

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

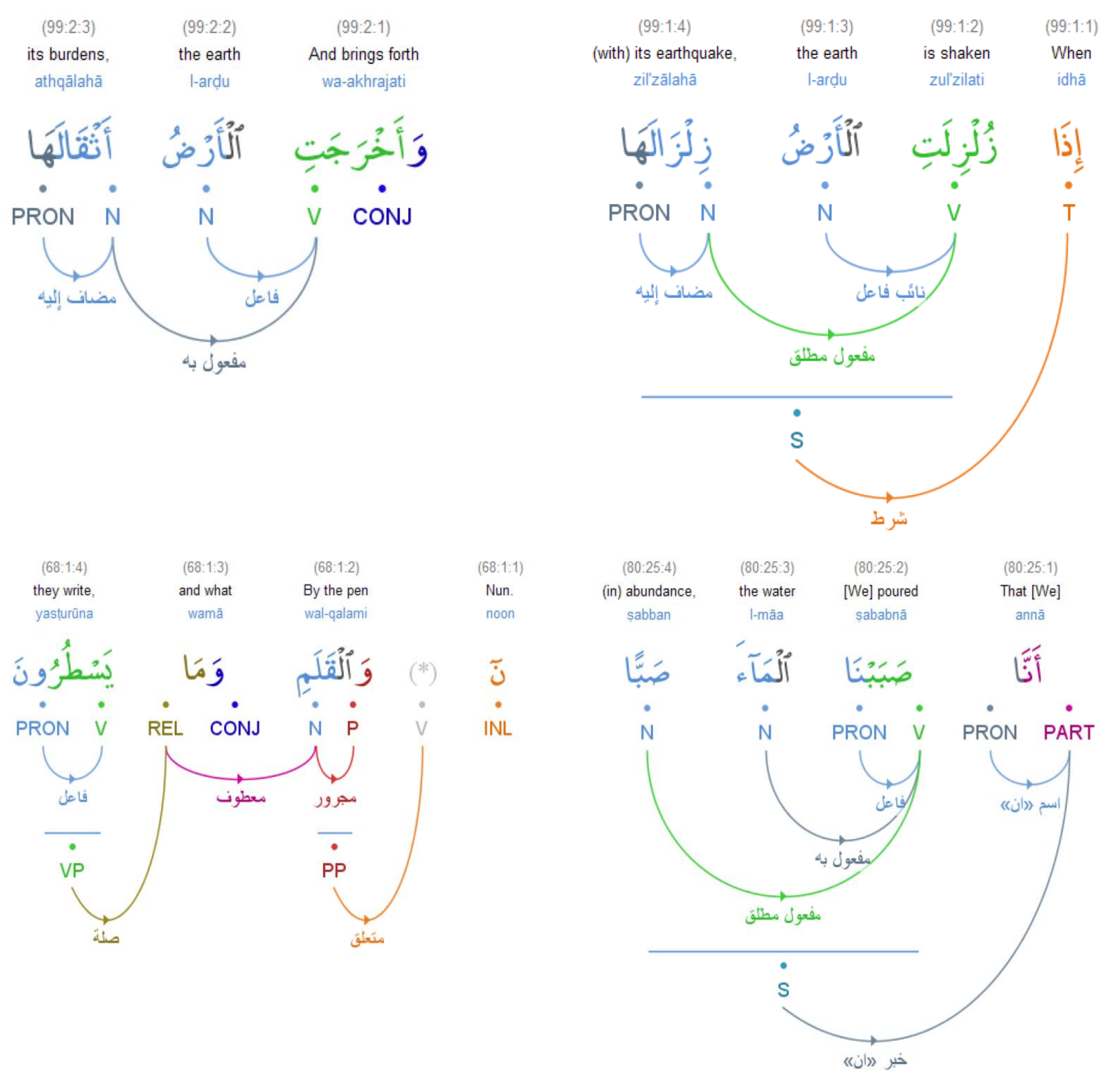
<i>emph</i>	توكيد	Emphasis
<i>intg</i>	استفهام	Interrogation
<i>neg</i>	نفي	Negation
<i>fut</i>	استقبال	Future clause
<i>voc</i>	منادي	Vocative
<i>exp</i>	مستثني	Exceptive
<i>res</i>	حصر	Restriction
<i>avr</i>	ردع	Aversion
<i>cert</i>	تحقيق	Certainty
<i>ret</i>	اضراب	Retraction
<i>prev</i>	كاف	Preventive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

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

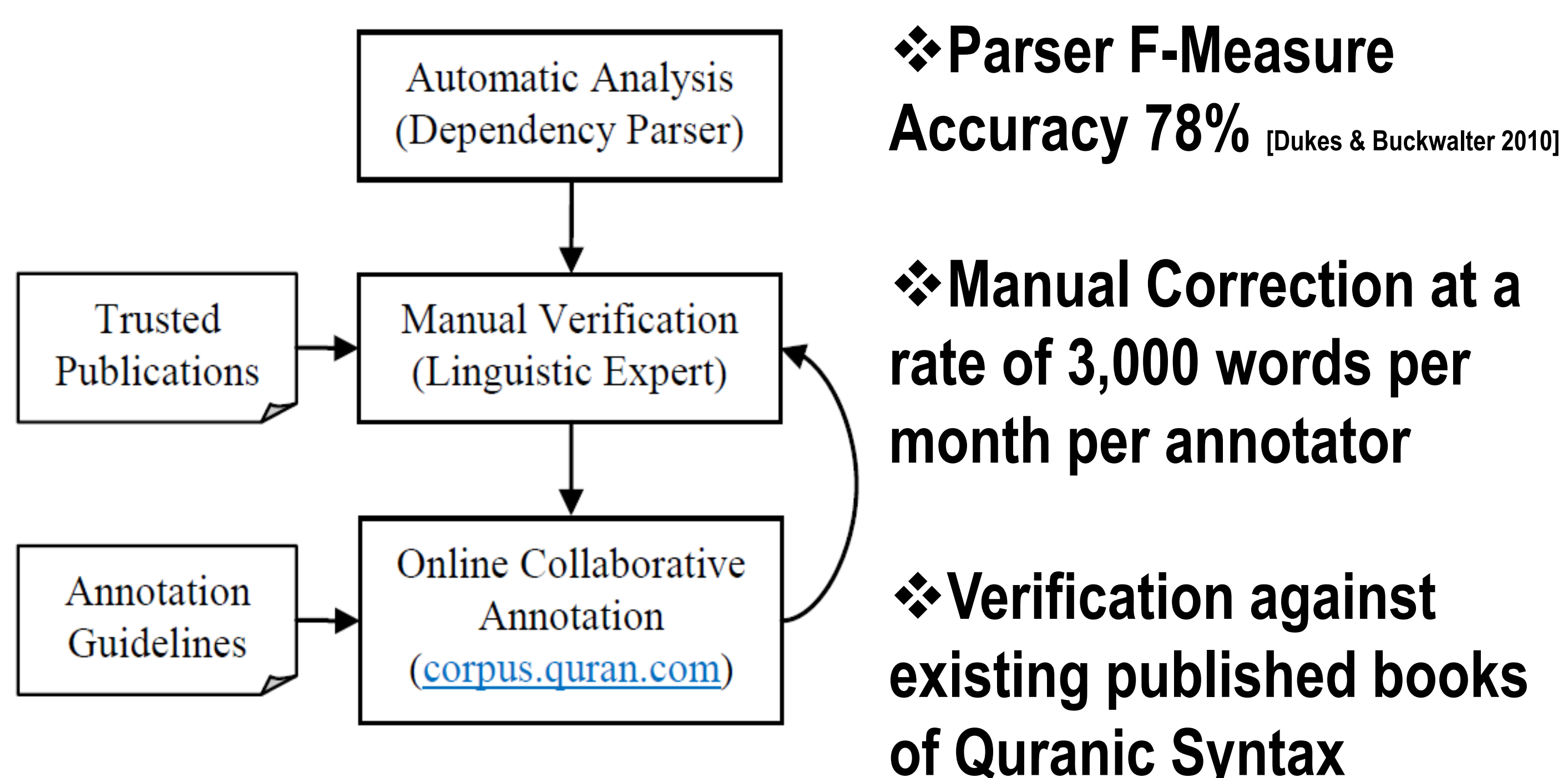
Traditional Arabic Grammar and I'rab

- 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 [Kruijff 2006]
- 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>

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

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.
- Kruijff, G. (2006) "Dependency Grammar" The Encyclopedia of Language and Linguistics 2nd Ed. Elsevier.