No single exhaustive list of Arabic discourse connectives exist for MSA.

Collection Procedure:

(i) Manual extraction of all connectives from Arabic literature, 50 randomly selected texts from the Penn Arabic Treebank and from 10 different web articles.

(ii) Automatic extraction of all text items with connective-specific POS tags (such as conjunctions) from the Penn Arabic Treebank.

(a) Manual verification by two Arabic native speakers.

We concentrate in this first study for Arabic on signalled relations.

Applications of Discourse corpus:

• Automatic summarization
• Question answering
• Text generation.

We followed the annotation principles in the PDTB (The English project of annotating discourse connectives in PTB, Prasad et al., 2008) as far as possible.

The Main attractions of PDTB approach:

• PDTB annotation is theory-neutral.

• It has also been shown to be extensible to other languages.

We concentrate in this first study for Arabic on signalled relations.

Necessary adaptations were made to take into account properties specific to Arabic.

We introduce two new relations at subcategorial level as we found them necessary in our pilot annotation for Arabic. These are EXPANSION.Background and COMPARISON.Similarity.

They had expected that the task would be taken up by a manager who has a more important record regarding the international (just as) the Saudi union did before.

We consider any text segments expressing abstract objects as arguments. They might be one or more of:

• Tensed or untensed verbal sentences or clauses prenex (prenex)
• Noun sentences prenex (prenex)
• Anaphoric expressions
• Verbs.

Al-Mašur (the nominalization):

• Al-Mašur is a well-defined noun category expressing an event.

• All morphological forms such as fīra'm/ašur/mašur is the Al-Mašur form of fīra'm/ašur/mašur.

• The common noun phrase play as an argument in LADTB.

• Interestingly what Al-Mašur in Arabic is not necessarily a nominalization in English.

For example:

The equivalent of agriculture (في الزراعة) is Al-Mašur form in Arabic, namely الزراعة. However, it is not a nominalization in English.

We consider any text segments expressing abstract objects as arguments.

CONCLUSION:

The Leeds Arabic Discourse Treebank (LADTB):

The Leeds Arabic Discourse Treebank (LADTB) is a gold standard corpus for Arabic discourse.

The Leeds Arabic Discourse Treebank (LADTB) is the first gold standard Arabic discourse treebank.

The Leeds Arabic Discourse Treebank (LADTB) is an important tool for research into the comparison of Arabic and English discourse.

The Leeds Arabic Discourse Treebank (LADTB) is a gold standard corpus for Arabic discourse.

The Leeds Arabic Discourse Treebank (LADTB) is the first gold standard Arabic discourse treebank.

The Leeds Arabic Discourse Treebank (LADTB) is an important tool for research into the comparison of Arabic and English discourse.

The Leeds Arabic Discourse Treebank (LADTB) is a gold standard corpus for Arabic discourse.

The Leeds Arabic Discourse Treebank (LADTB) is the first gold standard Arabic discourse treebank.

The Leeds Arabic Discourse Treebank (LADTB) is an important tool for research into the comparison of Arabic and English discourse.