Textual Entailment for Modern Standard Arabic

By: **Maytham Alabbas,** Formal Methods Group.   Supervisor: **Prof. Allan Ramsay**

The aim of project

To determine whether one Modern Standard Arabic sentence **entails** another using a 'Textual Entailment'-based approach.

Motivations

- Text entailment (TE) can be looked upon as mapping between variable language forms.
- Mapping is possible at lexical, syntactic and semantic levels of the language.
- TE is considered as a framework for other NLP applications like Question Answering, Summarization, …etc.

Entailment

*A text T **entails** a hypothesis H iff every situation that makes T true, makes H true* [1].

T1: The couple is divorced. entails
H1: The couple was married.

T2: No student came to class early. does not entail
H2: No student came to class.

Logical Entailment (see Figure 1)

Difficulties [2]:
- The translation of natural sentences into logic is difficult because of issues, such as **ambiguity** and **extragrammaticality**.
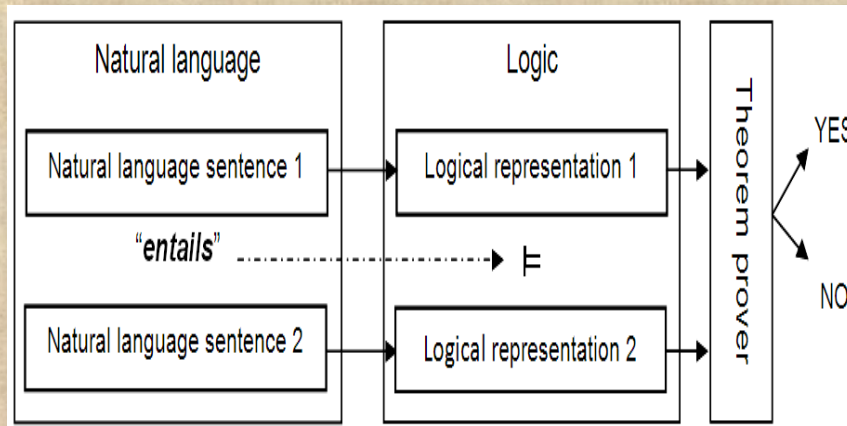- It needs vast additional knowledge (e.g. about word meaning), also it takes a lot of computation.

Figure 1: logical entailment technique

Textual Entailment

- **Textual entailment** is concerned with developing *approximate inference techniques* for natural language, using inference rules based on directly matching dependency trees and fragments of dependency trees.

Example rules:
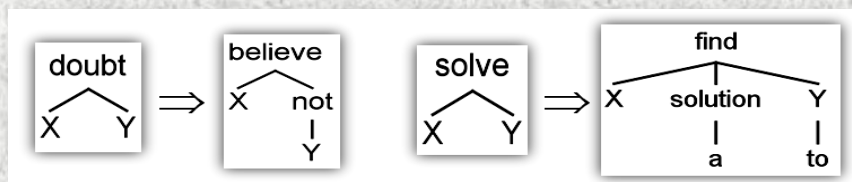
These can be hand-coded, but are often obtained by machine learning.

Modern Standard Arabic (MSA)

MSA is massively more ambiguous than English.
- The lack of diacritics (see Figure 2).
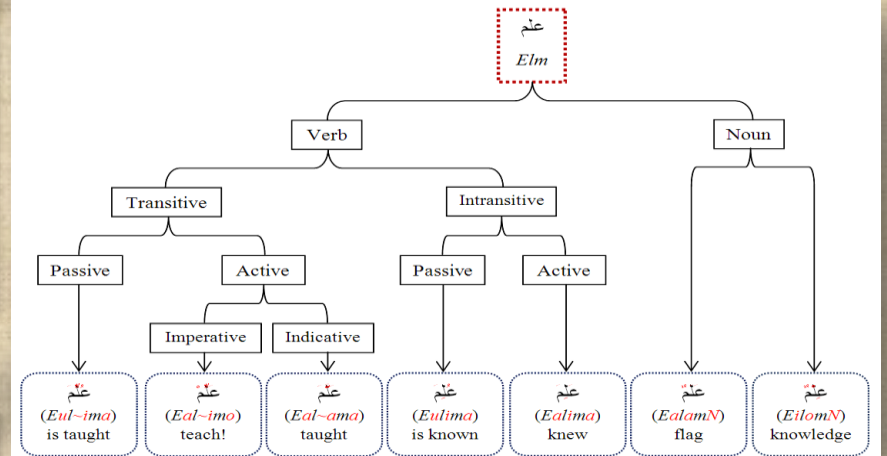- Free word order. ▪ Zero items(e.g., copulas)

Figure 2: ambiguity caused by the lack of diacritics

Current technique (see Figure 3)

Arabic linguistics analysis:
- Create dependency tree for both *T-H*.

Forward inference rules :
Expand *H* using syntactic templates, e.g.
X travel to $Y \Rightarrow X$ visit Y
X finds a solution to $Y \Rightarrow Y$ is solved by X

Structural rules:
- Find the minimum distance between two trees.
- Find the best sequence of editing operations (delete, insert and rename) for both nodes/subtrees.
- Determine cost function for dependency tree edit operations, including using hyponym rules.
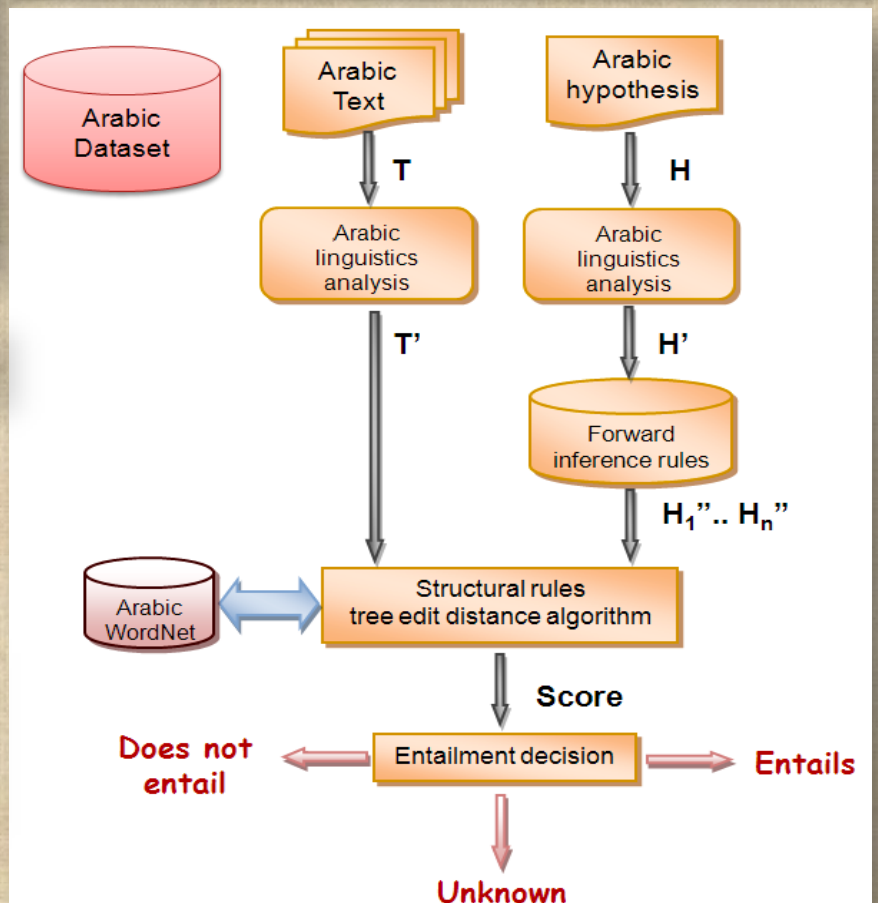

Figure 3: current technique

[1]. Chierchia, G., and McConnell-Ginet, S. (2001). *Meaning and grammar: An introduction to semantics*: The MIT Press.
[2]. Blackburn, P., Bos, J., Kohlhase, M., and de Nivelle, H. (2001) Inference and computational semantics. *Studies in Linguistics and Philosophy, Computing Meaning* 77: 11-28.