

Indice

Introduzione	7
I. I fondamenti della linguistica dei corpora	13
1. Premesse teoriche	13
2. Il corpus nella tradizione	22
3. Il corpus elettronico	29
II. Struttura di un corpus	43
1. Il corpus come campione	43
2. I tratti definitivi di un corpus	52
III. Codifica e annotazione	73
1. Codifica del carattere e markup testuale	73
2. Le unità linguistiche e la loro annotazione	84
IV. L'analisi dei corpora: strumenti di base	109
1. La distribuzione del lessico	109
2. Analisi delle concordanze	118
3. Associazioni lessicali: le collocazioni	126

V. La situazione italiana	141
1. I centri italiani per la linguistica dei corpora e le principali risorse scritte	141
2. I corpora di italiano parlato	155
<hr/>	
Quadri	
2.1. Corpora e questioni legali	69
3.1. La tokenizzazione	87
3.2. Le espressioni polirematiche	91
3.3. Il formato di trascrizione CHAT-LABLITA	100
3.4. La pragmatica e lo studio del parlato	104
4.1. Vocabolario del corpus e <i>type/token ratio</i>	111
4.2. Classi aperte e classi chiuse	115
4.3. Ricerche grammaticali <i>corpus-based</i>	137
<hr/>	
Appendice	179
<hr/>	
Riferimenti bibliografici e sitografia	187
<hr/>	
Indice analitico	207