

CONTENTS

1. Introduction	9
Part I. THE RULES OF THE GAME	15
2. Preliminaries	17
2.1. Frequencies and the Zipf's Law	17
2.2. Basics of Probability Calculus	18
2.2.1. <i>Probability Estimation</i>	19
2.2.2. <i>N-gram Models</i>	20
2.2.3. <i>Probabilistic Graphical Models</i> [◊]	21
2.3. Fundamental Paradigms: Rule-Based and Machine Learning Models . .	27
2.3.1. <i>Rule-Based Systems</i>	28
2.3.2. <i>Machine-Learning Systems</i>	28
3. Deep Neural Networks	35
3.1. The basic unit	35
3.1.1. <i>Biological Neuron Functionality</i>	35
3.1.2. <i>The Artificial Neuron</i>	36
3.2. Networks of Neurons: Basic Architectures	37
3.2.1. <i>Multilayer Perceptrons</i>	38
3.2.2. <i>Convolutional Networks</i>	42
3.2.3. <i>Recurrent Networks</i>	44
3.3. Networks of Neurons: Advanced Architectures	47
3.3.1. <i>Encoder/Decoder Networks</i>	47
3.3.2. <i>Transformers</i>	49

4. Vector Representations (a.k.a. Embeddings)	51
4.1. Count-based methods	52
4.2. Don't count, predict!	54
4.3. Contextual Word-embeddings	57
4.4. Sentence/Text Embeddings	59
4.5. A final consideration on vector representations	62
5. Evaluation: the Will and the Way	65
5.1. System Evaluation	65
5.1.1. <i>Instance Classification - Discrete Outputs</i>	65
5.1.2. <i>Textual Output</i>	67
5.1.3. <i>Interactive Systems</i>	67
5.2. Steps in the evaluation process	67
5.3. International Evaluation Challenges	68
5.4. An actual problem: SOTA chasing	69
Part II. THE AUTOMATIC ANALYSIS OF ITALIAN	73
6. Focus on Words	75
6.1. Italian “BERTology”	75
6.2. Morphological Analysis	76
6.2.1. <i>Italian Morphology</i>	76
6.2.2. <i>Computational tools to handle Italian Morphology</i>	77
6.2.3. <i>The AnIta Morphological Analyser</i>	78
6.2.4. <i>Evaluating AnIta Coverage</i>	83
6.3. Part-of-Speech Tagging	87
6.3.1. <i>Yes, but, which categories should we use?</i>	88
6.3.2. <i>PoS-tagging benchmarks for Italian</i>	93
6.3.3. <i>PoS-Tagging with Stochastic Models</i>	94
6.3.4. <i>Neural PoS-Taggers</i>	98
6.4. Lemmatisation	102
6.4.1. <i>The AnIta Lemmatiser</i>	105
6.4.2. <i>AnIta Lemmatiser Results at EVALITA 2011</i>	105
6.4.3. <i>A Neural Lemmatiser</i>	107
6.5. Named Entity Recognition	108
6.6. Word Sense Disambiguation	110
6.6.1. <i>Baseline methods for WSD</i>	110
6.6.2. <i>A Quantum-Like Approach to Word Sense Disambiguation</i>	112
6.7. Frame Semantic Parsing	124
6.7.1. <i>Electra-AGE_FE System Architecture</i>	126
6.7.2. <i>The IFrameNet project</i>	134
6.7.3. <i>Electra-AGE_FE on Italian</i>	135
7. Sentence and Text Analysis	137
7.1. Syntactic Analysis	137

7.1.1.	<i>General Parsing Strategies</i>	138
7.1.2.	<i>Neural Dependency Parsers</i>	139
7.1.3.	<i>Italian Treebanks</i>	145
7.1.4.	<i>Neural Parsers Evaluation</i>	146
7.1.5.	<i>An Ensemble of Neural Parsers</i>	149
7.1.6.	<i>Transformer-based Parsing</i>	160
7.2.	Two words on Neural Machine Translation	162
7.3.	Text Classification: Opinions, Polarity, Irony and Emotions	164
7.3.1.	<i>Sentiment Analysis</i>	164
7.3.2.	<i>Hate Speech Detection</i>	166
7.3.3.	<i>Emotion Detection</i>	166
7.4.	Text Generation/Summarisation	171
8.	Let the Speech Speak!	179
8.1.	Automatic Speech Recognition	179
8.1.1.	<i>Italian Spoken Corpora for ASR</i>	180
8.1.2.	<i>State of the Art for Italian ASR</i>	181
8.1.3.	<i>NVIDIA NeMo ASR</i>	182
8.1.4.	<i>Model setup and Results</i>	184
8.2.	Pitch Tracking	187
8.2.1.	<i>A Neural PDA Smoother</i>	188
8.2.2.	<i>Experiments and Results</i>	189
8.3.	Prosodic Prominence Identification	192
8.3.1.	<i>A Rule-based System</i>	198
8.3.2.	<i>Solution based on Machine/Deep-Learning</i>	200
8.3.3.	<i>A reflection on Evaluation</i>	208
8.4.	Pathological Language Analysis	210
8.4.1.	<i>Linguistic Features as Digital Biomarkers</i>	211
8.4.2.	<i>Mild Cognitive Impairment Detection</i>	219
8.4.3.	<i>Anorexia Nervosa & Developmental Language Disorder detection</i>	227
Part III.	SURFING ON “THE NEXT WAVE”	229
	Bibliography	237
	Acknowledgements	269
	Index	270
	Acronyms	273