

The Quranic Arabic Dependency Treebank

Syntactic Annotation Guidelines

Kais Dukes, Eric Atwell and Abdul-Baqee Sharaf

School of Computing, University of Leeds, UK

{sckd,csc6ea,scsams}@leeds.ac.uk



Introduction

- The Quranic Arabic Corpus (QAC) is an online linguistic resource [<http://corpus.quran.com>]
- The Quran is the central religious text of Islam, over 1,400 years old, a unique genre written in Quranic Arabic
- Annotated Layers:
 - Morphological Features [completed] [Dukes & Habash 2010]
 - Part-of-speech [completed]
 - Word-by-word translation
 - Syntactic Treebank [in-progress, this poster]
 - Anaphora resolution [started → in-progress]
 - Conceptual Ontology [started → in-progress]
 - Semantic Frames [planned]
- QAC includes infrastructure for collaborative annotation.
- A popular learning resource: 50,000 users per month.

Tagset for Dependency Grammar Relations

Cat*	Rel	Arabic	Description
1	adj	صفة	Adjective
	poss	مضاف إليه	Possessive construction
	pred	مبتدأ وخبر	Predicate of a subject
	app	بدل	Apposition
	spec	تمييز	Specification
	cpnd	مركب	Compound (numbers)
2	subj	فاعل	Subject of a verb
	pass	نائب فاعل	Passive subject
	obj	مفعول به	Object of a verb
	subjx	اسم كان	Subject of a special verb
	predx	خبر كان	Predicate of a special verb
	impv	أمر	Imperative
	imrs	جواب أمر	Imperative result
	pro	نهي	Prohibition
	gen	جار ومجرور	Preposition phrase (PP)
	link	متعلق	PP attachment
3	conj	معتوف	Coordinating conjunction
	sub	صلة	Subordinate clause
	cond	شرط	Condition
	rslt	جواب شرط	Result
	circ	حال	Circumstantial accusative
4	cog	مفعول مطلق	Cognate accusative
	prp	المفعول لأجله	Accusative of purpose
	com	المفعول معه	Comitative object

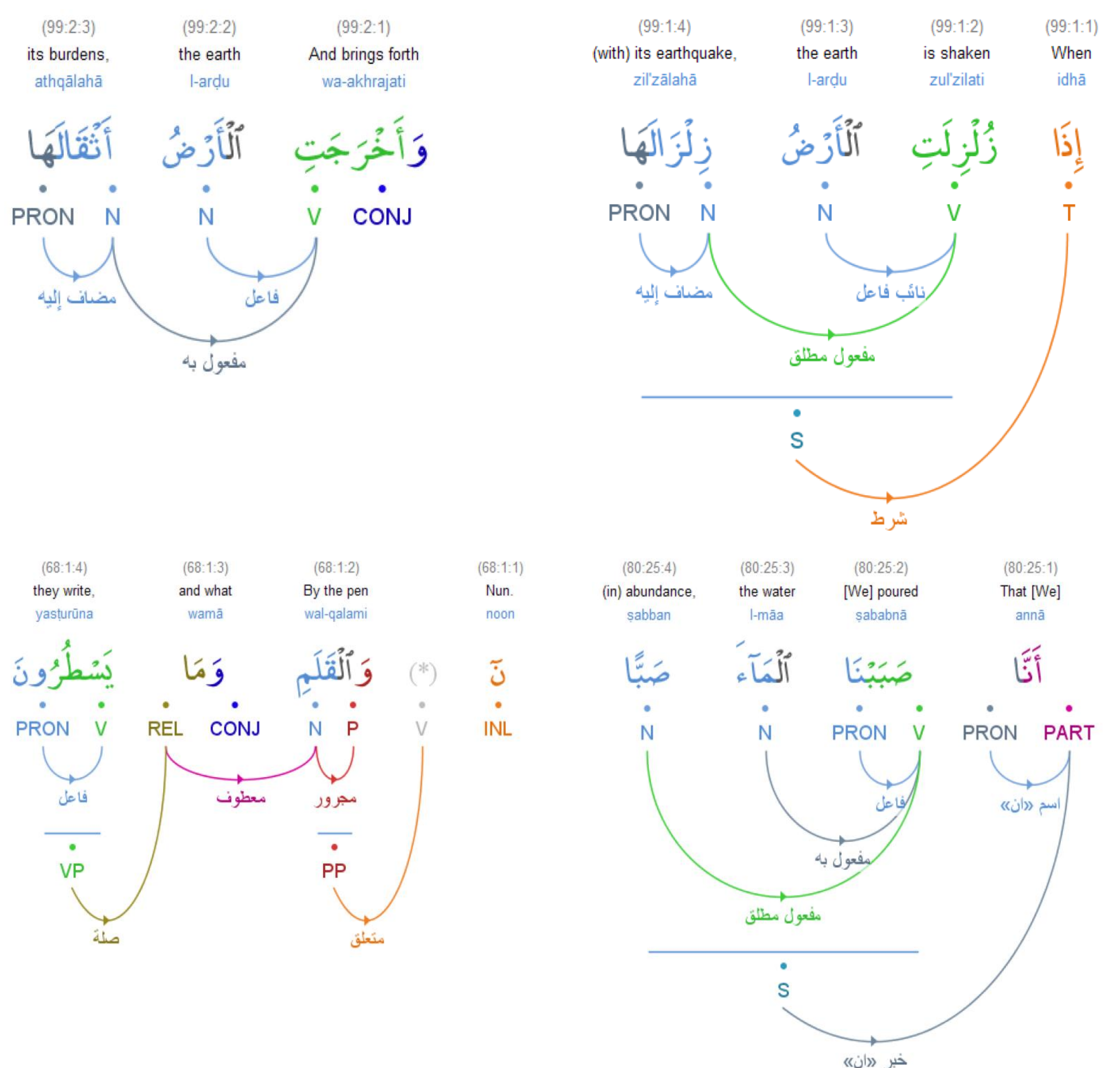
emph	توكيد	Emphasis
intg	استفهام	Interrogation
neg	نفي	Negation
fut	استقبال	Future clause
voc	منادي	Vocative
exp	مستثني	Exceptive
res	حصر	Restriction
avr	ردع	Aversion
cert	تحقيق	Certainty
ret	اضراب	Retraction
prev	كاف	Preventive
ans	جواب	Answer
inc	ابتداء	Inceptive
sup	فجأة	Surprise
exh	تحضيض	Exhortation
exl	تفصيل	Explanation
eq	تسوية	Equalization
caus	سببية	Cause
amd	استدراك	Amendment

*Categories: 1=Nominal dependencies, 2=Verbal dependencies, 3=Phrases and clauses, 4=Adverbial dependencies, 5=Particle Dependencies

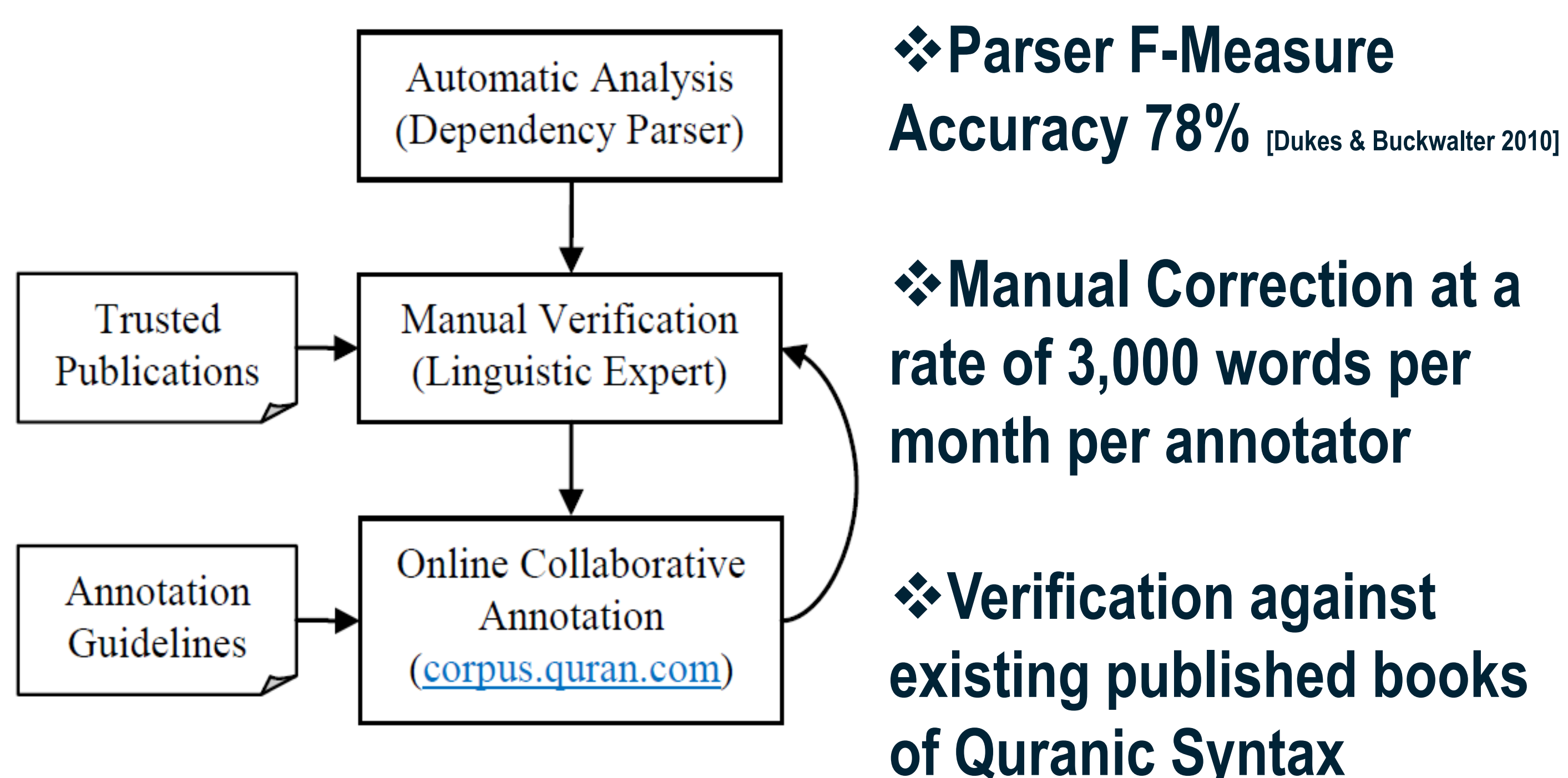
Traditional Arabic Grammar [l'rab]

- Early Arabic linguists studied classical Arabic syntax for over 1,000 years.
- Traditional Arabic grammar is considered to be one of the origins of modern dependency grammar
- Developing syntactic treebank for the Quran requires catering for rich morphological and highly derivational forms of Arabic. Moreover, the syntactic unit of analysis is not a word but a morphological segment.
- Syntactic representation adopted is a hybrid dependency constituency phrase structure model capable of showing relationship between words as well as phrases through non-terminal nodes.

Syntax of the Quran as Dependency Graphs



Annotation Process



Current Status

Morphological annotation is 100% complete. Syntactic dependency treebank is available for Chapters 1-3 and 67 – 114 covering approx. 11,000 words (14%).

Correction and discussion is available online:

<http://corpus.quran.com/messageboard.jsp>

References

Dukes, K. and Buckwalter, T. (2010) "A dependency treebank of the Quran using Traditional Arabic Grammar". 7th Int. conf. on Informatics and Systems. Cairo, Egypt.

Kais Dukes and Nizar Habash. (2010) Morphological Annotation of Quranic Arabic. LREC-2010 Valletta, Malta.

Applications of QAC

- Better visualization of the Syntax of the Quran
- Educational resource for students of Arabic and Quran
- A machine readable representation of the Quranic grammar
- Dataset is publicly available and used in academic research