpose of integrity, duplicate records are kept in the lexicon and marked with the appropriate flags in the (status) field.

Duplicates are records that look similar on the surface form level but they are grammatically different, some examples of the source of duplication are the following specific cases:

- Pronominal

Forms related to gender-neutral pronouns, example of this case is the first person singular, dual and plural.

- Affixes

Some of those are orthographically similar but semantically different like Prepositional Lam, and other similar particles (possessive Lam, emphatic Lams) that linked to nouns and verbs.

- Derivatives

Whether nominal or verbal, examples include infinitives (Mimatic infinitive, pseudoverbal adjective adjective, passive participle, locative noun), and also (superlative, hyperbolic participle).

The above similarities are only related to orthography they may not have the same grammatical function, the lexicon includes all those duplicates unless otherwise specified by the client.



Lexicon downsizing

Having few hundreds of millions of records may exert special challenges in many ways, in its full size the lexicon will be perfect for corpus analysis, morphological analyzers, POS taggers etc used as a training corpus, but for such applications like word cloud generation and sentiment analysis a downsized version is necessary. The following are figures calculated using our Conjugator and Inflector tools.

Category	Actively used	Maximum
Nominal templates	10	12
Noun cases	3	3
Noun forms	10	17
Plurals (1)	10	30
Noun affixes	6	12
Verb tense	3	3
Verb moods	3	5
Verb voices	3	3
Verb forms	10	15
Verb affixes	3	6
Arabic roots	3,000	7,500

(1) This includes Broken Plurals which are not applicable to the most of forms and can be ignored for the statistical purposes.

Note: the number below are estimation.

The minimal noun surface forms generated using a single set i.e. one template, one case one affix etc. is 100 forms for nouns, the figure is 250 forms for verbs. Based on the figures in the table above, and considering that verbs are considerably less than nouns, also not all verbs take the whole set and not all nouns accept all nominal templates the maximum number of active surface forms per paradigm for nouns is estimated 30K and 50K for verbs (total 80K), theoretically, since each root has its own characteristics which may reduce this number significantly.



Quantative downsizing

Removing the short vowels will render some surface forms morphologically identical e.g. the sound plural masculine and the dual masculine in the genitive case, the active participle and the pseudoverbal adjective, and the first person singular masculine past and third person feminine past; this will shrink noun total forms by 50% and 40% for verbs, thus reducing the numbers to 15K and 20K for nouns and verbs respectively, total is (35K) for a single root.

Having around 3K active roots in the Arabic language this method will downsize the lexicon to 105M surface forms instead of 240M.

Qualitative downsizing

The goal of this method is to keep only the most common Arabic vocabulary by removing the following subcategories which will have the corresponding effect in reducing the amount of records given in percentage next to each subcategory.

1. the dual (40%), both nouns and verbs
2. 1st person singular inflections (5%)
3. 2nd person singular inflections (5%)
4. interrogatives (3%)
5. the imperative, emphatic, and intensive moods of verbs (6%)
6. semi-classical noun templates and verb forms (10%)

Almost 70% size reduction of both nouns and verbs, this gives 72M surface forms instead of the maximum 240M.

Lexicon minimal set

Removing all double suffixes leaving only the base forms of the verbs with tenses and voices plus the verbal noun (infinitive form) with some nominal templates with few affixes will shrink the lexicon down to 30M surface forms though poor coverage.



Lexicon optimization

Though highly regular in terms of inflection and derivation morphology, Arabic nouns and verbs do not share even distribution over the available inflectional paradigms, this is better shown below where the four sample nouns do not have a template in common except for (AiC&tiCaAC) which is a verbal noun “Masdar” inflected from Form VIII, the same goes for all nominal categories e.g. Active Participle, Adjectives, Diminutive etc., this phenomenon is also occurs in verb paradigms.

		writing	judgement	pulling	burning
	Template	كتب	حكم	جذب	حرق
Form I	CaC&C		حُكْم	جَدَّب	حَرَق
Form II	taC&CiyC		تحكيم		
Form III	muCaACaCa:	مُكَاتِبَة	مُحَاكِمَة	مُجَادِبَة	
Form IV	EiC&CaAC		إحكام		إحراق
Form V	taCaC~uC		تَحْكَم		
Form VI	taCaACuC		تَحَاكُم	تَجَادِب	
Form VII	AinCiCaAC			إِنجذاب	
Form VIII	AiC&tiCaAC	إِكْتِتَاب	إِحْتِكَام	إِحْتِذَاب	إِحْتِرَاق
Form IX	AiC&CiCaAC				
Form X	Ais&tiC&CaAC	إِسْتِكْتَاب	إِسْتِحْكَام		

Some morphemes are more productive than the others (for the example the one highlighted in purple) and this is solely depends on the root i.e. no way to guess or automatically generate every form regardless of whether it actually exists, the blank (or full) cells need to be predicted for every single surface forms.

Kalmasoft Conjugation and Inflection tools are set to generate only those “common” inflections and exclude the ones that are not while reserving the possibility that the excluded forms may already been used in some regional Arabic varieties.

The algorithm implemented is based on heuristics and also some post-generation thorough check by language experts, this is not a perfect or optimal solution but practical enough to solve the NLP problem before hands.



Orthographic considerations

Some strict orthographic rules are followed in preparation of this lexicon, some of them are not common but kept for consistency reasons.

1. Sukun is shown over Lam in the Definite Article (ال).
2. The accusative indefinite case ending Tanwiin or “Fat’hatain” (َ) is set to come before the additional suffixed Alif (e.g. كَسْرًا), where in many occasions this “Fat’hatain” is written after the final Alif (e.g. كَسْرًا).
3. Accusative case ending short vowel “Fat’ha” before Alif (َ) is mostly unnecessary, it is added in the lexicon since it follows the full vocalization scheme (e.g. كَاسِر).
4. Indicative/nominative case ending short vowel “Damma” before Waw (ُ) is mostly unnecessary, it is to indicate long vowel since it follows the full vocalization scheme (e.g. مَكْسُور).
5. Genitive case ending short vowel “Kasra” before Yaa' (ِ) is mostly unnecessary, it is added to indicate long vowel since it follows the full vocalization scheme (e.g. تَكْسِير).
6. Fat’ha before final Taa' Marbuta or Yaa' Maqsura is unnecessary, it is kept for consistency.
7. Fat’ha over Hamza (أَ) is mostly unnecessary, it is kept for consistency.
8. Kasra under Hamza (إِ) is mostly unnecessary, it is kept for consistency.
9. Conjunctional Waw (وَ) which occasionally comes separated from the following surface form is kept linked with the form.
10. The glottal stop “Hamza” is usually written using the form (ء), the form (أ) with Hamza mounted over Alif is used instead for the roots if shown in the final database.



Database samples

Samples of the dictionary are available for evaluation purposes and will be provided on request with the following minimal format:

Verb paradigm sample

ID	Cat	POS	Root	Lemma	Vocalized	KATS	Arguments	Affixes	Gloss
V1	V	VPIA	كسر	كَسَرَ	كَسَرَتْهَا	kasar&tuhaA	1SM-3SF	[-, &tuhaA]	Active- Perfective- Indicative
V2	V	VPIA	كسر	كَسَرَ	كَسَرَتْهَا	kasar&tihaA	2SF-3SF	[-, tihaA]	Active- Perfective- Indicative
V3	V	VPIA	كسر	كَسَرَ	كَسَرَتَاهَا	kasar&naAha A	1PM-3SF	[-, naAhaA]	Active- Perfective- Indicative

Noun paradigm sample

ID	Cat	POS	Root	Lemma	Vocalized	KATS	Arguments	Affixes	Gloss
N1	N	NCN	نصر	نَاصِر	نَاصِرَ	naASira	SM	[-, a]	Active participle- Accusative
N2	N	NCN	نصر	نَاصِر	نَاصِرُونَ	naASiruwna	PM	[-, uwna]	Active participle- Nominative
N3	N	NPN	نصر	مُنْتَصِر	مُنْتَصِرَةٌ	muntaSira:i	SF	[mun, :i]	Passive participle- Genitive
N4	N	NPN	نصر	مَنْصُور	مَنْصُورُكُمْ	manSuwruku m&	SM-2PM	[ma-ukum]	Passive participle- Nominative



Dataset delivery considerations

Having a dataset of some 200 million records in a single text file with every 1 million records represent roughly 6.5MB of the file size (Windows 1256 character encoding) that is, the final file size will occupy around 1.3GB media size of data, compressing the data will shrink the size further to 130MB which is fairly easy to move around and also possible to load in today's computer memories but, opening such large file for checking any specific record may not be as efficient as needed, and will also be hard to manipulate and process for even a simple copy operation.

A division strategy is necessary for the sake of simplicity and efficiency, the following methods of dividing the contents before delivery are suggested:

1. Root based division

- Quantity of roots

Providing large sets of data related to fully inflected roots , this will ensure a consistent blocks of data as for each single root there will be a complete set of all derivations related to that root in the delivered dataset.

- Root class

Same as above except that each time a special class of roots are delivered eg all triconsonontals that start with specific letter.

2. POS based division

- Delivery of specific grammatical part of speech regardless of roots involved eg all the perfective verbs, verbal nouns or adjectives.



Statistical facts

The data given below are just a prediction, the exact statistical figures are subject to change according to many factors that are decided upon discussion and mutual agreement.

Item	Total records	Notes
Total active Arabic roots	4,000	
Total estimated surface tokens	200,000,000	surface forms
Total number of surface verb forms	80,000,000	
<ul style="list-style-type: none"> All forms 		
Total number of surface noun forms	150,000,000	
<ul style="list-style-type: none"> Adjectives 		
<ul style="list-style-type: none"> Present participles 		
<ul style="list-style-type: none"> Past participles 		
Database text file size	1.3GB	Windows1256 encoding
Compressed file size	130MB	RAR archive
Minimized version		
Total estimated surface tokens	40,000,000	surface forms
Total number of surface verb forms	25,000,000	
<ul style="list-style-type: none"> All forms 		
Total number of surface noun forms	15,000,000	
<ul style="list-style-type: none"> Adjectives 		
<ul style="list-style-type: none"> Present participles 		
<ul style="list-style-type: none"> Past participles 		



Appendixes

Appendix (A): Arabic verb forms "templates"

No.	Form	Code	Example	Notes
1	I: Form I	V01	kataba (كَتَبَ)	
2	II: Form II	V02	kattaba (كَتَّبَ)	
3	III: Form III	V03	kātaba (كَاتَبَ)	
4	IV: Form IV	V04	'aktaba (أَكْتَبَ)	
5	V: Form V	V05	takattaba (تَكَتَّبَ)	
6	VI: Form VI	V06	takātaba (تَكَاتَبَ)	
7	VII: Form VII	V07	inkataba (اِنْكَتَبَ)	
8	VIII: Form VIII	V08	iktataba (اِكْتَتَبَ)	
9	IX: Form IX	V09	iHmarra (اِحْمَرَ)	
10	X : Form X	V10	istaktaba (اِسْتَكْتَبَ)	

Appendix (B): Imperfect verb inflectional patterns

The following imperfective verb patterns will be included in the lexicon.

No.	Pattern	Code	Example	Notes
1	CaCaCa-yaCCaCu	T01	فتح	فَعَلَ-يَفْعُلُ
2	CaCaCa-yaCCiCu	T02	ضرب	فَعَلَ-يَفْعِلُ
3	CaCaCa-yaCCuCu	T03	نصر	فَعَلَ-يَفْعُلُ
4	CaCiCa-yaCCaCu	T04	فرح	فَعَلَ-يَفْعَلُ
5	CaCiCa-yaCCiCu	T05	وثق	فَعَلَ-يَفْعِلُ
6	CaCuCa-yaCCuCu	T06	حسن	فَعَلَ-يَفْعُلُ



Appendix (C): Affixed particles

The following are the particles prefixed to various noun and verb forms, they are generated only upon request with their all legitimate combinations.

No.	Particle	Code	Function	Notes
1	ال	R01	Definite article	nouns only
2	ك	R02	Simile Kaf	nouns only
3	ب	R03	Prepositional Ba'	nouns only
4	ل	R04	Affirmative Lam	nouns only
5	ل	R05	Imperative Lam	verbs only
6	ل	R06	Causative Lam	verbs only
7	ل	R07	Swear Lam	verbs only
8	و	R08	Conjunctive Waw	nouns and verbs
9	ف	R09	Conjunctive Fa'	nouns and verbs
10	ل	R10	Prepositional Lam	nouns and verbs
11	أ	R11	Interrogative Hamza	nouns and verbs

Including the Tanween is subject to discussion, combined particles will come stacked in the same order they appear in the surface form eg. (R02R07R05R01) denotes (أولل).

Appendix (D): Diminutive forms

The following are the diminutive forms used with nouns.

No.	Form	Code	Example	Notes
1	CuCayC	D01	حسين	فُعَيْلٌ
2	CuCayCiC	D02	مديخر	فُعَيْعِلٌ
3	CuCayCiyC	D03	مسيمير	فُعَيْعِيلٌ



Appendix (E): Verbal noun forms

The following are the verbal noun forms derived from verbs. (*)

No.	Form	Code	Example	Notes
1	CaCC	M01	نصر	فَعْلٌ
2	CiCC	M02	سحر	فِعْلٌ
3	CuCC	M03	حزن	فُعْلٌ
4	CaCaC	M04	طرب	فَعَلٌ
5	CaCiyC	M05	رحيل	فَعِيلٌ
6	CaCuWC	M06	قبول	فَعُولٌ
7	CuCuWC	M07	دخول	فُعُولٌ
8	CaCaAC	M08	كساد	فَعَالٌ
9	CiCaAC	M09	قيام	فِعَالٌ
10	CuCaAC	M10	سُعال	فُعَالٌ
11	taCCiCa	M11	تجربة	تَفْعِلَةٌ
12	taCCuCa	M12	تهلكة	تَفْعُلَةٌ
13	CaCaACah	M13	سماحة	فَعَالَةٌ
14	CiCaACah	M14	ملاحة	فِعَالَةٌ
15	CuCaACah	M15	حوامة	فُعَالَةٌ
16	CaCaAC	M16	عذاب	فَعَالٌ
17	CaCaCAn	M17	غليان	فَعْلَانٌ
18	CuCCAn	M18	غفران	فُعْلَانٌ
19	CiCCAn	M19	عرفان	فِعْلَانٌ
20	CaCCah	M20	صنعة	فَعْلَةٌ
21	CuCCah	M21	سمرة	فُعْلَةٌ
22	CiCCah	M22	حرفة	فِعْلَةٌ
23	CaCuWCah	M23	ضرورة	فَعُولَةٌ



No.	Form	Code	Example	Notes
24	CuCuWCah	M24	صعوبة	فُعُولَة
25	CaCCaCah	M25	دحرجة	فَعَّلَة
26	CiCCAC	M26	وسواس	فِعْلَال
27	tiCaAC	M27	تبيان	تَفَعَّال
28	ECCaAC	M28	إخراج	إِفْعَال
29	taCCiyC	M29	ترميم	تَفْعِيل
30	muCaACaCah	M30	مراجعة	مُقَاعَلَة
31	taCaCCuC	M31	تبعثر	تَفَعَّل
32	taCaCCuC	M32	تقدم	تَفَعَّل
33	taCaACuC	M33	تراجع	تَفَاعَّل
34	EnCiCaAC	M34	انكسار	إِنْفِعَال
35	EctiCaAC	M35	ارتجاج	إِفْعِيَال
36	ECCaCaAC	M36	احمرار	إِفْعَالَال
37	ECCiCCaAC	M37		إِفْعَالَال
38	ECCinCAC	M38	اطمئنان	إِفْعِيَالَال
39	EstCCaAC	M39	استعداد	إِسْتَفْعَال
40	ECCiyCaAC	M40	اخشيشان	إِفْعِيَعَال
41	ECCiWWaAC	M41		إِفْعَوَال

(*) Arabic infinitives differ from English ones, they are rather derivations (verbal nouns).



Appendix (F): Hyperbolic participle

The following are the hyperbolic participle forms.

No.	Form	Code	Example	Notes
1	CaC~AC	H01	كذاب	فَعَّال
2	CaCuWC	H02	طروب	فَعُول
3	CaCiyC	H03	عليم	فَعِيل
4	CiC~iyC	H04	قديس	فَعِيل
5	CaCiC	H05	فطن	فَعِل
6	miCCaAC	H06	متلاف	مِفْعَال

Appendix (G): Broken Plurals

No	Pattern	Code	Example	Notes
1	AaCCaAC	P01	أنجال	أَفْعَال
2	AaCCuC	P02	أنجم	أَفْعُل
3	CiCCh	P03		فِعْلَة
4	AaCCiCh	P04	أوردة	أَفْعِلَة
5	CiCCAn	P05	حملان	فِعْلَان
6	CuCC	P06	فلك	فُعُل
7	CuCCAn	P07	بعران	فُعْلَان
8	CuCuC	P08	غير	فُعُل
9	CuCaCA'	P09	بخلاء	فُعَلَاء
10	CuCaC	P10	غرف	فُعَل
11	AaCCiCA'	P11	أقوياء	أَفْعِلَاء
12	CiCaC	P12	نعم	فِعَل
13	CaWACiC	P13	كوارث	فَوَاعِل
14	CuCaCh	P14		فُعْلَة



No	Pattern	Code	Example	Notes
15	CaCaAeiC	P15	صحائف	فَعَائِل
16	CaCaCh	P16	بررة	فَعَلَة
17	CaCaACiC	P17	عجائب	فَعَائِل
18	CiCaCh	P18	فيلة	فَعَلَة
19	CaCaACiy	P19	سحالي	فَعَالِي
20	CaCCY	P20	جرحى	فَعَلَى
21	CuC~aC	P21	خلص	فُعِّل
22	CuC~AC	P22	كتاب	فُعَّال
23	CaCaACiyC	P23	محاليل	فَعَالِيل

Appendix (H): Instrumental noun

The following are the instrumental noun forms.

No.	Form	Code	Example	Notes
1	miCCaC	S01	مرجل	مِفْعَل
2	miCCAC	S02	منشار	مِفْعَال
3	miCCaCh	S03	مكنسة	مِفْعَلَة
4	CACiCh	S04	حاصدة	فَاعِلَة
5	CACuWC	S05	ساطور	فَاعُول
6	CACuWCh	S06	طاخونة	فَاعُولَة
7	CaC~AC	S07	برّاد	فُعَّال
8	CaC~ACh	S08	غسّالة	فُعَّالَة



Appendix (I): Pseudoverbal adjective

The following are the pseudoverbal adjective forms.

No.	Form	Code	Example	Notes
1	ACCaC	A01	أحور	أَفْعَل
2	CaCCAn	A02	عطشان	فَعَلَان
3	CaCiC	A03	فرح	فَعِل
4	CuCaAC	A04	شجاع	فُعَال
5	CaCaAC	A05	جبان	فَعَال
6	CaCC	A06	سهل	فَعَل
7	CuCC	A07	حلو	فُعَل
8	CaCaC	A08	حسن	فَعَل
9	CuCuC	A09	كفو	فُعَل
10	CayCiC	A10	طيب	فَيَعِل

Appendix (J): Grammatical terms

Please check the full list here <http://www.kalmasoft.com/adevtool.htm#gloss>

	English	Arabic	Abbreviation	Example
	noun	اسم	N	
	verb	فعل	V	
	particle	حرف		عن
	proper noun	اسم علم		الخرطوم
	exclusion	الاستثناء		
	negation	النفى		
	negative	منفي		
	affirmative	مثبت		
	neuter	عديم الجنس		



English	Arabic	Abbreviation	Example
masculine	مذكر	m	
feminine	مؤنث	f	
unisex	مستوي الجنس	u	
first person	متكلم	1st	
second person	مخاطب	2nd	
third person	غائب	3rd	
singular	مفرد	s	
dual	مثنى	d	
plural	جمع	p	
inanimate	جماد		
animate	حي		
agent, subject	فاعل		
patient, object	مفعول		
active participle	اسم فاعل		
passive participle	اسم مفعول		
adjective	صفة	Adj	
comparative	صيغة التفضيل		أكبر
proper adjective	صفة منسوبة		عبقري
hyperbolic participle	صيغة المبالغة		كسّار
Pseudoverbal adjective	الصفة المشبهة		كسير
possessive pronoun			لي، لك
Broken plural	جمع التكسير		
Sound Masculine Plural	جمع المذكر السالم		مؤمنون
Sound Feminine Plural	جمع المؤنث السالم		مؤمنات



	English	Arabic	Abbreviation	Example
	reflexive pronoun	ضمير ذاتي		نفس، ذات
	pronoun	ضمير		
	adverb	ظرف	Adv	
	intensifier adverb	مفعول مطلق		
	adverbial phrase	حال		
	subordinate	مبتدأ		
	predicate	خبر		
	diptote	الممنوع من الصرف		
	triptote	المنصرف		
	genitive construct	إضافة		
	superlative	اسم التفضيل		أكبر
	agreement	الإتباع اللغوي		
	accusative of vocation	منادى		
	conjunction	عطف، أداة عطف		
	preposition	حرف جر		
	relative pronoun	اسم الموصول		الذي
	numerative noun	اسم المرة		ضربة
	temporal noun	اسم الزمان		
	locative noun	اسم المكان		
	noun of utilization	اسم الآلة		
	gerund	مصدر مؤول		
	infinitive	مصدر		
	possessive pronoun	ضمير ملكية		
	diminutive	صيغة تصغير		



English	Arabic	Abbreviation	Example
case	حالة الاسم		
nominative	حالة الرفع، مرفوع		
accusative	حالة النصب، منصوب		
Nunation	التنوين		
genitive, possessive	حالة الجر، مجرور		
modals, invariables	الجوامد		هيهات
tense	زمان الفعل		
active phrase	جملة فعلية		
active voice	المبني للمعلوم		
passive voice	المبني للمجهول		
verb conjugation	تصريف الأفعال		
mood	الحالة الإعرابية		
indicative	الرفع		
subjunctive	جملة الشرط		
jussive	المجزوم		
emphatic	الفعل المؤكد		لتكسرنَّ
perfective, past	ماض		
imperfective, present	مضارع		
future	مستقبل		
imperative	أمر		أكسر
intransitive	لازم		
transitive	متعد		كسر
ditransitive	متعد لمفعولين		أعطى
ambitransitive	متعد ولازم		



English	Arabic	Abbreviation	Example
mood	حالة الفعل		
assimilated verb	الفعل المثال		وعد
hollow verb	الفعل الأجوف		قال
deficient verb	الفعل الناقص		كسا
derivation	الاشتقاق		
derived	مشتق		
static	جامد		
root	جذر		
sound	صحيح		كسر
defective	معتل		
triconsonantal root	جذر ثلاثي صحيح		كسر
quadrilateral root	جذر رباعي		عشعش
Hamzated	جذر مهموز		أمر
geminated	جذر مضعف		مدّ
symmetric			قلق
part of speech	قسم الكلام		
definite article	أداة التعريف		ال
prepositional Lam	لام الابتداء		
swear Lam	لام القسم		
simile Kaf	كاف التشبيه		
prepositional Ba'	باء الجر		
clitic	طرف صرفي		
proclitic	بادئة صرفية		
enclitic	خاتمة صرفية		



	English	Arabic	Abbreviation	Example
	agreement	الإتباع اللغوي		
	affixes	الزوائد		
	prefix	السوابق		
	suffexes	اللواحق		
	gender	النوع		
	diphthong	الإدغام		
	number	العدد		مفرد، مثنى، جمع
	lexicon	معجم		