U LISBOA | UNIVERSIDADE DE LISBOA

MWEs in the QTLeap Corpus of Online Helpdesk Interactions

Andreia Querido, Rita de Carvalho, João Rodrigues, António Branco

Annotation of a corpus from a specific domain: online helpdesk interactions (as part of QTLeap Project);
- Double-blind annotation followed by adjudication.

Corpus characteristics:
- 9959 sentences in Portuguese;
- Real interactions between clients and experts in a chat channel.

Theoretical guidance: Sag et al., 2002

Only 3 subclasses of MWEs have ocurences in this corpus:

Distribution of MWEs		
Fixed Expressions	**Compound Nominals**	**Light Verbs**
287 occurrences	799 occurrences	46 occurrences
25%	71%	4%
87 types	116 types	13 types