

CONTENTS

Abstract / Resum / Resumen / Résumé	iii
Contents	vii
1 Motivation, objectives and structure	1
1.1 Motivation	1
1.2 Objectives	3
1.3 Structure of this PhD thesis	5
2 Preliminary matters	7
2.1 Machine learning and natural language processing	7
2.1.1 Basic definitions: ASR, MT, NLP	7
2.1.2 Machine learning, deep learning and neural networks	8
2.2 Automatic speech recognition	9
2.3 Machine translation	9
2.3.1 Introduction	10
2.3.2 Phrase-based machine translation	11
2.3.3 Neural machine translation	11
2.3.4 This PhD thesis in the context of the MT state of the art	12
2.4 Evaluation of ASR quality	13
2.5 Evaluation of MT quality	14
2.5.1 Introduction	14
2.5.2 Automatic evaluation	15
2.5.3 Human evaluation	18
2.5.4 Relation between automatic and human evaluations in MT	20
2.6 Evaluating time and space efficiency	21
2.7 Software for ASR and MT used in this work	22
3 Evaluation and post-editing of automatic transcriptions and translations of Open Educational Resources	25
3.1 Introduction. Evaluation framework	26
3.1.1 The project: transLectures	28
3.1.2 The pilot: poliMédia	29
3.1.3 The poliMédia datasets for ASR and MT	30
3.1.4 Automatic transcription and translation systems	31
3.1.5 Types of evaluation	34
3.2 Automatic evaluations	36
3.3 Human evaluations by language experts	37

3.3.1	Human evaluations by language experts, transLectures Year 1	38
3.3.2	Human evaluations by language experts, transLectures Year 2	40
3.3.3	Conclusions on human evals. by language experts in transLectures	42
3.4	Human evaluations by users	43
3.4.1	Evaluating intelligent interfaces for post-editing automatic transcriptions of online video lectures (transLectures Y2)	44
3.4.2	Human evals. of auto. transc. and translations by users (transLectures Y3)	53
3.5	Conclusions	59
3.5.1	Publications and contributions by the author	62
4	The transition from phrase-based to neural machine translation of Open Educational Resources	65
4.1	Introduction	66
4.1.1	The EU project X5gon	66
4.1.2	WMT: The Conference on Machine Translation	68
4.2	Datasets	68
4.2.1	WMT17 bilingual and monolingual datasets	69
4.2.2	WMT18 bilingual and monolingual datasets	71
4.3	The MLLP-UPV German-English MT System for WMT18	74
4.3.1	Data preparation	75
4.3.2	System description	77
4.3.3	Experimental evaluation	77
4.3.4	Global results of the WMT18 DE→EN News Task	81
4.3.5	Conclusions for WMT18	82
4.4	Comparing Phrase-Based and Neural MT systems	83
4.4.1	Experimental setup	83
4.4.2	System descriptions	83
4.4.3	System results	86
4.4.4	System results comparison	88
4.5	Impact of NMT systems on real OER scenarios	90
4.5.1	X5gon NMT systems	91
4.5.2	X5gon NMT system results	91
4.6	Conclusions	93
4.6.1	Publications and contributions by the author	94
5	Speech data filtering and verbatimization for streaming automatic speech recognition of parliamentary debates	95
5.1	Introduction	96
5.2	Creating the Europarl-ASR corpus: data gathering, definition of ASR tasks and evaluation sets	98
5.2.1	Data gathering and selection	98
5.2.2	Data partition and task definition	99
5.2.3	Manual revision of transcriptions	101
5.3	Speech data filtering and speech data verbatimization	101

5.4	Baseline experiments and results	103
5.5	Conclusions	105
5.5.1	Publications and contributions by the author	105
6	Live and offline computer assisted subtitling of broadcast media	107
6.1	Introduction and context	108
6.1.1	Public broadcasting and media accessibility in Spain and in València: À Punt Mèdia	109
6.1.2	The IberSpeech-RTVE Speech-to-Text Challenge 2018 and 2021: the benchmark for automatic subtitling of media contents in Spain	111
6.1.3	The À Punt–UPV R&D collaboration agreement for real-time computer assisted subtitling of media contents	113
6.2	Preliminary concepts: media accessibility and NLP	115
6.2.1	Measures for media accessibility	115
6.2.2	Media subtitling: live versus pre-recorded	117
6.3	Accessibility and subtitling for the deaf and hard-of-hearing at À Punt	120
6.3.1	The SDH workflow at À Punt	120
6.3.2	SDH for pre-recorded broadcasts at À Punt	121
6.3.3	SDH for live broadcasts at À Punt	122
6.4	Datasets	122
6.4.1	Reference evaluation datasets	122
6.4.2	Complete evaluation datasets	125
6.4.3	Complete training datasets	126
6.5	Description of ASR systems	128
6.5.1	ASR system architecture	129
6.5.2	Decoding strategy	130
6.5.3	Initial systems	131
6.5.4	Improved systems	133
6.6	Experimental results	135
6.6.1	Experimental setup	135
6.6.2	Catalan ASR system results	136
6.6.3	Spanish ASR system results	137
6.6.4	Initial comparison with a commercial ASR provider	138
6.7	Integration of ASR systems for media subtitling in À Punt	139
6.7.1	Software for the integration of ASR systems for media subtitling	139
6.7.2	The À Punt computing cluster for computer assisted subtitling	140
6.7.3	Integration of ASR systems for pre-recorded media subtitling .	141
6.7.4	Integration of ASR systems for live media subtitling	142
6.7.5	Postprocessing ASR for media subtitling	142
6.8	Conclusions	143
6.8.1	Publications and contributions by the author	143

7 Achievements and conclusions	145
7.1 Achievements	145
7.2 Contribution to the UN Sustainable Development Goals	150
7.3 Final conclusions	151
Bibliography	157
List of Abbreviations	181
List of Figures	183
List of Tables	185
Agraïments / Acknowledgements	187
Funding acknowledgements	189
A Automatic evaluation of MT quality: further details	191
A.1 Preliminary concepts: n-grams	191
A.2 BLEU: further details	192
B Tools for automatic speech recognition and machine translation	193
B.1 Hardware for ASR and MT	194
B.2 Software for ASR and MT	195
B.2.1 Automatic speech recognition software	195
B.2.2 Neural MT and NLP software	196
B.2.3 Phrase-based machine translation software	196
B.2.4 Language modelling software	197
B.2.5 Deep learning frameworks	199
B.2.6 Deep learning GPGPU frameworks and libraries	199
B.2.7 Automatic speech recognition evaluation software	200
B.2.8 Machine translation evaluation software	201