

D1.24
Report on the
Luxembourgish Language

Author	Dimitra Anastasiou
Dissemination level	Public
Date	28-02-2022



About this document

Project	European Language Equality (ELE)
Grant agreement no.	LC-01641480 – 101018166 ELE
Coordinator Co-coordinator	Prof. Dr. Andy Way (DCU)
Start date, duration	Prof. Dr. Georg Rehm (DFKI) 01-01-2021, 18 months
<u> </u>	1
Deliverable number	D1.24
Deliverable title	Report on the Luxembourgish Language
Туре	Report
Number of pages	24
Status and version	Final
Dissemination level	Public
Date of delivery	Contractual: 28-02-2022 – Actual: 28-02-2022
Work package	WP1: European Language Equality – Status Quo in 2020/2021
Task	Task 1.3 Language Technology Support of Europe's Languages in
A 4 la	2020/2021
Author Reviewers	Dimitra Anastasiou
Editors	Annika Grützner-Zahn, Khalid Choukri Maria Giagkou, Stelios Piperidis, Georg Rehm, Jane Dunne
EC project officers	Susan Fraser, Miklos Druskoczi
Contact	European Language Equality (ELE)
	ADAPT Centre, Dublin City University
	Glasnevin, Dublin 9, Ireland
	Prof. Dr. Andy Way – andy.way@adaptcentre.ie
	European Language Equality (ELE)
	DFKI GmbH
	Alt-Moabit 91c, 10559 Berlin, Germany
	Prof. Dr. Georg Rehm – georg.rehm@dfki.de
	http://www.european-language-equality.eu
	© 2022 ELE Consortium



Consortium

1 Dublin City University (Coordinator) DCU IE 2 Deutsches Forschungszerturm für Künstliche Intelligenz GmbH (Co-coordinator) DFKI DE 3 Univerzita Karlova (Charles University) CUNI CZ 4 Athina-Fernviltice Kentro Kainotomias Stis Technologies Tis Pliroforias, Ton Epikoinonion Kai Tis Gnosis ILSP GR 5 Universidad De Pais Vasco/ Euskal Herriko Unibertsitatea (University of the Basque Country) UPWEHU ES 6 CROSSLANG W EUROPean Federation of National Institutes for Language ERNII LU 7 European Federation of National Institutes for Language Equality Network) ELRY FEX.N 8 Réseau européen pour Pégalité des langues (European Language Equality Network) ELRY FEX.N 9 European Civil Society Platform for Multilingualism ELRY FEX.N N. 10 CLARIN ERIC - Common Language Resources and Technology Infrastructure as a European Research CLARIN N. 11 Université Leiden (University of Leiden) ULE N. 12 Leves com (European Institute for Research and Strategic Studies in Telecommunications GmbH EESCM				
Indiverzita Karlova (Charles University) CUIN CZ	1	Dublin City University (Coordinator)	DCU	IE
4. Athina-Frevnitiko Kentro Kainotomias Sits Technologies Tis Pitroforias, Ton Epikoinonion Kai Tis Gnoss ILIVPX-ILI RS 5. Indiversidad Del Pais Vasco/ Euskal Herriko Unibertsitatea (University of the Basque Country) UPVX-ILI RS 6. CROSSLANG NY European Federation of National Institutes for Language ELL UNIVERSITY TR 7. European Cival Society Platform for Multilingualism ELEN TR 8. European Cival Society Platform for Multilingualism CLARIN REIC - Common Language Resources and Technology Infrastructure as a European Research Infrastructure Consortium CLARIN REIC - Common Language Resources and Technology Infrastructure as a European Research Infrastructure Consortium NL LEI NL LEI NL LEI NL 10. Universitie Lieden (University of Leiden) ULIE NL LIBER NL NL NL LIBER NL NL NL LIBER NL NL NL NL LIBER NL NL NL NL NL NL LIBER NL NL NL NL LIBER NL NL NL NL NL NL LIBER NL NL LIBER NL NL	2	Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (Co-coordinator)	DFKI	DE
5 Universidad Del Pais Vasco/ Euskal Herriko Unibertsitatea (University of the Basque Country) UPUVEHU ES 6 CROSSLAMG NY ELTOPEAN FEDERAL (SESLAMG NY ELTOPEAN FEDERAL (SESLAMG NY ELTOPEAN FEDERAL (UNIVERSITY OF MULTIME AUGUST OF	3	Univerzita Karlova (Charles University)	CUNI	CZ
6 CROSSLANG NV CRSIAGG BE 7 European Federation of National Institutes for Language ELIN L IV 8 Réseau européen pour l'égalité des langues (European Language Equality Network) ELEN PR 9 European Civil Society Platform for Multilingualism CLARIN RINGUE 10 CLARIN ERIC - Common Language Resources and Technology Infrastructure as a European Research CLARIN NI 11 Universiteit Léclén (University of Leiden) ULEI NL 12 Eurescom (European Institute for Research and Strategic Studies in Telecommunications GmbH) ERSCM DE DE 14 Wikimedia Deutschland (Gesellschaft zur Förderung freien Wissens e. V.) WMD DE LU LIBER (Association of European Research Libraries) LIBER (Association of European Research Libraries) LIBER (Association of European Research Libraries) MWD DE LU LIBER (Association of European Research Libraries) MWD DE LU LIBER (Association of European Research Libraries) MWD DE LU LIBER (Association of European Research Libraries) MWD DE LIBER (Association of European Research Libraries) MWD DE LIBER (Association Association of European Research Libraries) THIDE LIBER (Association Association Research European Research Libraries) LIBER (Association Association Research Research Libraries) FEAR SASOCIA RESEARCH RESEARCH RESEARCH R	4	Athina-Erevnitiko Kentro Kainotomias Stis Technologies Tis Pliroforias, Ton Epikoinonion Kai Tis Gnosis	ILSP	GR
7 Duropean Federation of National Institutes for Language Expentation of National Institutes for Language Equality Network) ELLN ELLN Bear Bear européen pour l'égalité des language (European Language Equality Network) ELLN NB 10 CLARIN ERIC - Common Language Resources and Technology Infrastructure as a European Research Infrastructure Consortium ULEI NL 11 Universiteit Leiden (University of Leiden) ULEI LIBRO NL 12 Eurescone (European Institute for Research and Strategic Studies in Telecommunications GmbH) ERSCM DE 13 Stichting LIBER (Association of European Research Libraries) UMD DE 14 Wikinedia Deutschland (Geselbschaft aur Fröderung Freien Wissens e. V.) WMD DE 15 Tilde SLA LENDA (ELDA) ELDA FR 16 Evaluations and Language Resources Distribution Agency ELDA FR EXpert System Iberia SL EXPSYS ES 17 Expert System Iberia SL Vesign Statute SL EXPSYS EX EXPSYS EX 18 HENDOLDT Analytics GmbH SCA AT SCA AT SCA <td< td=""><td>5</td><td>Universidad Del Pais Vasco/ Euskal Herriko Unibertsitatea (University of the Basque Country)</td><td>UPV/EHU</td><td>ES</td></td<>	5	Universidad Del Pais Vasco/ Euskal Herriko Unibertsitatea (University of the Basque Country)	UPV/EHU	ES
8 Réseau européen pour l'égalité des langues Européan Language Equality Network) ELEN FR 9 European Civil Society Platform for Multilingualism CLARIN ERIC – Common Language Resources and Technology Infrastructure as a European Research CLARIN NI 10 CLARIN ERIC – Common Language Resources and Technology Infrastructure as a European Research ULEI NI 11 Universiteit Leiden (University of Leiden) ULEI NI 12 Eurescom (European Institute for Research and Strategic Studies in Telecommunications GmbH) ERESCM DE 14 Wikimedia Deutschland (Gesellschaft zur Förderung freien Wissens e. V.) WMD DE 15 EILED FILE IX EXPEST ELDE IX 16 Evaluations and Language Resources Distribution Agency ELDA R EXPEST ES 17 Expert System Beria SL EXPEST System Beria SL EXPEST System Beria SL EXPEST SE EXPEST ES 18 HENSOLDT Analytics GmbH EXPEST System Beria SL EXPEST System Beria SL EXPEST SW EXPEST SE 18 HENSOLDT Analytics GmbH EXPEST SW EXPEST SW <td>6</td> <td>CROSSLANG NV</td> <td>CRSLNG</td> <td>BE</td>	6	CROSSLANG NV	CRSLNG	BE
Buropean Civil Society Platform for Multilingualism	7	European Federation of National Institutes for Language	EFNIL	LU
10 CLARIN ERIC - Common Language Resources and Technology Infrastructure as a European Research Universitiet Lieden (University of Leiden) VILEI NL 11 Universitiet Lieden (University of Leiden) LIBER NL 12 Eurscom (European Institute for Research and Strategic Studies in Telecommunications GmbH) ERSCM DE 13 Stichting LIBER (Association of European Research Libraries) LIBER NL 14 Wikimedia Deutschland (Gesellschaft zur Förderung freien Wissens e. V.) WMD DE 16 Evaluations and Language Resources Distribution Agency ELDA FR 17 Experts System Iberia SI. EXPSYS ES 18 HENSOLDT Analytics GmbH HENS AT 19 Xcelerator Machine Translations Ltd. (KantanMT) KNTN IK 20 PANGEANIC-B. I. Europa SLU RSW AT 21 SIRMA AI EAD (Intotext) ONTO BG 23 SAP SE Universität Wien (University of Venna) UVE AT 24 Universiteit Antwerpen (University of Antwerp) UVE AT 25 Institute for Bulgarian Language "Prof. Lyubomir Andreychin" IBL BG 26 Institute for Bulgarian Language "Prof. Lyubomir Andreychin" IBL BG 27 Sevuečil	8	Réseau européen pour l'égalité des langues (European Language Equality Network)	ELEN	FR
Infrastructure Consortium of University of Leiden) 1 Universitier Leiden (University of Leiden) 2 Eurescom (European Institute for Research and Strategic Studies in Telecommunications GmbH) ERSCM DE Stichting LIBER (Association of European Research Libraries) LIBER NL UNIVERSE (Association of European Research Libraries) LIBER NL UNIVERSE (Association of European Research Libraries) Title SIA TILDE LV Expert System Iberia SI. Expert System	9		ECSPM	DK
12Eurescom (European Institute for Research and Strategic Studies in Telecommunications GmbH)ERSCMDE13Stichting LIBER (Association of European Research Libraries)NLNL14Wikimedia Deutschland (Gesellschaft zur Förderung freien Wissens e.V.)TILDELV15Tilde SIATILDELV16Evaluations and Language Resources Distribution AgencyELDATILDELV17Expert System Iberia SI.EXPSYSES18HENSOLDT Analytics GmbHKNTNIE20PANGEANIC-B. I. Europa SLUPANES21Separt Sille Web Company GmbHSCAT22SIRMA AI EAD (Ontotext)ONTOBG23SAP SEONTOBG24Universitiet Aintwerpen (University of Vienna)UVAAT25Universitiet Antwerpen (University of Antwerp)UANTWBE26Institute for Bulgarian Language "Prof. Lyubomir Andreychin"IBLBG27Seveučilište u Zagrebu Filozofski fakultet (Univ. of Zagreb, Faculty of Hum. and Social Sciences)FFZGHR28Kobenhavus Universitet (University of Copenhagen)UCPHDK27Tartu Ulikool (University of Helsinki)UTARTEF28Contre National de la Recherche ScientifiqueCNBNYTKHU30Stofhun Árna Magnússonar i Islenskum fræðum SAM (Árni Magnússon Inst. for Icelandic Studies)SAMIS41Luxingia University of Latvia)LKIILIL42Luxingi		Infrastructure Consortium		
13Stichting LIBER (Association of European Research Libraries)LIBERNL14Wiklimedia Deutschland (Gesellschaft zur Förderung freien Wissens e.V.)WMDDE15Tilde SIATILDELIV16Evaluations and Language Resources Distribution AgencyELDAFR17Expert System Iberia SI.EKPSES18HENSOLDT Analytics GmbHHENSAT19ACcelerator Machine Translations Ltd. (KantanMT)KNTNER21Semantic Web Company GmbHSWCAT23SIRMA I EAD (Intotext)SWCAT24Universität Wien (University of Vienna)UVIEAT25Universität Wien (University of Antwerp)UVIEAT26Institute for Bulgarian Language "Prof. Lyubomir Andreychin"IBLBG27Sveučitiše u Zagreb Hilozofski fakultet (Univ. of Zagreb, Faculty of Hum. and Social Sciences)IFCAHE28Kobenhavns Universitet (University of Copenhagen)UCPHDK29Tartu Ulikool (University of Tartu)UCPHDK31Centre National de la Recherche ScientifiqueCNRSFR4Produzione Bruno KesslerKNRIS5Liviyas Universitates Matemātikas un Informātikas instituts (Institute of Mathematics and ComputeKNRIS6Liviyas University of LatviajLKILLLU7Luxembourg Institute of Science and TechnologyLKILCNORNO8Liviyus Kalbos Institutudas (Institute of Mala) <td></td> <td></td> <td></td> <td></td>				
14Wikimedia Deutschland (Gesellschaft zur Förderung freien Wissens e.V.)WMDDE15Tilde SIATILDELV6Expert System Iberia SLEXPSYSES17Expert System Iberia SLEXPSYSES18HENSOLDT Analytics GmbHHENSAT19Xcelerator Machine Translations Ltd. (KantanMT)KNTNIE20PANGEANIC-B.I. Europa SLUPANES21SEMAM AI EAD (Ontotext)ONTOBG22SIRMA AI EAD (Ontotext)ONTOBG23SAP SEONTOBG24Universitiet Wein (University of Vienna)UVIEAT25Universitiet Antwerpen (University of Antwerp)UVIEAT26Institute for Bulgarian Language Prof. I Lyubomir Andreychin"IBLBG27Sveucilište u Zagrebu Filozofski fakultet (Univ. of Zagreb, Faculty of Hum. and Social Sciences)FFZGHR28Københavns University of Tartu)UTARTEE30Helsingin Yilopisto (University of Helsinki)UTHARTEF31Centre National de la Recherche ScientifiqueNYTKHI32Nyelvtudományi Kutatóközpont (Research Institute for Linguistics)NYTKHI31Latvijas Universitates Matemátikas un Informátikas instituts (Institute of Mathematics and ComputerMCSLV32Luvijus Slivitus (Jabos Instituta Of Norway)LIXLV33Luviere, University of JatviaLVLVLV44Livituri (Jabos Institut		, 1		
15Tilde SIAITILDEIV16Evaluations and Language Resources Distribution AgencyELDAFR7Expert System Iberia SLEXPSYSES18HENSOLDT Analytics GmbHHENSAT20PANGEANIC-B.I. Europa SLURNTNIE21Semantic Web Company GmbHSWCAT22SERNA AI EAD (Ontotext)ONTOBG23SAP SESAPDE24Universitità Wien (University of Vienna)UVIEAT25Universitità Wien (University of Antwerp)UANTWBE26Institute for Bulgarian Language "Prof. Lyubomir Andreychin"IBLBG27Sveučilište u Zagrebu Fliozofski fakultet (Univ. of Zagreb, Faculty of Hum. and Social Sciences)FFZGHR28Kobenhavns University of TartuUCPHDK29Tartu Ulikool (University of Helsinki)UCPHDK31Centre National de la Recherche ScientifiqueCNRSFR32Nyelvtudományi Kutatóközpont (Research Institute for Linguistics)NYTKHU33Stoffund Arna Magnússonar í Islenskum fræðum SAM (Árni Magnússon Inst. for Icelandic Studies)SAMIS34Fondazione Bruno KesslerISLixLT35Lietuvij Kalbos Institutas (Institute of the Lithuanian Language)LKILT36Lietuvij Kalbos Institutas (Institute of the Lithuanian Language)LKILT37Liviversidade de Lisboa, Faculdade de Ciências (University of Lisbon, Faculty of Sciences)IPI		· · · · · · · · · · · · · · · · · · ·		
16Evaluations and Language Resources Distribution AgencyELDAFR17Expert System Iberia SLEXPSYSEX18HENSOLDT Analytics GmbHHENSAT19Xcelerator Machine Translations Ltd. (KantanMT)KNTNIE20PANGEANIC-B. I. Europa SLUPANES21Semantic Web Company GmbHSWCAT22SIRMA AI EAD (Ontotext)ONTOBG23SAP SEONTOBG24Universität Wien (University of Vienna)UANTWBE25Universität Wien (University of Antwerp)UANTWBE26Institute for Bulgarian Language "Prof. Lyubomir Andreychin"IBLBG27Sveučilište u Zagrebu Filozofski fakulter (Univ. of Zagreb, Faculty of Hum. and Social Sciences)FFZGHR28Kabenhavus Universitet (University of Copenhagen)UCPHDK29Tartu Ulikool (University of Tartu)UTARTEE30Helsingin Yilopisto (University of Helsinki)UHELFI41Centre National de la Recherche ScientifiqueNYTKHI42Nyelvtudományi Kutatóközpont (Research Institute for Linguistics)NYTKHI43Istofnun Árna Magnússonar í Íslenskur fræðum SAM (Árni Magnússon Inst. for Icelandic Studies)STLY46Lietuvij Kalbos Institutes of Matematikas un Informātika instituts (Institute of Mathematics and Computer Science, University of Latvia)LKILT47Luxembourg Institutes of Science and TechnologyLKILISTLU </td <td></td> <td>•</td> <td></td> <td></td>		•		
17Expert System Iberia SLEXPSYSES18I HENSOLDT Analytics GmbHHENSOLDT Analytics GmbHKNTNIE20PANGEANIC-B. I. Europa SLUPANES21Semantic Web Company GmbHSWCAT22SIRMA AI EAD (Ontotext)ONTOBG23SAP SESAPDE24Universität Wien (University of Vienna)UVIEAT25Universiteit Antwerpen (University of Antwerp)UANTWBE26Institute for Bulgarian Language "Prof. Lyubomir Andreychin"IBLBG27Swedilište u Zagrebu Flüozofski fakultet (Univ. of Zagreb, Faculty of Hum. and Social Sciences)FFZGHR28Københavns Universitet (University of Copenhagen)UCPHDK29Tartu Ulikool (University of Tartu)UTARTEF31Centre National de la Recherche ScientifiqueUHELFI32Nyelvtudományi Kutatóközpont (Research Institute for Linguistics)NYTKHU34Fondazione Bruno KesslerFBKIT35Latvijas Universitias Matematikas un Informátikas instituts (Institute of Mathematics and Computersity of Latvia)IMCSIV36Lietuvių Kalbos Institutas (Institute of the Lithuanian Language)LKILKILKI37Lietuvių Kalbos Institutas (Institute of Mathematics and Computersity of Malta)Università at Malta (University of Malta)Università at Malta (University of Malta)Università at Malta (University of Malta)INTNL38Vichituri Podstaw Informatyki P				
18HENSOLDT Analytics GmbHHENSAT19Xcelerator Machine Translations Ltd. (KantanMT)KNTNES20PANGEANIC-B. I. Europa SLUPANES21Semantic Web Company GmbHSWCAT22SIRMA AI EAD (Ontotext)ONTOBG23SAP SESAPDE24Universität Wien (University of Vienna)UVIEAT25Universiteit Antwerpen (University of Antwerp)UANTWBE26Institute for Bulgarian Language "Prof. Lyubomir Andreychin"IBLBG27Sveučilište u Zagrebu Filozofski fakultet (Univ. of Zagreb, Faculty of Hum. and Social Sciences)FEGHR28Københavns University of Tartu)UCPHDK29Tartu Ulikool (University of Tartu)UTARTEE30Helsingin Yilopisto (University of Helsinki)UHELFI31Centre National de la Recherche ScientifiqueCNRSFR32Stofnun Árna Magnússonar í íslenskum fræðum SAM (Árni Magnússon Inst. for Icelandic Studies)NYTKHU33Stofnun Árna Magnússonar í íslenskum fræðum SAM (Árni Magnússon Inst. for Icelandic Studies)LKILV34Livujas Universitátes Matematikas un Informátikas instituts (Institute of Mathematics and ComputerIMCSLV35Livujas Universitátes Matematikas un Informátikas instituts (Institute of Mathematics and ComputerIMCSLV36Livujujut Kalbos Institutas (Institute of the Lithuanian Language)LKIILIL37Luxembourg Inst		0 0		
19Xcelerator Machine Translations Ltd. (KantanMT)KNTNIE20PANGEANIC-B. I. Europa SLUPANSS21Semantic Web Company GmbHSWCAT22SIRMA AI EAD (Ontotext)ONTOBG23SAP SESAPDE24Universitât Wien (University of Vienna)UVIEAT25Universiteit Antwerpen (University of Antwerp)UANTWBE26Institute for Bulgarian Language "Prof. Lyubomir Andreychin"IBLBG27Sveülište u Zagrebu Filozofski fakultet (Univ. of Zagreb, Faculty of Hum. and Social Sciences)FEZGHB28Københavns Universitet (University of Copenhagen)UCPHDK29Tartu Ulikool (University of Tartu)UTARTEE30Helsingin Yliopisto (University of Helsinki)UHLIT31Centre National de la Recherche ScientifiqueCNRSFR32Nyelvtudományi Kutatóközpont (Research Institute for Linguistics)NYTKHU34Fondazione Bruno KesslerFBKIT35Latvijas Universitates Matemátikas un Informátikas institúts (Institute of Mathematics and ComputerIMMCIX36Lietuvių Kalbos Institutas (Institute of the Lithuanian Language)LKIIL37Luxembourg Institutu of Science and TechnologyLKIIN38Università ta Malta (University of Malta)UMMT39Stichting Instituut voor de Nederlandse Taal (Dutch Language Institute)INTINT40Spräkrådet (Langua		1 ,		
20PANGEANIC-B.I. Europa SLUESES21Semantic Web Company GmbHSWCAT22SIRMA AI EAD (Ontotext)ONTOBG23SAP SESAP SEUniversität Wien (University of Vienna)UVIEAT24Universiteit Antwerpen (University of Antwerp)UNATWBE25Institute for Bulgarian Language "Prof. Lyubomir Andreychin"IBLBG27Sveučilište u Zagrebu Filozofski fakultet (Univ. of Zagreb, Faculty of Hum. and Social Sciences)FFZGHR28Københavns Universitet (University of Copenhagen)UCPMDK29Tartu Ulikool (University of Tartu)UTARTEE30Helsingin Yilopisto (University of Helsinki)UHELTI31Centre National de la Recherche ScientifiqueCRRSFR32Nyelvtudományi Kutatóközpont (Research Institute for Linguistics)NYTKHU33Stofnun Árna Magnússonar í íslenskum fræðum SAM (Árni Magnússon Inst. for Icelandic Studies)NYTKHU34Latvijas Universitates Matemātikas un Informātikas instituts (Institute of Mathematics and ComputerIRCLV35Luxembourg Institute of Science and TechnologyLULU40Språkrådet (Language Council of Norway)LUNO41Università ta Malta (University of Malta)MANO42Universidade de Lisboa, Faculdade de Ciências (University of Lisbon, Faculty of Sciences)IPIPAP43Institutul de Cercetāri Pentru Inteligență Artificială (Romanian Academy)ICIA<				
21Semantic Web Company GmbHSWCAT22SIRMA AI EAD (Ontotext)ONTOBC23SAP SESAP SEUniversität Wien (University of Vienna)UVIEAT25Universiteit Antwerpen (University of Antwerp)UANTWBE26Institute for Bulgarian Language "Prof. Lyubomir Andreychin"IBLBG27Sveučilište u Zagrebu Filozofski fakultet (Univ. of Zagreb, Faculty of Hum. and Social Sciences)FFZGHR28Københavns Universitet (University of Copenhagen)UCPHDK29Tartu Ulikool (University of Tartu)UTARTEF31Centre National de la Recherche ScientifiqueCNRSFR32Nyelvtudományi Kutatóközpont (Research Institute for Linguistics)NYTKHU31Fondazione Bruno KesslerFBKIT32Latvijas Universitátes Matemátikas un Informátikas institúts (Institute of Mathematics and Computer Science, University of Latvia)LKILT33Listuvių Kalbos Institutas (Institute of the Lithuanian Language)LKILT34Livienvijas Università ta Malta (University of Malta)UMMT39Stichting Institutut voor de Nederlandse Taal (Dutch Language Institute)INTNL40Språkrådet (Language Council of Norway)LCNORNP41Institutul de Cercetari Pentru Inteligență Artificială (Romanian Academy)ICIARP42University of Cyprus, French and European StudiesICIARO43Institut Jožef Stefan (Jozef Stefan Institute) <t< td=""><td></td><td></td><td></td><td></td></t<>				
22SIRMA AI EAD (Ontotext)BG23SAP SESAPDE24Universität Wien (University of Vienna)UVIEAT25Universiteit Antwerpen (University of Antwerp)UANTWBE26Institute for Bulgarian Language "Prof. Lyubomir Andreychin"IBLBG27Sveučilište u Zagrebu Filozofski fakultet (Univ. of Zagreb, Faculty of Hum. and Social Sciences)FFZGHR28Københavns Universitet (University of Copenhagen)UCPHDK29Tartu Ulikool (University of Tartu)UTARTEE30Helsingin Yliopisto (University of Helsinki)UHELFI31Centre National de la Recherche ScientifiqueCNRSFR32Nyelvtudományi Kutatóközpont (Research Institute for Linguistics)NYTKHU33Stofnun Árna Magnússonar í Íslenskum fræðum SAM (Árni Magnússon Inst. for Icelandic Studies)SAMIS4Fondazione Bruno KesslerFEKIT5Lietuviy Kalbos Institutas (Institute of the Lithuanian Language)LKIIT4Lietuviy Kalbos Institutas (Institute of the Lithuanian Language)LKIIT3Stichting Institutu vor de Nederlandse Taal (Dutch Language Institute)INTN4Università ta Malta (University of Malta)UNN5Institut Jožef Stefan (Jozef Stefan Institutie)INTN4Universidade de Lisboa, Faculdade de Ciências (University of Lisbon, Faculty of Sciences)IPIPANP4University of Cyprus, French and European S		•		
23SAP SESAPDE24Universität Wien (University of Vienna)UVIEAT25Universiteit Antwerpen (University of Antwerp)UANTWBE26Institute for Bulgarian Language "Prof. Lyubomir Andreychin"IBLBG27Sveučilište u Zagrebu Filozofski fakultet (Univ. of Zagreb, Faculty of Hum. and Social Sciences)FFZGHR28Københavns Universitet (University of Copenhagen)UCPHDYUTARTEE30Helsingin Vliopisto (University of Tartu)UHELFI31Centre National de la Recherche ScientifiqueCNRSFR31Centre National de la Recherche ScientifiqueNYTKHU32Nyelvtudományi Kutatóközpont (Research Institute for Linguistics)SAMIS34Fondazione Bruno KesslerFBKIT35Lateityiajs Universitates Matemátikas un Informátikas instituts (Institute of Mathematics and Computer Indices University of Latvia)IMCSLV36Lietuvių Kalbos Institutas (Institute of the Lithuanian Language)LKILT37Luxembourg Institute of Science and TechnologyLKILT38Università ta Malta (University of Malta)UMMT39Stichting Institutu voor de Nederlandse Taal (Dutch Language Institute)INTNL40Språkrådet (Language Council of Norway)LCNORNO41Institutul de Cercetări Pentru Inteligență Artificială (Romanian Academy)ICIAR42University of Cyprus, French and European StudiesUCY<		• •		
24Universität Wien (University of Vienna)UVIEAT25Universiteit Antwerpen (University of Antwerp)UANTWBE26Institute for Bulgarian Language "Prof. Lyubomir Andreychin"IBLBE27Sveučilište u Zagrebu Filozofski fakultet (Univ. of Zagreb, Faculty of Hum. and Social Sciences)FFZGHR28Københavns Universitet (University of Copenhagen)UCPHDK29Tartu Ulikool (University of Tartu)UTARTEE31Centre National de la Recherche ScientifiqueCNRSFR32Nyelvtudományi Kutatóközpont (Research Institute for Linguistics)NYTKHU33Stofnun Árna Magnússonar í íslenskum fræðum SAM (Árni Magnússon Inst. for Icelandic Studies)SAMIT34Fondazione Bruno KesslerFBKIT35Latvijas Universitàtes Matemātikas un Informātikas instituts (Institute of Mathematics and Computer Science, University of Latvia)LKL36Lietwity Kalbos Institutas (Institute of the Lithuanian Language)LKLKL37Luxembourg Institute of Science and TechnologyLISTLU38Università ta Malta (University of Malta)UMMT40Språkrådet (Language Council of Norway)LCNORNO41Institutul doc Cercetări Pentru Inteligență Artificială (Romanian Academy)ICIARO42University of Cyprus, French and European StudiesUCYCY45Jazykovedný ústav Ludovita Stúra Slovenskej akadémie vied (Slovak Academy of Sciences)JULSSK <td></td> <td></td> <td></td> <td></td>				
25Universiteit Antwerpen (University of Antwerp)UANTWBE26Institute for Bulgarian Language "Prof. Lyubomir Andreychin"IBLBG27Sveučilište u Zagrebu Filozofski fakultet (Univ. of Zagreb, Faculty of Hum. and Social Sciences)FFZGHR28Københavns Universitet (University of Copenhagen)UCPHDK29Tartu Ulikool (University of Tartu)UTARTEE30Helsingin Yliopisto (University of Helsinki)UHELFI31Centre National de la Recherche ScientifiqueCNRSFR32Nyelvtudományi Kutatóközpont (Research Institute for Linguistics)NYTKHU33Stofnun Árna Magnússonar í íslenskum fræðum SAM (Árni Magnússon Inst. for Icelandic Studies)SAMIS4Fondazione Bruno KesslerFBKIT5Latvijas Universitetes Matemätikas un Informátikas institúts (Institute of Mathematics and ComputerIMCLV36Lietuvių Kalbos Institutas (Institute of the Lithuanian Language)LKIL37Luxembourg Institute of Science and TechnologyLKILT38Università ta Malta (University of Malta)UMMT39Stichting Institutut voor de Nederlandse Taal (Dutch Language Institute)INTLCNORNO40Spräkrådet (Language Council of Norway)LCNORNOINT41Institutul de Cercetaři Pentru Inteligență Artificială (Romanian Academy)ICIAICIATCIA42University of Cyprus, French and European StudiesUCYCY43 <td></td> <td></td> <td></td> <td></td>				
26Institute for Bulgarian Language "Prof. Lyubomir Andreychin"IBLBG27Sveučilište u Zagrebu Filozofski fakultet (Univ. of Zagreb, Faculty of Hum. and Social Sciences)FFZGHR28Københavns Universitet (University of Copenhagen)UCPHDVCPH29Tartu Ulikool (University of Tartu)UTARTEE30Helsingin Yliopisto (University of Helsinki)UHELFI31Centre National de la Recherche ScientifiqueCNRSFR31Stofnun Árna Magnússonar í íslenskum fræðum SAM (Árni Magnússon Inst. for Icelandic Studies)SAMIS34Fondazione Bruno KesslerFBKIT35Latvijas Universitates Matemātikas un Informātikas institūts (Institute of Mathematics and Computer Science, University of Latvia)LKILT36Lietuvių Kalbos Institutas (Institute of the Lithuanian Language)LKILT37Luxembourg Institute of Science and TechnologyLSTLU38Università ta Malta (University of Malta)UMMT39Stichting Instituut voor de Nederlandse Taal (Dutch Language Institute)INTNL40Språkrådet (Language Council of Norway)LCNORNO41Institutul de Cercetäri Pentru Inteligență Artificială (Romanian Academy)ECU, isonFCULisbonFCU43Institutul de Cercetări Pentru Inteligență Artificială (Romanian Academy)ICIARC44Universidade de Lisboa, Faculdade de Cièncias (University of Lisbon, Faculty of Sciences)JULSSK45Jazykovedný ústav Ludovíta				
27Sveučilište u Zagrebu Filozofski fakultet (Univ. of Zagreb, Faculty of Hum. and Social Sciences)FFZGHR28Københavns Universitet (University of Copenhagen)UCPHDK29Tartu Ulikool (University of Tartu)UTARTEE30Helsingin Yliopisto (University of Helsinki)UHELFI31Centre National de la Recherche ScientifiqueCNRSFR32Nyelvtudományi Kutatóközpont (Research Institute for Linguistics)NYTKHU33Stofnun Árna Magnússonar í íslenskum fræðum SAM (Árni Magnússon Inst. for Icelandic Studies)SAMIS34Fondazione Bruno KesslerFBKIT52Latvijas Universitátes Matemátikas un Informátikas institúts (Institute of Mathematics and ComputerIMCSLV36Lietuvi Kalbos Institutes (Institute of the Lithuanian Language)LKILT31Luxembourg Institute of Science and TechnologyLISTLU38Università ta Malta (University of Malta)UMMT39Stichting Institutu voor de Nederlandse Taal (Dutch Language Institute)INTNL40Språkrådet (Language Council of Norway)LCNORNO41Institut Podstaw Informatyki Polskiej Akademii Nauk (Polish Academy of Sciences)IPPANPL42Universidade de Lisboa, Faculdade de Ciências (University of Lisbon, Faculty of Science)FCULisbonPT43Institut Jožef Stefan (Jozef Stefan Institute)SISI44University of Cyprus, French and European StudiesUCYCY45 <td></td> <td></td> <td></td> <td></td>				
28Københavns Universitet (University of Copenhagen)UCPHDK29Tartu Ulikool (University of Tartu)UTARTEE30Helsingin Yliopisto (University of Helsinki)UHELFI31Centre National de la Recherche ScientifiqueCNRSFR32Nyelvtudományi Kutatóközpont (Research Institute for Linguistics)NYTKHU33Stofnun Árna Magnússonar í íslenskum fræðum SAM (Árni Magnússon Inst. for Icelandic Studies)SAMIS34Fondazione Bruno KesslerFBKIT35Latyias Universitates Matemātikas un Informātikas institūts (Institute of Mathematics and Computer Science, University of Latvia)LKILT36Lietuvių Kalbos Institutas (Institute of the Lithuanian Language)LKILT37Luxembourg Institute of Science and TechnologyLISTUMMT39Stichting Institutut voor de Nederlandse Taal (Dutch Language Institute)INTNL40Språkrådet (Language Council of Norway)LCNORNO41Institut Podstaw Informatyki Polskiej Akademii Nauk (Polish Academy of Sciences)IPIPANPT43Institut de Cercetări Pentru Inteligență Artificială (Romanian Academy)ICIARO44University of Cyprus, French and European StudiesUCYCY45Jazykovedný ústav L'udovíta Štúra Slovenskej akadémie vied (Slovak Academy of Sciences)JULSSK46Institut Jožef Stefan (Jozef Stefan Institute)ISISI47Centro Nacional de Supercomputación (Barcelona Supercomputing Center) <td></td> <td></td> <td></td> <td></td>				
29Tartu Ulikool (University of Tartu)UTARTEB30Helsingin Yliopisto (University of Helsinki)UHELFI31Centre National de la Recherche ScientifiqueCNRSFR32Nyelvtudományi Kutatóközpont (Research Institute for Linguistics)NYTKHU33Stofnun Árna Magnússonar í íslenskum fræðum SAM (Árni Magnússon Inst. for Icelandic Studies)SAMIS34Fondazione Bruno KesslerFBKIT35Latvijas Universitátes Matemátikas un Informátikas institúts (Institute of Mathematics and Computer Science, University of Latvia)LKILKI36Lietuvių Kalbos Institutas (Institute of the Lithuanian Language)LKILKI37Luxembourg Institute of Science and TechnologyLISTLU38Università ta Malta (University of Malta)UMMT39Stichting Institutu voor de Nederlandse Taal (Dutch Language Institute)INTNL40Språkrådet (Language Council of Norway)LCNORNO41Institut Podstaw Informatyki Polskiej Akademii Nauk (Polish Academy of Sciences)IPIPANPL42Universidade de Lisboa, Faculdade de Ciências (University of Lisbon, Faculty of Science)FCULisbonPT43Institut Jožef Stefan (Jozef Stefan Institute)ICIARO44University of Cyprus, French and European StudiesUCYCY45Jazykovedný ústav Ľudovíta Štúra Slovenskej akadémie vied (Slovak Academy of Sciences)JULSSK46Institut Jožef Stefan (Jozef Stefan Institute)ISISI<				
30Helsingin Yliopisto (University of Helsinki)UHELF131Centre National de la Recherche ScientifiqueCNRSFR32Nyelvtudományi Kutatóközpont (Research Institute for Linguistics)NYTKHU33Stofnun Árna Magnússonar í íslenskum fræðum SAM (Árni Magnússon Inst. for Icelandic Studies)SAMIS34Fondazione Bruno KesslerFBKIT35Latvijas Universitátes Matemātikas un Informātikas institūts (Institute of Mathematics and Computer Science, University of Latvia)IMCSLV36Lietuvių Kalbos Institutas (Institute of the Lithuanian Language)LKIL37Luxembourg Institute of Science and TechnologyLISTU39Stichting Institutu voor de Nederlandse Taal (Dutch Language Institute)INTNL40Språkrådet (Language Council of Norway)LCNORNO41Instytut Podstaw Informatyki Polskiej Akademii Nauk (Polish Academy of Sciences)IPIPANPL42Universidade de Lisboa, Faculdade de Ciências (University of Lisbon, Faculty of Science)FCULisbonPT43Institutul de Cercetări Pentru Inteligență Artificială (Romanian Academy)ICIARO44University of Cyprus, French and European StudiesUCYCY45Jazykovedný ústav Ludovíta Štúra Slovenskej akadémie vied (Slovak Academy of Sciences)JULSSK46Institut Jožef Stefan (Jozef Stefan Institute)SISI47Centro Nacional de Supercomputación (Barcelona Supercomputing Center)BSCES48Kungliga		, , , , , , , , , , , , , , , , , , , ,		
31Centre National de la Recherche ScientifiqueCNRSFR32Nyelvtudományi Kutatóközpont (Research Institute for Linguistics)NYTKHU33Stofnun Árna Magnússonar í íslenskum fræðum SAM (Árni Magnússon Inst. for Icelandic Studies)SAMIS34Fondazione Bruno KesslerFBKIT35Latvijas Universitates Matemātikas un Informātikas institūts (Institute of Mathematics and Computer Science, University of Latvia)LKILT36Lietuvių Kalbos Institutas (Institute of the Lithuanian Language)LKILT37Luxembourg Institute of Science and TechnologyLISTLU39Stichting Institutut voor de Nederlandse Taal (Dutch Language Institute)INTNI40Språkrådet (Language Council of Norway)LCNORNO41Instytut Podstaw Informatyki Polskiej Akademii Nauk (Polish Academy of Sciences)IPIPANPL42Universidade de Lisboa, Faculdade de Ciências (University of Lisbon, Faculty of Science)FCULisbonPT43Institutul de Cercetări Pentru Inteligență Artificială (Romanian Academy)ICIARO44University of Cyprus, French and European StudiesUCYCY45Jazykovedný ústav Ludovíta Štúra Slovenskej akadémie vied (Slovak Academy of Sciences)JULSSK46Institut Jožef Stefan (Jozef Stefan Institute)SRES48Kungliga Tekniska högskolan (Royal Institute of Technology)KTHSE49University of SheffieldUSPUZHCH50University of Sheffield <td></td> <td></td> <td></td> <td></td>				
32Nyelvtudományi Kutatóközpont (Research Institute for Linguistics)NYTKHU33Stofnun Árna Magnússonar í íslenskum fræðum SAM (Árni Magnússon Inst. for Icelandic Studies)SAMIS34Fondazione Bruno KesslerFBKIT35Latvijas Universitates Matemātikas un Informātikas institūts (Institute of Mathematics and Computer Science, University of Latvia)LKILT36Lietuvių Kalbos Institutas (Institute of the Lithuanian Language)LKILT37Luxembourg Institute of Science and TechnologyLISTLU38Università ta Malta (University of Malta)UMMT39Stichting Institutut voor de Nederlandse Taal (Dutch Language Institute)INTNL40Språkrådet (Language Council of Norway)LCNORNO41Instytut Podstaw Informatyki Polskiej Akademii Nauk (Polish Academy of Sciences)IPIPANPL42Universidade de Lisboa, Faculdade de Ciências (University of Lisbon, Faculty of Science)FCULisbonPT43Institutul de Cercetări Pentru Inteligență Artificială (Romanian Academy)ICIARO44University of Cyprus, French and European StudiesUCYCY45Jazykovedný ústav Ludovíta Štúra Slovenskej akadémie vied (Slovak Academy of Sciences)JULSSK46Institut Jožef Stefan (Jozef Stefan Institute)JSISI47Centro Nacional de Supercomputación (Barcelona Supercomputing Center)BSCES48Kungliga Tekniska högskolan (Royal Institute of Technology)KTHSE49<				
Stofnun Árna Magnússonar í íslenskum fræðum SAM (Árni Magnússon Inst. for Icelandic Studies) FBK IT FBK IT Latvijas Universitates Matemātikas un Informātikas institūts (Institute of Mathematics and Computer Science, University of Latvia) Lietuvių Kalbos Institutas (Institute of the Lithuanian Language) List Luxembourg Institute of Science and Technology LIST LU Nuiversità ta Malta (University of Malta) Um MT Stichting Instituut voor de Nederlandse Taal (Dutch Language Institute) Institut I Podstaw Informatyki Polskiej Akademii Nauk (Polish Academy of Science) Institutul de Cercetări Pentru Inteligență Artificială (Romanian Academy) Institutul de Cercetări Pentru Inteligență Artificială (Romanian Academy of Sciences) Institutul de Cercetări Pentru Inteligență Artificială (Romanian Academy of Sciences) JULS SK Jazykovedný ústav Ludovíta Štúra Slovenskej akadémie vied (Slovak Academy of Sciences) Institut Jožef Stefan (Jozef Stefan Institute) Schamber Institut Jožef Stefan (Jozef Stefan Institute) Latvija University of Cyprus, French and European Supercomputing Center) KTH SE Kungliga Tekniska högskolan (Royal Institute of Technology) KTH SE University of Sheffield USFD UK University of Sheffield University of Sheffield UNIGO ES		•		
FBK IT Latvijas Universitātes Matemātikas un Informātikas institūts (Institute of Mathematics and Computer Science, University of Latvija) Litvių Kalbos Institutas (Institute of the Lithuanian Language) Liki LT Luxembourg Institute of Science and Technology LIST LUS Università ta Malta (University of Malta) UM MT Stichting Institutu voor de Nederlandse Taal (Dutch Language Institute) INT NL US Språkrådet (Language Council of Norway) LCNOR NO Instytut Podstaw Informatyki Polskiej Akademii Nauk (Polish Academy of Sciences) IPIPAN PL LUXI Institutul de Cercetări Pentru Inteligență Artificială (Romanian Academy) ICIA RO University of Cyprus, French and European Studies Jazykovedný ústav Ludovíta Štúra Slovenskej akadémie vied (Slovak Academy of Sciences) ISI SI Centro Nacional de Supercomputación (Barcelona Supercomputing Center) Superior Steffan (Jozef Steffan Institute) Universität Zürich (University of Zurich) University of Sheffield University of Sheffield USFD UK				
Latvijas Universitātes Matemātikas un Informātikas institūts (Institute of Mathematics and Computer Science, University of Latvia) Lietuvių Kalbos Institutas (Institute of the Lithuanian Language) LKI LT LUXembourg Institute of Science and Technology LIST LUST LUXEMS Università ta Malta (University of Malta) UM MT Stichting Instituut voor de Nederlandse Taal (Dutch Language Institute) INT NL VENOR NO LCNOR NO IPIPAN PL Universidade de Lisboa, Faculdade de Ciências (University of Lisbon, Faculty of Sciences) IPIPAN PL Universidade de Lisboa, Faculdade de Ciências (University of Lisbon, Faculty of Science) ICIA RO UCY CY LIST LU MAT INSTITUTION NI LONOR NO INSTITUTION PL Universidade de Lisboa, Faculdade de Ciências (University of Lisbon, Faculty of Science) ICIA RO LIST NI LUXONOR NO IPIPAN PL Universidade de Lisboa, Faculdade de Ciências (University of Lisbon, Faculty of Science) ICIA RO LIST NI LUXONOR NO IVIA RO LCNOR NO IVIA RO LUCY CY LUCY		· · · · · · · · · · · · · · · · · · ·		
Science, University of Latvia) 36 Lietuvių Kalbos Institutas (Institute of the Lithuanian Language) LKI LT 37 Luxembourg Institute of Science and Technology LIST LU 38 Università ta Malta (University of Malta) UM MT 39 Stichting Institutu voor de Nederlandse Taal (Dutch Language Institute) INT NL 40 Språkrådet (Language Council of Norway) LCNOR NO 41 Instytut Podstaw Informatyki Polskiej Akademii Nauk (Polish Academy of Sciences) IPIPAN PL 42 Universidade de Lisboa, Faculdade de Ciências (University of Lisbon, Faculty of Science) FCULisbon PT 43 Institutul de Cercetări Pentru Inteligență Artificială (Romanian Academy) LICIA RO 44 University of Cyprus, French and European Studies UCY CY 45 Jazykovedný ústav Ludovíta Štúra Slovenskej akadémie vied (Slovak Academy of Sciences) JULS SK 46 Institut Jožef Stefan (Jozef Stefan Institute) ISI SI 47 Centro Nacional de Supercomputación (Barcelona Supercomputing Center) BSC ES 48 Kungliga Tekniska högskolan (Royal Institute of Technology) KTH SE 49 University of Sheffield USFD UK 50 University of Sheffield USFD UK				
LIST LU LUXembourg Institute of Science and Technology LIST LU LU38 Università ta Malta (University of Malta) UM MT Stichting Instituut voor de Nederlandse Taal (Dutch Language Institute) INT NL LUA0 Språkrådet (Language Council of Norway) LCNOR NO LCNOR NO LCNOR NO LUSY Podstaw Informatyki Polskiej Akademii Nauk (Polish Academy of Sciences) IPIPAN PL LUXE Universidade de Lisboa, Faculdade de Ciências (University of Lisbon, Faculty of Science) LIST UNIVERSIDADE LOST PORTO LOST PORTO LOST PORTO LOST PORTO LUSY CY		Science, University of Latvia)		
38Università ta Malta (University of Malta)UMMT39Stichting Instituut voor de Nederlandse Taal (Dutch Language Institute)INTNL40Språkrådet (Language Council of Norway)LCNORNO41Instytut Podstaw Informatyki Polskiej Akademii Nauk (Polish Academy of Sciences)IPIPANPL42Universidade de Lisboa, Faculdade de Ciências (University of Lisbon, Faculty of Science)FCULisbonPT43Institutul de Cercetări Pentru Inteligență Artificială (Romanian Academy)ICIARO44University of Cyprus, French and European StudiesUCYCY45Jazykovedný ústav Ludovítă Štúras Slovenskej akadémie vied (Slovak Academy of Sciences)JULSSK46Institut Jožef Stefan (Jozef Stefan Institute)JSISI47Centro Nacional de Supercomputación (Barcelona Supercomputing Center)BSCES48Kungliga Tekniska högskolan (Royal Institute of Technology)KTHSE49Università Zürich (University of Zurich)UZHCH50University of SheffieldUSFDUK51Universidad de Vigo (University of Vigo)UVIGOES				
Stichting Institutut voor de Nederlandse Taal (Dutch Language Institute) Språkrådet (Language Council of Norway) LCNOR NO 1 Instytut Podstaw Informatyki Polskiej Akademii Nauk (Polish Academy of Sciences) IPIPAN PL 1 Universidade de Lisboa, Faculdade de Ciências (University of Lisbon, Faculty of Science) ICIA RO VICY VIV Språkrådet (Language Council of Norway) ICIA RO LUSY CY Språkrådet (Language Council of Norway) ICIA RO LUSY CY LUSY CY LUSY LUCY CY LUSY LUCY LUC		6,		
40Språkrådet (Language Council of Norway)LCNORNO41Instytut Podstaw Informatyki Polskiej Akademii Nauk (Polish Academy of Sciences)IPIPANPL42Universidade de Lisboa, Faculdade de Ciências (University of Lisbon, Faculty of Science)FCULisbonPT43Institutul de Cercetări Pentru Inteligență Artificială (Romanian Academy)ICIARO44University of Cyprus, French and European StudiesUCYCY45Jazykovedný ústav Ludovíta Štúra Slovenskej akadémie vied (Slovak Academy of Sciences)JULSSK46Institut Jožef Stefan (Jozef Stefan Institute)JSISI47Centro Nacional de Supercomputación (Barcelona Supercomputing Center)BSCES48Kungliga Tekniska högskolan (Royal Institute of Technology)KTHSE49Universität Zürich (University of Zurich)UZHCH50University of SheffieldUSFDUK51Universidad de Vigo (University of Vigo)UVIGOES				
42Universidade de Lisboa, Faculdade de Ciências (University of Lisbon, Faculty of Science)FCULisbonPT43Institutul de Cercetări Pentru Inteligență Artificială (Romanian Academy)ICIARO44University of Cyprus, French and European StudiesUCYCY45Jazykovedný ústav Ludovíta Štúra Slovenskej akadémie vied (Slovak Academy of Sciences)JULSSK46Institut Jožef Stefan (Jozef Stefan Institute)JSISI47Centro Nacional de Supercomputación (Barcelona Supercomputing Center)BSCES48Kungliga Tekniska högskolan (Royal Institute of Technology)KTHSE49Universität Zürich (University of Zurich)UZHCH50University of SheffieldUSFDUK51Universidad de Vigo (University of Vigo)UVIGOES	40		LCNOR	NO
43Institutul de Cercetări Pentru Inteligență Artificială (Romanian Academy)ICIARO44University of Cyprus, French and European StudiesUCYCY45Jazykovedný ústav Ludovíta Štúra Slovenskej akadémie vied (Slovak Academy of Sciences)JULSSK46Institut Jožef Stefan (Jozef Stefan Institute)JSISI47Centro Nacional de Supercomputación (Barcelona Supercomputing Center)BSCES48Kungliga Tekniska högskolan (Royal Institute of Technology)KTHSE49Universität Zürich (University of Zurich)UZHCH50University of SheffieldUSFDUK51Universidad de Vigo (University of Vigo)UVIGOES	41	Instytut Podstaw Informatyki Polskiej Akademii Nauk (Polish Academy of Sciences)	IPIPAN	PL
44University of Cyprus, French and European StudiesUCYCY45Jazykovedný ústav Ľudovíta Štúra Slovenskej akadémie vied (Slovak Academy of Sciences)JULSSK46Institut Jožef Stefan (Jozef Stefan Institute)JSISI47Centro Nacional de Supercomputación (Barcelona Supercomputing Center)BSCES48Kungliga Tekniska högskolan (Royal Institute of Technology)KTHSE49Universität Zürich (University of Zurich)UZHCH50University of SheffieldUSFDUK51Universidad de Vigo (University of Vigo)UVIGOES	42	Universidade de Lisboa, Faculdade de Ciências (University of Lisbon, Faculty of Science)	FCULisbon	PT
45Jazykovedný ústav Ľudovíta Štúra Slovenskej akadémie vied (Slovak Academy of Sciences)JULSSK46Institut Jožef Stefan (Jozef Stefan Institute)JSISI47Centro Nacional de Supercomputación (Barcelona Supercomputing Center)BSCES48Kungliga Tekniska högskolan (Royal Institute of Technology)KTHSE49Universität Zürich (University of Zurich)UZHCH50University of SheffieldUSFDUK51Universidad de Vigo (University of Vigo)UVIGOES	43	Institutul de Cercetări Pentru Inteligență Artificială (Romanian Academy)	ICIA	RO
46Institut Jožef Stefan (Jozef Stefan Institute)JSISI47Centro Nacional de Supercomputación (Barcelona Supercomputing Center)BSCES48Kungliga Tekniska högskolan (Royal Institute of Technology)KTHSE49Universität Zürich (University of Zurich)UZHCH50University of SheffieldUSFDUK51Universidad de Vigo (University of Vigo)UVIGOES	44	University of Cyprus, French and European Studies	UCY	CY
47 Centro Nacional de Supercomputación (Barcelona Supercomputing Center) 48 Kungliga Tekniska högskolan (Royal Institute of Technology) 49 Universität Zürich (University of Zurich) 50 University of Sheffield 51 Universidad de Vigo (University of Vigo) 52 UVIGO 53 UVIGO 54 UVIGO 55 ES	45	Jazykovedný ústav Ľudovíta Štúra Slovenskej akadémie vied (Slovak Academy of Sciences)	JULS	SK
48Kungliga Tekniska högskolan (Royal Institute of Technology)KTHSE49Universität Zürich (University of Zurich)UZHCH50University of SheffieldUSFDUK51Universidad de Vigo (University of Vigo)UVIGOES	46	Institut Jožef Stefan (Jozef Stefan Institute)	JSI	SI
49 Universität Zürich (University of Zurich) 50 University of Sheffield 51 Universidad de Vigo (University of Vigo) 52 UVIGO 53 UVIGO 54 UVIGO 55 UVIGO	47	Centro Nacional de Supercomputación (Barcelona Supercomputing Center)	BSC	ES
50 University of Sheffield USFD UK 51 Universidad de Vigo (University of Vigo) UVIGO ES	48		KTH	SE
51 Universidad de Vigo (University of Vigo) UVIGO ES	49	·	UZH	CH
		,		
52 Bangor University BNGR UK				
	52	Bangor University	BNGR	UK



Contents

1	Introduction	
2	The Luxembourgish Language in the Digital Age 2.1 General Facts	2 2 4
3	What is Language Technology?	5
4	Language Technology for Luxembourgish4.1Language Data Tools4.2Projects, Initiatives, Stakeholders	7 7 9
5	5.2 Levels of Technology Support	14
6	Summary and Conclusions	17



List of Figures

1 Overall state of technology support for selected European languages (2022) . . 16

List of Tables

List of Acronyms

AI Artificial Intelligence

CEF AT Connecting Europe Facility, Automated Translation CTIE Centre des technologies de l'information de l'État

DGT Directorate-General for Translation

DSG Deutsche Gebärdensprache EC European Commission

ELE European Language Equality (this project)

ELE Programme European Language Equality Programme (the long-term, large-scale fund-

ing programme specified by the ELE project)

ELG European Language Grid (EU project, 2019-2022)
ELRC European Language Resource Coordination

GPU Graphics Processing Unit

HCI Human Computer Interaction (see HMI)

HPC High-Performance Computing INL Institut National des Langues

LACS Luxembourg American Cultural Society & Center

LIH Luxembourg Institute of Health

LISER Luxembourg Institute of Socio-Economic Research
LIST Luxembourg Institute of Science and Technology

LOD Luxembourgish Online Dictionary
LR Language Resources/Resources
LT Language Technology/Technologies

Machine Learning ML**Machine Translation** MT Named Entity Recognition NER NCC National Competence Centre NLG Natural Language Generation Natural Language Processing NLP Natural Language Understanding NLU **Optical Character Recognition** OCR OLO ofizjel lezebuurjer ortografi

TTS Text-to-Speech SR Speaker Recognition

ZLS Zenter fir d'Lëtzebuerger Sprooch



Abstract

Luxembourg, officially the Grand Duchy of Luxembourg, is a small, but plurilingual country. The national language is Luxembourgish, while French is the legislative language; French, German and Luxembourgish are the three administrative and judicial languages.

There are about 650,000 inhabitants and the majority of Luxembourgers speak four languages, while the rate of foreign residents almost reaches 50% of the total population. According to STATEC (as of May 2019), French is the mostly spoken language at work (78%), followed by English (51%) and Luxembourgish (48%), whereas Luxembourgish is the most widely spoken language at home (53%).

Luxembourgish is present in newspapers, TV, radio and the Internet. As of March 2021, there were 59,000 Wikipedia articles written in Luxembourgish. Luxembourgish will become more important in secondary schools with changes to be incorporated in the academic school years 2021-2022 and 2022-2023.

However, Luxembourgish is still an under-resourced language when it comes to Luxembourgish data resources and tools, such as lexicons, glossaries, corpora, mobile applications, etc. There is need for dedicated funding programmes and strategies, training and education programmes as well as research – industry – government collaborations.

Natural language is the most common and versatile way for humans to convey information. The computational processing of human languages has been established as a specialised field known as Language Technology. Awareness needs to be raised and shared among various stakeholders about the benefits of Language Technology. This study provides an overview of the current status of the level of support that Luxembourgish receives through technology. We refer to the AI strategy of Luxembourg, to the long-term strategy of the Luxembourg State to promote Luxembourgish, as well as to available language data and tools in Luxembourgish, which are listed in the European Language Grid (ELG) – a catalogue that has been populated by language informants across all European member states.

Zesummefaassung

Lëtzebuerg, offiziell Groussherzogtum Lëtzebuerg, ass e klengt, awer méisproochegt Land. D'Nationalsprooch ass Lëtzebuergesch, a Franséisch ass déi legislativ Sprooch. Franséisch, Däitsch a Lëtzebuergesch sinn déi dräi Verwaltungs – a Justizsproochen. Et gi ronn 640.000 Awunner an d'Majoritéit vun de Lëtzebuerger schwätzt véier Sproochen, während den Taux vun Auslänner bal 50% vun der Gesamtbevëlkerung erreecht. Laut STATEC (Stand Mee 2019) ass Franséisch déi meescht geschwat Sprooch op der Aarbecht (78%), gefollegt vun Englesch (51%) a Lëtzebuergesch (48%), woubäi Lëtzebuergesch déi meescht geschwat Sprooch doheem ass (53%). D'Lëtzebuergescht ass an den Zeitungen, op der Tëlee, am Radio, an um Internet präsent. Zanter Mäerz 2021 goufen 59.000 Artikelen op Wikipedia op Lëtzebuergesch geschriwwen. D'Lëtzebuergescht gëtt och am Lycée méi wichteg mat Ännerunge déi fir déi akademescht Schouljoeren 2021 – 2022 an 2022 – 2023 geplangt sinn.

Leider ginn et de Moment nach net vill lëtzebuergesch Datenressourcen an Tools, wéi z.B. Lexikonen, Glossarer, Corpa, mobil Uwendungen, asw. Fir d'Situatioun ze verbesseren, gi méi Finanzéierungsprogrammer oder Strategien, Trainings – an Ausbildungsprogrammer souwéi Zesummenaarbecht tëschent Fuerschung, Industrie a Regierung gebraucht.

Natierlech Sprooch ass eng allgemeng a villsäiteg Manéier fir de Mënschen Informatioun ze vermëttelen. D'Computer Veraarbechtung vu mënschleche Sprooche gouf als e spezialiséiert Feld etabléiert bekannt als Sproochtechnologie. Et muss e Bewosstsinn iwwer d'Virdeeler vun der Sproochtechnologie entstoen an ënnert verschiddenen Akteuren gedeelt ginn.



Dës Etude gëtt en Iwwerbléck iwwert den aktuellen Zoustand vum Niveau vun der Ënnerstëtzung, déi d'Lëtzebuergescht duerch Technologie kritt. Mir bezéien eis op d'AI-Strategie vu Lëtzebuerg, op d'laangfristeg Strategie vum Lëtzebuerger Staat fir d'Lëtzebuergescht ze förderen, souwéi op verfügbare Sproochdaten an Tools op Lëtzebuergesch, déi am European Language Grid (ELG) opgezielt sinn – e Katalog déi vu Sproochinformanten an allen europäesche Memberstaaten erstallt gouf.

1 Introduction

This study is part of a series that reports on the results of an investigation of the level of support the European languages receive through technology. It is addressed to decision makers at the European and national/regional levels, language communities, journalists, etc. and it seeks to not only delineate the current state of affairs for each of the European languages covered in this series, but to additionally – and most importantly – to identify the gaps and factors that hinder further development of research and technology. Identifying such weaknesses will lay the grounds for a comprehensive, evidence-based, proposal of required measures for achieving Digital Language Equality in Europe by 2030.

To this end, more than 40 research partners, who are experts in more than 30 European languages have conducted an enormous and exhaustive data collection procedure that provides a detailed, empirical and dynamic map of technology support for our languages.¹

The report has been developed in the frame of the European Language Equality (ELE) project. With a large and all-encompassing consortium consisting of 52 partners covering all European countries, research and industry and all major pan-European initiatives, the ELE project develops a strategic research, innovation and implementation agenda as well as a roadmap for achieving full digital language equality in Europe by 2030.

2 The Luxembourgish Language in the Digital Age

2.1 General Facts

Luxembourg is a very small, but highly multilingual country. Luxembourg was part of changing European empires. The current territory came into being through several divisions, most recently in 1839 as an outcome of the Belgian Revolution. In 1867, the country was declared neutral in the Second London Treaty after another dispute over its political affiliation (the "Luxembourg Crisis"). In 1890, after the death of King William III of the Netherlands, the country became completely independent. In the two World Wars of the 20th century Luxembourg was occupied by the German Empire, first in 1914 and later in 1940, when, in both cases, Luxembourg marched to France. More information on the history of Luxembourg can be found at Pauly (2011) and Kreins (2021). The historical and sociolinguistic background of the Grand Duchy of Luxembourg, particularly with regards to street naming practices, can be found at Purschke (2021b).

Today, Luxembourg is the third European capital – along with Brussels and Strasbourg. It has the honour of hosting many of the European Union's important institutions, including the Publications Office of the EU, the Directorate-General for Translation (DGT), and the Translation Centre for the Bodies of the EU.

As for the population of Luxembourg, the Statistics portal of the Grand Duchy of Luxembourg (henceforth STATEC) published in February 2019 a demographic atlas of Luxembourg

The results of this data collection procedure have been integrated into the European Language Grid so that they can be discovered, browsed and further investigated by means of comparative visualisations across languages.

(in French)². According to this atlas, between 1981 and 2018, the Luxembourgish population increased by about 65%, from 364,597 to 602,005. Currently, the population is 649,574.³ There are 12 officially declared towns (Diekirch, Differdange, Dudelange, Echternach, Eschsur-Alzette, Ettelbruck, Grevenmacher, Luxembourg, Remich, Rumelange, Vianden, Wiltz) and 102 municipalities. Luxembourg City has the highest percentage of foreigners with 70.8%.

The languages spoken vary depending on the social situation or the region. From a geographical point of view, the regions with the highest density of Luxembourgish speakers include the north (85%) and the east (81%) of the country. According to the STATEC⁴, three out of four residents work in a multilingual environment and 25% of the population has to speak four or more languages at work. French is the mostly spoken language at work (78%), followed by English (51%) and Luxembourgish (48%). Luxembourgish is the most widely spoken language at home (53%), followed by French (32%) and Portuguese (19%). More information on multilingualism in Luxembourg can be found at the government's website.⁵

It should be noted that Luxembourgish and Turkish, which have official status in Luxembourg and Cyprus respectively, are the only two official languages of EU member states that are not official languages of the EU.

The Luxembourgish language is a Moselle-Franconian dialect, which was historically the mainly spoken language up to the 19th century in Luxembourg. On 24 February 1984, a law was first enacted and made Luxembourgish an officially recognised language ("Sproochege-setz"). In September 2018, the law was amended to add German sign language as an official language of the grand duchy of Luxembourg. The German sign language was chosen because the current "Deutsche Gebärdensprache" (DSG) is used by the majority of the community, and follows the recommendations by Daaflux (Deaf and hearing-impaired Association) and experts and NGOs in the field. More information on the German Sign Language can be found at the German report.

According to the provisions of the Languages Law of 1984, French, German or Luxembourgish may be used in administrative and judicial matters. This means that citizens can interact with the administration in any of these three languages, and that officials must attempt 'as far as possible' to respond in the language used by the applicant. Legislative documents are written in French and an important consequence of this on the judicial level is that only the text in French is deemed authentic for all levels of public administration.

As far as the vocabulary of Luxembourgish is concerned, it has a substantial number of loan words from French and German, but the morpho-syntax follows Germanic patterns (Gilles and Trouvain, 2013). With the exception of the alveolo-palatal fricatives and the approximant [w], the consonant inventory of Luxembourgish is quite similar to Standard German. The vowel inventory contains the phones [i: i e: e θ ϵ : æ a: a θ o: o u: u] as monophthongs. In addition, Luxembourgish has a set of eight diphthongs, which is considerably larger than the Standard German one (eight compared to three) (Gilles and Trouvain, 2013). A recent article on the complex language situation of Luxembourg with focus on dialectology can be found at Gilles (2019).

There is an officially recognised system with regards to the orthography of Luxembourgish, called "OLO" (ofizjel lezebuurjer ortografi)⁷ since 5 June 1946. The latest version of the official Luxembourgish orthography can be found at the Zenter fir d'Lëtzebuerger Sprooch

https://statistiques.public.lu/fr/actualites/population/population/2019/02/20190228/index.html

³ https://countrymeters.info/de/Luxembourg

⁴ https://statistiques.public.lu/fr/actualites/conditions-sociales/conditions-vie/2019/05/20190520/index.html

https://luxembourg.public.lu/en/society-and-culture/languages/languages-spoken-luxembourg.html

⁶ https://today.rtl.lu/news/luxembourg/a/1214521.html

https://web.archive.org/web/20050426112604/http://www.legilux.public.lu/leg/a/archives/1946/0400709/index. html



(ZLS)/Centre for the Luxembourgish Language⁸ and also downloaded as a PDF file (Conseil fir d'Lëtzebuerger Sprooch (CPLL) an Zenter fir d'Lëtzebuerger Sprooch (ZLS), 2021). The Luxembourgish orthography officially regulates the spelling of the Luxembourgish language for the areas for which the Luxembourg State is responsible (administrations, schools). The complete set of rules is based on a synthesis of the orthography dating back to 1975 and 1999. The codification and subsequent implementation of orthography in Luxembourgish can be found at Gilles (2015). More information on the languages spoken in Luxembourg can be found at Lulling et al. (2010). The Wikipedia page on the Luxembourgish language⁹ includes many updated references on the grammar or lexicons of Luxembourgish. The recent status of Luxembourgish, the cultural and linguistic influences as well as the language preferences in everyday practice can be found at Purschke (2020).

In order to promote the Luxembourgish language, the government has implemented a long-term strategy¹⁰ with four major objectives and concrete commitments:

- increasing the importance of Luxembourgish;
- advancing the standardisation, use and study of Luxembourgish;
- promoting learning Luxembourgish and learning about Luxembourg culture;
- promoting culture in the Luxembourgish language.

On this note, the aforementioned ZLS was founded in 2018 (see 4.2). Moreover, two measures are planned at the institutional level: on the one hand, the inscription of the Luxembourgish language and multilingualism into the Luxembourg Constitution, and on the other, the fact that Luxembourgish reaches a status that allows it to be used in many administrative processes. At the government declaration is stated that: "La place de la langue luxembourgeoise au niveau de l'Union européenne (UE) sera renforcée. Des négociations seront menées avec les partenaires européens pour élaborer une règle administrative sur l'utilisation du luxembourgeois." (translation: "The position of the Luxembourgish language at the level of the European Union (EU) will be strengthened. Negotiations will be conducted with European partners to develop an administrative rule on the use of Luxembourgish."

As for the education system in Luxembourg, it is plurilingual, with the use of German, French and Luxembourgish. While German is the main language used for teaching in elementary schools and in the lower classes of high schools, most subjects are taught in French in the higher classes in high schools. Public schools also offers international classes, namely in French and English. More information on which languages are spoken at elementary schools and in classical or secondary education can be found on the website of the Ministry of Education, Children, and Youth (Minstère de l'Éducation nationale, de l' Enfance et de la Jeunesse). For more information on the Luxembourgish education system, Weber and Horner (2012) provide an account of language-in-education policies in Luxembourg since the creation of the Luxembourgish state in the early nineteenth century.

2.2 Luxembourgish in the Digital Sphere

As far as the usage of Luxembourgish on the Internet is concerned, the Luxembourgish Wikipedia has more than 59,000 articles as of March 2021. In social media, the tweets of the

⁸ https://portal.education.lu/zls/ORTHOGRAFIE

⁹ https://de.wikipedia.org/wiki/Luxemburgische_Sprache

https://menej.gouvernement.lu/en/dossiers.gouvernement%2Ben%2Bdossiers%2B2018%2Blangue-luxembourgeoise.html

¹¹ https://men.public.lu/en/themes-transversaux/langues-ecole-luxembourgeoise.html



government of Luxembourg are often written in Luxembourgish with the exception that, when they deal with affairs in the EU domain, these are written in English.

Regarding the Luxembourg press, it has always been polyglot. Whilst German has always been the language par excellence of the press, slowly but surely, the country's dailies, such as the "Luxemburger Wort" founded in 1848, and the "Tageblatt" created in 1913, have seen French taking up about 20% to 30% of the editorial space. There are currently 14 newspapers in Luxembourg, 10 of which are in German, 1 in French, 1 in Portuguese, and 2 bilingual (German & French); all newspaper have digital versions. RTL Luxembourg should also be mentioned (with both news, television and radio programmes). As far as television is concerned, RTL Télé Lëtzebuerg¹² is the main television channel in Luxembourg, broadcasting in Luxembourgish; it is part of the RTL Group. Moreover, radio 100,7¹³ is a public service radio station, broadcasting in the Luxembourgish language. The Luxembourgish programmes of RTL, including Télé Lëtzebuerg, but also radio 100,7 are heavily financed by the government of Luxembourg.

The official websites which are available in Luxembourgish are mainly the governmental websites and include rules and regulations, news, information about the ministries, departments, topics, and political system. According to Rivera Pastor et al. (2017) in the study about equality in the digital age, Luxembourgish together with a few other languages (e. g. Irish, Maltese, Scottish Gaelic and Welsh) are the least supported in the internet services analysed. Luxembourgish is not supported in Google Account (as an input language on the Web),¹⁴ Google Developer Console, YouTube (interface), Facebook, Yahoo (web), Twitter, Bing or Instagram.

Regarding the percentage of population with Internet access, according to STATEC¹⁵ (Share of households) in 2020, 94% of the households in Luxembourg had Internet access at home. It represents an increase of about 30% compared to 2005. The highest rate of households with Internet access recorded was from 2015 to 2017, when it stood at 97%. Luxembourg has also recently published its ultra-high-speed broadband strategy 2021-2025. The main objectives of this strategy are the following:

- · make connectivity accessible to all;
- accelerate the transition of households and businesses to more efficient and sustainable technologies;
- accelerate the deployment of future-proof infrastructure, while respecting technological neutrality;
- improve transparency and strengthen consumer protection;
- develop Luxembourg as the launchpad of choice for ICT service providers of today and tomorrow.

3 What is Language Technology?

Natural language¹⁶ is the most common and versatile way for humans to convey information. We use language, our natural means of communication, to encode, store, transmit,

¹² https://www.rtl.lu

¹³ https://www.100komma7.lu

¹⁴ Luxembourgish is supported in Google Translate though.

https://statistiques.public.lu/stat/TableViewer/tableView.aspx?ReportId=12976&IF_Language=eng&MainTheme=3&FldrName=1&RFPath=31

¹⁶ This section has been provided by the editors. It is an adapted summary of Agerri et al. (2021) and of Sections 1 and 2 of Aldabe et al. (2021).



share and process information. Processing language is a non-trivial, intrinsically complex task, as language is subject to multiple interpretations (ambiguity), and its decoding requires knowledge about the context and the world, while in tandem language can elegantly use different representations to denote the same meaning (variation).

The computational processing of human languages has been established as a specialised field known as *Computational Linguistics* (CL), *Natural Language Processing* (NLP) or, more generally, Language Technology (LT). While there are differences in focus and orientation, since CL is more informed by linguistics and NLP by computer science, LT is a more neutral term. In fact, LT is largely multidisciplinary in nature; it combines linguistics, computer science (and notably AI), mathematics and psychology among others. In practice, these communities work closely together, combining methods and approaches inspired by both, together making up *language-centric AI*.

Language Technology is the multidisciplinary scientific and technological field that is concerned with studying and developing systems capable of processing, analysing, producing and understanding human languages, whether they are written, spoken or embodied.

With its starting point in the 1950s with Turing's renowned intelligent machine (Turing, 1950) and Chomsky's generative grammar (Chomsky, 1957), LT enjoyed its first boost in the 1990s. This period was signalled by intense efforts to create wide-coverage linguistic resources, such as annotated corpora, thesauri, etc. which were manually labelled for various linguistic phenomena and used to elicit machine readable rules which dictated how language can be automatically analysed and/or produced. Gradually, with the evolution and advances in machine learning, rule-based systems have been displaced by data-based ones, i. e. systems that learn implicitly from examples. In the recent decade of 2010s we observed a radical technological change in NLP: the use of multilayer neural networks able to solve various sequential labelling problems. The success of this approach lies in the ability of neural networks to learn continuous vector representations of the words (or word embeddings) using vast amounts of unlabelled data and using only some labelled data for fine-tuning.

In recent years, the LT community has been witnessing the emergence of powerful new deep learning techniques and tools that are revolutionising the way in which LT tasks are approached. We are gradually moving from a methodology in which a pipeline of multiple modules was the typical way to implement LT solutions, to architectures based on complex neural networks trained with vast amounts of data, be it text, audio or multimodal. The success in these areas of AI has been possible because of the conjunction of four different research trends: 1) mature deep neural network technology, 2) large amounts of data (and for NLP processing large and diverse multilingual data), 3) increase in high performance computing (HPC) power in the form of GPUs, and 4) application of simple but effective self-learning approaches.

LT is trying to provide solutions for the following main application areas:

- **Text Analysis** which aims at identifying and labelling the linguistic information underlying any text in natural language. This includes the recognition of word, phrase, sentence and section boundaries, recognition of morphological features of words, of syntactic and semantic roles as well as capturing the relations that link text constituents together.
- Speech processing aims at allowing humans to communicate with electronic devices through voice. Some of the main areas in Speech Technology are Text to Speech Synthesis, i. e. the generation of speech given a piece of text, Automatic Speech Recognition, i. e. the conversion of speech signal into text, and Speaker Recognition (SR).



- Machine Translation, i.e. the automatic translation from one natural language into another.
- Information Extraction and Information Retrieval which aim at extracting structured information from unstructured documents, finding appropriate pieces of information in large collections of unstructured material, such as the internet, and providing the documents or text snippets that include the answer to a user's query.
- Natural Language Generation (NLG). NLG is the task of automatically generating texts. Summarisation, i. e. the generation of a summary, the generation of paraphrases, text re-writing, simplification and generation of questions are some example applications of NLG.
- Human-Computer Interaction which aims at developing systems that allow the user to converse with computers using natural language (text, speech and non-verbal communication signals, such as gestures and facial expressions). A very popular application within this area are conversational agents (better known as chatbots).

LT is already fused in our everyday lives. As individual users we may be using it without even realising it, when we check our texts for spelling errors, when we use internet search engines or when we call our bank to perform a transaction. It is an important, but often invisible, ingredient of applications that cut across various sectors and domains. To name just a few, in the *health* domain, LT contributes for instance to the automatic recognition and classification of medical terms or to the diagnosis of speech and cognitive disorders. It is more and more integrated in *educational* settings and applications, for instance for educational content mining, for the automatic assessment of free text answers, for providing feedback to learners and teachers, for the evaluation of pronunciation in a foreign language and much more. In the *law/legal* domain, LT proves an indispensable component for several tasks, from search, classification and codification of huge legal databases to legal question answering and prediction of court decisions.

The wide scope of LT applications evidences not only that LT is one of the most relevant technologies for society, but also one of the most important AI areas with a fast growing economic impact.¹⁷

4 Language Technology for Luxembourgish

In the 2nd half of 2021, we have been identifying resources for the Luxembourgish language in the framework of ELE. For this purpose, we have contacted all relevant stakeholders, such as the University of Luxembourg, companies/SMEs, research institutions, as well as ministries/government and the Center for the Luxembourgish Language (ZLS).

4.1 Language Data Tools

As far as **tools** are concerned, 33% of the identified tools are specific to Luxembourgish. These Luxembourgish-specific tools include: a Grapheme-to-Phoneme conversion for Luxembourgish based on 30,000 manually phonetically transcribed words, two spellcheckers,

In a recent report from 2021, the global LT market was already valued at USD 9.2 billion in 2019 and is anticipated to grow at an annual rate of 18.4% from 2020 to 2028 (https://www.globenewswire.com/news-release/2021/03/22/2196622/0/en/Global-Natural-Language-Processing-Market-to-Grow-at-a-CAGR-of-18-4-from-2020-to-2028.html). A different report from 2021 estimates that amid the COVID-19 crisis, the global market for NLP was at USD 13 billion in the year 2020 and is projected to reach USD 25.7 billion by 2027, growing at an annual rate of 10.3% (https://www.researchandmarkets.com/reports/3502818/natural-language-processing-nlp-global-market).



a PoS-tagger/Tokenizer/Lemmatizer, and Sentence Splitter (Sirajzade and Schommer, 2019), and a mobile application called "Schnëssen". This app is intended to establish a state-of-the-art digital platform to collect data on the present-day language situation of Luxembourgish by means of crowdsourcing: users can participate in a large set of audio recordings tasks and in sociolinguistic surveys. A recently published tool, LëtzRead, is a free browser extension to integrate Luxembourgish-learning just by browsing the web (displaying certain words in Luxembourgish). Moreover, The library spaCy for advanced NLP has been trained for Luxembourgish. The remaining share of the tools are mostly language-independent. To name just a few examples: OCR tools (e.g. Ocular) or word aligning (e.g. Tagaligner). The text-to-speech (TTS) tool MaryTTS has been extended to support the Luxembourgish language (Steiner et al., 2017). As for support in MT, Luxembourgish is supported by Google Translate, but not by other MT systems available.

Contemporaneity is well represented in Luxembourg. The web service eLuxemburgensia has an interface providing access to historical archives from 1841 to 2021. The content of the archives are mainly text, but also images, including 17,421 postcards. For some collections, there is limited access indicated by "Access only in the National Library" or "Limited access BnL". Users may make a private copy of all eluxemburgensia.lu content for personal use. Private copying is one of the limitations and exceptions of Luxembourg copyright.

As for **data resources** available in Luxembourgish, 21% is specific to Luxembourgish only. Out of these resources, most are monolingual corpora, but there is also a Luxembourgish COVID glossary as well as an orthography trainer. The biggest text corpus in Luxembourgish contains 170 million words from a wide range of genres (Parliamentary debates, literature, transcripts of conversations, media texts (articles from news outlets like RTL.lu, radio100,7, eldoradio, social media)). All texts are annotated (sentences, lemmatised, PoS) and orthographically normalised. This corpus is owned by the University of Luxembourg and is used for internal use only.

The Luxembourgish online dictionary $(LOD)^{21}$ is a multilingual dictionary with 30,000 entries, in which Luxembourgish words are translated into German, French, English and Portuguese and illustrated by examples. There are also verb and adjective tables as well as an audio button to listen to the pronunciation.

As for the mobile application Schnëssen, in the first year of data collection, around 210,000 recordings were collected for numerous phenomena from all linguistic levels and over 2,800 sociolinguistic questionnaires have been filled out. The app allowed the authors to compile the largest spoken language corpus of Luxembourgish (Entringer et al., 2021).

Most of the remaining 79% of our data resources collection are corpora which include texts in multiple languages (often >100), also including also Luxembourgish.

Many lexical Luxembourgish specific resources, including corpora, dictionaries, material for phonetics, applications, etc. are available at **Infolux**²², which is the research portal about Luxembourgish, developed and maintained by the Institute of the Luxembourgish Language and Literature at the University of Luxembourg.

Identifying some **gaps** in the availability of Luxembourgish tools and data, one can say that mainly because of the lack of underlying data resources, there are gaps on many aspects of LT. What is currently missing in the Luxembourgish LT landscape are available bilingual corpora, e.g. Luxembourgish – English/German/French. The availability of such data sets would facilitate the development of many LT applications, such as NER, MT, NLU, virtual agents, recommender systems, etc. All of these applications are mainly statistically-based, therefore they typically require a large amount of manually annotated training data. Re-

¹⁸ https://infolux.uni.lu/schnessen/

¹⁹ https://www.letzread.com

²⁰ https://infolux.uni.lu/letzebuergesch-lo-als-sprooch-an-de-spacy-nlp-tools/

²¹ https://www.lod.lu

²² https://infolux.uni.lu



garding language models which can be used for NLU and NLG, the multilingual BERT covers many languages, including Luxembourgish; however, a BERT model trained specifically on large Luxembourgish data would yield better results.

Another important aspect is that written Luxembourgish is not well standardised; consequently many people write Luxembourgish without following the official spelling rules. While both German and French are intensively taught in schools, Luxembourgish, although the first language of approximately 60% of the population, forms part of the school curriculum only rudimentarily (Gilles, 2015). This has an impact on the correctness of Luxembourgish in the development of LT applications. According to Purschke (2020), "in view of the ongoing standardisation of Luxembourgish and a lack of rule knowledge in the population, orthographic variation — among other factors like code-switching or regional dialects – poses a great challenge for the automatic processing of text data".

It is noteworthy, though, that Luxembourgish will become more important in secondary schools with changes incorporated for the academic school year 2021-2022 and 2022-2023. "In the last years, the interest and the meaning of the Luxembourgish language have clearly grown," education minister Meisch said during a press conference: "With the integration of Luxembourgish in the upper secondary school cycle, and the introduction of Luxembourgish as the 4th language, we are mirroring this change in schools too."

4.2 Projects, Initiatives, Stakeholders

In this section, we will describe the national AI infrastructures and strategies currently available in Luxembourg.

Digitalisation plays a big role at the government of Luxembourg, therefore the Ministry for Digitalisation was created on 11th December 2018, headed by the Prime Minister, Xavier Bettel, and the Deputy Minister for Digitalisation, Marc Hansen. Some of its strategic axes are: i) Developing eGovernment, ii) Advancing administrative reform, iii) Promoting digital inclusion, and iv) Integrating new technologies. The Government IT Centre (Centre des technologies de l'information de l'État, CTIE), directly subordinated to the Ministry for Digitalisation is in charge of the setting up and development of eGovernment. GovTech Lab²⁶ is a recent joint initiative of the Ministry for Digitalisation and the CTIE. GovTech is an innovation laboratory that uses open innovation to work with internal (ministries, administrations, public actors) and external actors in the development of innovative solutions.

Digital Luxembourg²⁷ was founded in 2014 as a collaborative government initiative overseen by a dedicated team at the Ministry of State's Department of Media, Telecommunications & Digital Policy. It presents public-private partnerships & government initiatives driving the nation's digital progress.

Concerning the citizens' online services provided by the State, Guichet.lu²⁸ is an information portal that simplifies citizens' exchanges with the State and offers them quick and user-friendly access to all the information, procedures and services offered by Luxembourg public bodies. A mobile application has recently been developed for myguichet.lu to make online administrative procedures more accessible. The website of Guichet.lu is available in French, German, and English, but it communicates in French, German, English and Luxembourgish on social media. Guichet.lu is offering several documents in plain language (a clear and easy-to-understand language for persons with reading and comprehension difficulties).

²³ https://men.public.lu/en/actualites/communiques-conference-presse/2022/01/26-letzebuergesch-4.html

²⁴ https://delano.lu/article/more-luxembourgish-in-high-sch

²⁵ Strategic axes of the Ministry for Digitalisation: https://digital.gouvernement.lu/en/axes.html

²⁶ https://govtechlab.public.lu/en.html

²⁷ https://digital-luxembourg.public.lu

²⁸ https://guichet.public.lu/en.html



As far as the AI strategy of Luxembourg is concerned, Luxembourg's national AI Vision initiative underlines the country's unique ability to become a living lab of real-world AI applications, while developing leading AI regulations that put people first. In this AI vision, NLP is referred to in the section under AI for Employment: "AI will impact both repetitive administrative tasks and major intellectual challenges. Without a doubt, it could be an excellent tool in cognitive tasks – such as analyzing and reading legal texts or translating, via natural language processing tools." (AI vision).

Furthermore, the Ministry for Digitalisation and the CTIE have drawn recently the 'Electronic Governance 2021-2025' strategy, which was adopted by the Government Council in early 2021 (Electronic Governance 2021-2025). This strategy analyses the essential conditions required by the central civil service administration enabling it to meet society's needs efficiently. Six key principles make it possible to guide and support the digitalisation of public services and ensure that online public services meet the needs of society: i) Once Only, ii) Digital by Default, iii) Inclusion and Accessibility, iv) Openness and Transparency, v) Reliability and Security, vi) Interoperability and Standardisation.

Besides the official governmental strategies, there was also a large-scale survey at the end of 2020 about "What the public thinks about AI" conducted by the Luxembourg Institute of Socio-Economic Research (LISER report). The survey was filled in by 2,400 randomly selected citizens by the end of 2020. When asked which sector they trusted most with data and AI, the public sector won, with 77% of respondents voicing a high or very high level of trust. This is linked to the next response, i. e. that citizens embrace digital tools to interact with government: 73% of people use digital tools for public sector tasks (online payments, MyGuichet.lu, etc.), while second was mobility (Google Maps, Waze, etc.).

One important stakeholder for the promotion of the Luxembourgish language is the ZLS, created by the law of 20 July 2018. The ZLS Centre contributes to the realisation of government policy on the Luxembourgish language. The ZLS may, on its own initiative or at the request of the government, deal with any issues related to the Luxembourgish language. Some of its activities include:

- Publish rules on the spelling and grammar of the Luxembourgish language;
- Develop and adapt linguistic tools;
- Answer questions about the spelling, grammar, phonetics and good use of the Luxembourgish language;
- On behalf of the Minister, officially translate documents and communications to be published.

Another important stakeholder regarding multilingualism in Luxembourg is the Luxembourg National Languages Institute (Institut National des Langues, INL)²⁹. INL offers adult language courses as part of lifelong learning, also known as 'adult education'. This type of education is characterised by the resumption of classes or training to acquire skills or qualifications that go beyond initial school or academic education. INL offers evening classes in eight languages, including the country's three official languages (French, German and Luxembourgish), as well as English, Chinese, Italian, Portuguese, and Spanish. The number of Luxembourgish classes for adult education in INL tripled from 2008 to 2018.

As far as the research landscape in Luxembourg is concerned, there are four public Research and Development Institutions in Luxembourg: 1) University of Luxembourg, 2) Luxembourg Institute of Health (LIH), 3) Luxembourg Institute of Science and Technology (LIST)

²⁹ https://www.inll.lu/en/



and 4) Luxembourg Institute of Socio-Economic Research (LISER). The full research land-scape in Luxembourg has been recently published (Research Luxembourg, 2021). Furthermore, the University of Sheffield is home to the international Centre for Luxembourg Studies, the only one of its kind outside of Luxembourg.³⁰ Moreover, outside Luxembourg, there is the Luxembourg American Cultural Society and Centre (LACS).³¹ LACS is a non profit organisation and its mission is to preserve the roots of the heritage and nurture the leaves of ongoing relationships between Luxembourg and America.

Back to Luxembourg, the language and culture-related disciplines are at the Faculty of Humanities, Education and Social Sciences, Department of Humanities, directed by Prof. Peter Gilles since 2020. At the Department of Humanities there are seven research institutes: Multilingualism, Philosophy, Institute for Luxembourgish Linguistics and Literature Studies, History, German studies, English studies, Romance Studies, Media and Arts. At the Institute for Luxembourgish Linguistics and Literature Studies, there are two programmes of study: i) the Master in 'Languages, Cultures and Media—Luxembourg Studies' and ii) the training course 'Luxembourgish Language and Culture'.

The research interests of the members of the department of Luxembourgish Language and Literature encompass:

- The structure and variation of Luxembourgish;
- The relationship between language and society in Luxembourg and the Greater Region;
- The characteristics and development of Luxembourgish literature in its multi-lingual and –cultural contexts;
- The study of Luxembourgish literature in light of literary and aesthetic theories.

At LIST, there is no dedicated Computational Linguistics group, however mainly open-source NLP frameworks are being used and customised in applied research projects in various domains, in Finance Technology, HCI & Mixed Reality or Health Informatics. On a more general note, LIST is working on a nation-wide digital twin³². "We like to think of our whole country as a digital testbed," confirms Thomas Kallstenius, CEO of LIST. A major challenge with the digital twin is to develop analytics methods capable of handling the enormous amount of data involved. "We are looking into a particular discipline of artificial intelligence (AI) called 'explainable and trustworthy AI", says Dr. Kallstenius. "We want to develop a technique that is highly accurate and detailed but whose outcomes can still be understood by decision makers." Furthermore, LIST has a Data Analytics platform³³, a hybrid infrastructure (external and internal Cloud) covering the entire range of data analytics activities. The platform supports LIST's research and innovation capacity and is based on three pillars: i) a HPC infrastructure, ii) a cognitive analytics pillar, iii) an interactive visualisation wall.

LIST is currently involved in three LT projects: it is the Technology National Anchor Point of the European Language Resources Coordination (ELRC) Project, the National Competence Centre (NCC) for the European Language Grid (ELG), and the European Language Equality (ELE) Project. The ELRC Whitepaper for Luxembourgish can be found on pages 113-116 of the ELRC White Paper (ELRC Consortium, 2019). The last ELRC Luxembourgish workshop³⁴ was held on 11/12/2020 with more than 80 participants from public administrations in Luxembourg, but also from European institutions as well as SMEs. The Belgian – Luxembourgish

³⁰ https://www.sheffield.ac.uk/luxembourgish/centre-luxembourg-studies

³¹ https://www.lacs.lu

³² https://www.tradeandinvest.lu/news/a-nation-wide-digital-twin/

³³ https://www.list.lu/en/institute/rd-infrastructures/data-analytics-platform/

³⁴ https://www.lr-coordination.eu/l2Luxembourg



ELG workshop was organised virtually by the university of Antwerp on 08/07/2021. There were 45 participants, coming from Belgium, Luxembourg, and the Netherlands. The presentations of all speakers are available.³⁵

As far as some recent research projects related to LT including Luxembourgish are concerned, we refer to ENRICH4ALL, STRIPS, and Lingscape. More information on these projects follow below.

ENRICH4ALL (E-goverNment [RI] CHatbot for ALL)³⁶ is a project funded by CEF-TC-2020-1: Automated Translation (2020-EU-IA-0088). The partners of the ENRICH4ALL Consortium are LIST, BEIA Consult International (Romania), Romanian Academy Research Institute for Artificial Intelligence "Mihai Drăgănescu", and SupWiz Aps (Denmark). The objective of ENRICH4ALL is to integrate the CEF AT core service platform eTranslation to an already existing AI-based chatbot technology. The chatbot service will be deployed in public administration. We primarily address the official languages in each Consortium member country: Luxembourgish, German, French, Romanian, and Danish. However, through the integration of eTranslation, we will support all languages currently supported by eTranslation, as well. Within the project, there have been various resources freely available, such as a language identifier for Luxembourgish as well as question-answering data sets.

STRIPS (A Semantic Search Toolbox for the Retrieve of Similar Patterns in Luxembourgish Documents)³⁷ was a three-year project (02/18-01/21), funded by the University of Luxembourg, that aimed to develop a semantic search toolbox for the retrieval of similar patterns in documents written in Luxembourgish. RTL (Radio Télévision Lëtzebuerg) was the project partner providing their online news and corresponding user comments (for the period covered by the news between 2008-2018) for the retrieval of similar patterns over different time spans. One result of the project was a Sentiment Annotation Tool (Sirajzade et al., 2020), which is a browser-based tool that enables the annotation of split sentences from the database. The Sentiment Engine, a separate module, is trained with this material in order to annotate the whole data set and analyze the sentiment of the comments over time and in relationship to the news articles.

Lingscape³⁸ is a project about corpus building and group projects. It is a mobile application researching linguistic landscapes all over the world by collecting photos of signs and lettering on an interactive map. All different kinds of signs and lettering form the linguistic landscape of a place or community. Often these signs display different languages, be they on the same sign or next to each other. Users may collect and upload their own photos or explore the map with all photos added by other users. Through this project, it will be evident in which languages signs are lettered in different countries and how language in public spaces changes over time. More information can be found at Purschke (2021a).

As for companies in Luxembourg related to LTs, most of them are LSPs and not R&D. There are a few exceptions with a company developing a translation management system and another one with MT applied for financial translations. Eight companies are currently listed in the ELG platform.³⁹

5 Cross-Language Comparison

The LT field⁴⁰ as a whole has evidenced remarkable progress during the last years. The advent of deep learning and neural networks over the past decade together with the consid-

 $^{^{35}\} https://www.uantwerpen.be/en/conferences/workshop-resources-for-luxembourgish-and-flemish/programme/$

³⁶ https://www.enrich4all.eu

³⁷ https://acc.uni.lu/Research/strips/

³⁸ https://lingscape.uni.lu

³⁹ https://www.european-language-grid.eu/catalogue/search/luxembourg?&entity_type__term=Organization

⁴⁰ This section has been provided by the editors.



erable increase in the number and quality of resources for many languages have yielded results unforeseeable before. However, is this remarkable progress equally evidenced across all languages? To compare the level of technology support across languages, we considered more than 11,500 language technology tools and resources in the catalogue of the European Language Grid platform (as of January 2022).

5.1 Dimensions and Types of Resources

The comparative evaluation was performed on various dimensions:

- The current state of technology support, as indicated by the availability of tools and services⁴¹ broadly categorised into a number of core LT application areas:
 - Text processing (e.g. part-of-speech tagging, syntactic parsing)
 - Information extraction and retrieval (e.g. search and information mining)
 - Translation technologies (e.g. machine translation, computer-aided translation)
 - Natural language generation (e.g. text summarisation, simplification)
 - Speech processing (e.g. speech synthesis, speech recognition)
 - Image/video processing (e.g. facial expression recognition)
 - Human-computer interaction (e.g. tools for conversational systems)
- The potential for short- and mid-term development of LT, insofar as this potential can be approximated by the current availability of resources that can be used as training or evaluation data. The availability of data was investigated with regard to a small number of basic types of resources:
 - Text corpora
 - Parallel corpora
 - Multimodal corpora (incl. speech, image, video)
 - Models
 - Lexical resources (incl. dictionaries, wordnets, ontologies etc.)

5.2 Levels of Technology Support

We measured the relative technology support for 87 national, regional and minority European languages with regard to each of the dimensions mentioned above based on their respective coverage in the ELG catalogue. For the types of resources and application areas, the respective percentage of resources that support a specific language over the total number of resources of the same type was calculated, as well as their average. Subsequently each language was assigned to one band per resource type and per application area and to an overall band, on a four-point scale, inspired by the scale used in the META-NET White Paper Series, as follows:

1. **Weak or no support**: the language is present (as content, input or output language) in <3% of the ELG resources of the same type

⁴¹ Tools tagged as "language independent" without mentioning any specific language are not taken into account. Such tools can certainly be applied to a number of languages, either as readily applicable or following fine-tuning, adaptation, training on language-specific data etc., yet their exact language coverage or readiness is difficult to ascertain.



- 2. **Fragmentary support**: the language is present in \geq 3% and <10% of the ELG resources of the same type
- 3. **Moderate support**: the language is present in \geq 10% and <30% of the ELG resources of the same type
- 4. **Good support**: the language is present in \geq 30% of the ELG resources of the same type⁴²

The overall level of support for a language was calculated based on the average coverage in all dimensions investigated.

5.3 European Language Grid as Ground Truth

At the time of writing (January 2022), the ELG catalogue comprises more than 11,500 metadata records, encompassing both data and tools/services, covering almost all European languages – both official and regional/minority ones. The ELG platform harvests several major LR/LT repositories⁴³ and, on top of that, more than 6,000 additional language resources and tools were identified and documented by language informants in the ELE consortium. These records contain multiple levels of metadata granularity as part of their descriptions.

It should be noted that due to the evolving nature of this extensive catalogue and differing approaches taken in documenting records, certain levels of metadata captured are not yet at the level of consistency required to carry out a reliable cross-lingual comparison at a granular level. For example, information captured on corpora size, annotation type, licensing type, size unit type, and so on, still varies across records for many languages, while numerous gaps exist for others. As the ELG catalogue is continuously growing, the comprehensiveness, accuracy and level of detail of the records will naturally improve over time. Moreover, the Digital Language Equality (DLE) metric will allow for dynamic analyses and calculations of digital readiness, based on the much finer granularity of ELG records as they mature.⁴⁴

For the purposes of high-level comparison in this report, the results presented here are based on relative counts of entries in the ELG for the varying types of data resources and tools/services for each language. As such, the positioning of each language into a specific level of technology support is subject to change and it reflects a snapshot of the available resources on January 2022.

That said, we consider the current status of the ELG repository and the higher level findings below adequately representative with regard to the current existence of LT resources for Europe's languages.

5.4 Results and Findings

As discussed above, our analysis takes into account a number of dimensions for data and tools/services. Table 1 reports the detailed results per language per dimension investigated and the classification of each language into an overall level of support.

⁴² The thresholds for defining the four bands were informed by an exploratory *k*-means 4-cluster analysis based on all data per application and resource type, in order to investigate the boundaries of naturally occurring clusters in the data. The boundaries of the clusters (i. e., 3%, 10% and 30%) were then used to define the bands per application area and resource type.

⁴³ At the time of writing, ELG harvests ELRC-SHARE, LINDAT/CLARIAH-CZ, CLARIN-SI, CLARIN-PL and HuggingFace.

⁴⁴ Interactive comparison visualisations of the technology support of Europe's languages will be possible on the ELG website using a dedicated dashboard, which dynamically analyses the resources available in the ELG repository, from the middle of 2022 onwards.

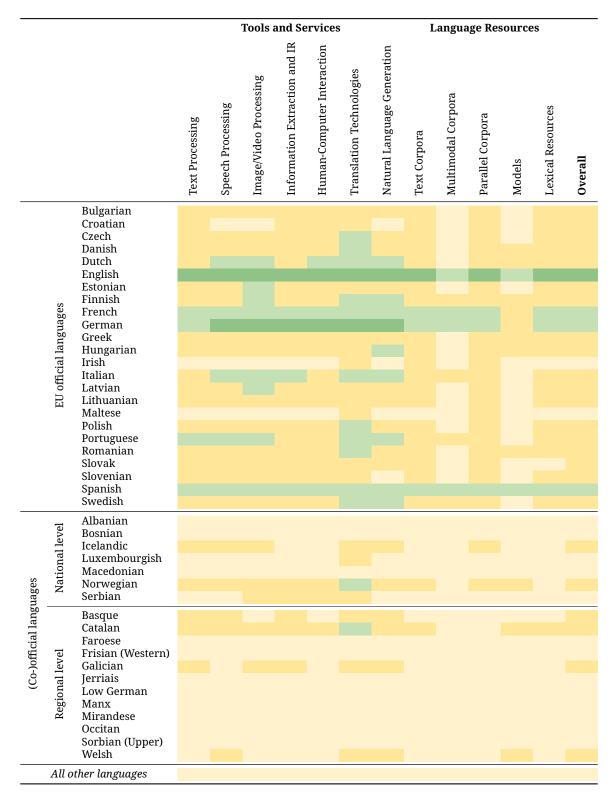


Table 1: State of technology support, in 2022, for selected European languages with regard to core Language Technology areas and data types as well as overall level of support (light yellow: weak/no support; yellow: fragmentary support; light green: moderate support; green: good support)

The best supported language is, as expected, English, the only language that is classified in the *good support* group. French, German and Spanish form a group of languages with *moderate support*. Although they are similar to English in some dimensions (e. g., German in terms of available speech technologies and Spanish in terms of available models), overall they have not yet reached the coverage that English has according to the ELG platform. All other official EU languages are clustered in the *fragmentary support* group, with the exception of Irish and Maltese, which have only *weak or no support*. From the remaining languages, (co-)official at national or regional level in at least one European country and other minority and lesser spoken languages, ⁴⁵ Norwegian and Catalan belong to the group of languages with *fragmentary support*. Basque, Galician, Icelandic and Welsh are borderline cases; while they are grouped in the *fragmentary support* level, they barely pass the threshold from the lowest level. All other languages are supported by technology either weakly or not at all. Figure 1 visualises our findings.

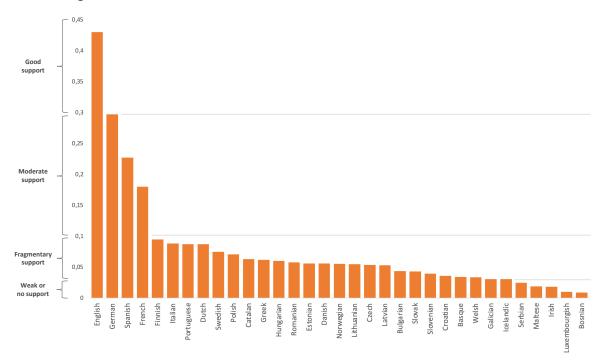


Figure 1: Overall state of technology support for selected European languages (2022)

While a fifth level, excellent support, could have been foreseen in addition to the four levels described in Section 5.2, we decided not to consider this level for the grouping of languages. Currently no natural language is optimally supported by technology, i.e. the goal of *Deep Natural Language Understanding* has not been reached yet for any language, not even for English, the best supported language according to our analysis. While recently there have been many breakthroughs in AI, Computer Vision, ML and LT, we are still far from the grand challenge of highly accurate deep language understanding, which is able to seamlessly integrate modalities, situational and linguistic context, general knowledge, meaning, reasoning,

⁴⁵ In addition to the languages listed in Table 1, ELE also investigated Alsatian, Aragonese, Arberesh, Aromanian, Asturian, Breton, Cimbrian, Continental Southern Italian (Neapolitan), Cornish, Eastern Frisian, Emilian, FrancoProvencal (Arpitan), Friulian, Gallo, Griko, Inari Sami, Karelian, Kashubian, Ladin, Latgalian, Ligurian, Lombard, Lower Sorbian, Lule Sami, Mocheno, Northern Frisian, Northern Sami, Picard, Piedmontese, Pite Sami, Romagnol, Romany, Rusyn, Sardinian, Scottish Gaelic, Sicilian, Skolt Sami, Southern Sami, Tatar, Tornedalian Finnish, Venetian, Võro, Walser, Yiddish.



emotion, irony, sarcasm, humour, culture, explain itself at request, and be done as required on the fly and at scale. A language can only be considered as excellently supported by technology if and when this goal of Deep Natural language Understanding has been reached.

The results of the present comparative evaluation reflect, in terms of distribution and imbalance, the results of the META-NET White Paper Series (Rehm and Uszkoreit, 2012). The complexities of the analyses clearly differ across the 2012 and 2022 studies, and as such, a direct comparison between the two studies can therefore not be made. However, we can instead compare the relative level of progress made for each language in the meantime. It is undebatable that the technology requirements for a language to be considered digitally supported today have changed significantly (e.g. the prevalent use of virtual assistants, chat bots, improved text analytics capabilities, etc.). Yet also the imbalance in distribution across languages still exists.

The results of this analysis are only informative of the relative positioning of languages, but not of the progress achieved within a specific language. The LT field as a whole has significantly progressed in the last ten years and remarkable progress has been achieved for specific languages in terms of quantity, quality and coverage of tools and language resources. Yet, the abysmal distance between the best supported languages and the minimally supported ones is still evidenced in 2022. It is exactly this distance that needs to be ideally eliminated, if not at least reduced, in order to move towards Digital Language Equality and avert the risks of digital extinction.

6 Summary and Conclusions

Luxembourg is a highly multilingual country with Luxembourgish as the national language, French as the legislative language, and French, German and Luxembourgish as the three administrative and judicial languages.

Luxembourgish has received an official status only since 1984, and moreover, is still not an official language of the EU. The number of Luxembourgish classes in INL tripled from 2008 to 2018.

The focus on the Luxembourgish language has increased during the last few years both from the governmental side with its long-term strategy and the research side, as a consequence. On the one hand, the government aims at increasing the importance of Luxembourgish by advancing the standardisation, use and study of Luxembourgish, promoting learning Luxembourgish and learning about Luxembourg culture, and promoting culture in the Luxembourgish language.

The ZLS, created by the law of 20 July 2018, contributes to the realisation of the government policy on the Luxembourgish language. On the other hand, we see that in the last few years, many research projects focus on Luxembourgish and some of them are based on crowdsourcing. This is an excellent example about creating large spoken, image, or written corpora quickly and by diverse users, which may contribute to developing further LT applications.

What is needed in Luxembourg are united forces for efficient collaboration. Since most people are multilingual, various stakeholders do not see the need to invest either time or budget in creating or sharing Luxembourgish resources. Both the government, the EU and the research institutions as well as LSPs have to work together to achieve the desired results. Resources and tools have to be openly available, following the principle "Data should be as open as possible and as closed as necessary". This collaboration brings certain benefits in a long-term strategy about LT education and training as well as investment in data collection and annotation.

Moreover, certain measures, such as dedicated LT programmes would certainly benefit



the Luxembourgish landscape in becoming broader. What is needed is dedicated education and training programmes, so that researchers, linguists and engineers have the necessary skills (e.g. Luxembourgish linguistics, computational linguistics, AI) to adapt to future job requirements.

Last but not least, if Luxembourgish is taught at schools, this would lead to future generations following the official spelling rules of Luxembourgish. Moreover, if Luxembourgish reaches a status where it can be used in many administrative procedures, this would have an enormous impact on creating multilingual written and spoken data resources.

To sum up, some important action points to increase the Language Technology with regards to Luxembourgish are:

- reaching a status that Luxembourgish can be used in many administrative procedures;
- raising awareness among various stakeholders in public and private sectors about the impact of having available spoken and written Luxembourgish data;
- advancing the standardisation, use and study of Luxembourgish;
- dedicated and collaborative national and EU funding programmes for both basic and applied research on Luxembourgish.

References

Rodrigo Agerri, Eneko Agirre, Itziar Aldabe, Nora Aranberri, Jose Maria Arriola, Aitziber Atutxa, Gorka Azkune, Arantza Casillas, Ainara Estarrona, Aritz Farwell, Iakes Goenaga, Josu Goikoetxea, Koldo Gojenola, Inma Hernaez, Mikel Iruskieta, Gorka Labaka, Oier Lopez de Lacalle, Eva Navas, Maite Oronoz, Arantxa Otegi, Alicia Pérez, Olatz Perez de Viñaspre, German Rigau, Jon Sanchez, Ibon Saratxaga, and Aitor Soroa. European Language Equality D1.2: Report on the state of the art in Language Technology and Language-centric AI, September 2021. URL https://european-language-equality.eu/wp-content/uploads/2021/10/ELE_Deliverable_D1_2.pdf.

AI vision. Artificial intelligence:a strategic vision for luxembourg. https://digital-luxembourg.public.lu/sites/default/files/2020-09/AI_EN_0.pdf.

Itziar Aldabe, Georg Rehm, German Rigau, and Andy Way. European Language Equality D3.1: Report on existing strategic documents and projects in LT/AI, November 2021. URL https://european-language-equality.eu/wp-content/uploads/2021/12/ELE__Deliverable_D3_1_revised_.pdf.

Noam Chomsky. Syntactic structures. The Hague: Mouton, 1957.

Conseil fir d'Lëtzebuerger Sprooch (CPLL) an Zenter fir d'Lëtzebuerger Sprooch (ZLS). D'lëtzebuerger orthografie, 2021. https://portal.education.lu/Portals/79/Documents/WEB_LetzOrtho_4Oplo_v01.pdf? ver=2021-01-13-085421-963.

Electronic Governance 2021-2025. Electronic governance 2021-2025 strategy. https://ctie.gouvernement.lu/dam-assets/documents/Strategie-gouvernance-electronique-2021-2025-vfin.pdf.

ELRC Consortium. ELRC White Paper, 2019. URL https://www.lr-coordination.eu/sites/default/files/Reports%202021/ELRCWhitePaper.pdf.

Nathalie Entringer, Peter Gilles, Sara Martin, and Christoph Purschke. Schnëssen. surveying language dynamics in luxembourgish with a mobile research app. *Linguistics Vanguard*, 7(s1):20190031, 2021. doi: doi:10.1515/lingvan-2019-0031. URL https://doi.org/10.1515/lingvan-2019-0031.

Peter Gilles. From status to corpus: Codification and implementation of spelling norms in luxembourgish. In *Language planning and microlinguistics*, pages 128–149. Springer, 2015.



- Peter Gilles. 39. komplexe überdachung ii: Luxemburg. die genese einer neuen nationalsprache. In *Deutsch*, pages 1039–1060. De Gruyter Mouton, 2019.
- Peter Gilles and Jürgen Trouvain. Luxembourgish. *Journal of the International Phonetic Association*, 43(1):67–74, 2013. doi: 10.1017/S0025100312000278.
- Jean-Marie Kreins. Histoire du Luxembourg. Que sais-je, 2021.
- LISER report. LISER report about what the public thinks about AI. https://smc.gouvernement.lu/dam-assets/Rapport-IA6-final.pdf.
- François Schanen Lulling et al. L'etzebuergesch: la langue nationale du grand-duché de luxembourg. Lengas. Revue de sociolinguistique, (60), 2010.
- Michel Pauly. Geschichte Luxemburgs, volume 2732. CH Beck, 2011.
- Christoph Purschke. Attitudes toward multilingualism in luxembourg. a comparative analysis of online news comments and crowdsourced questionnaire data. *Frontiers in Artificial Intelligence*, 3:79, 2020.
- Christoph Purschke. Crowdscapes. participatory research and the collaborative (re) construction of linguistic landscapes with lingscape. *Linguistics Vanguard*, 7(s1), 2021a.
- Christoph Purschke. Cultural representation in luxembourgish street naming practices. *Linguistics Vanguard*, 7(s5), 2021b.
- Georg Rehm and Hans Uszkoreit, editors. *META-NET White Paper Series: Europe's Languages in the Digital Age*, 32 volumes on 31 European languages, Heidelberg etc., 2012. Springer.
- Research Luxembourg. Research Landscape in Luxembourg, 2021. URL https://researchluxembourg.lu/wp-content/uploads/2021/09/ResearchLuxembourg_Booklet_Final.pdf.
- Rafael Rivera Pastor, Carlota Tarín Quirós, Juan Pablo Villar García, Toni Badia Cardús, and Maite Melero Nogués. Language equality in the digital age: Towards a human language project. Technical report, Report IP/G/STOA/FWC/2013-001/Lot4, 2017.
- Share of households. Share of households with internet access in luxembourg (2005-2020). https://statistiques.public.lu/stat/TableViewer/tableView.aspx?ReportId=12976&IF_Language=eng&MainTheme=3&FldrName=1&RFPath=31.
- Joshgun Sirajzade and Christoph Schommer. The luna open toolbox for the luxembourgish language. *Advances in Data Mining, Applications and Theoretical Aspects, Poster Proceedings 2019*, 2019.
- Joshgun Sirajzade, Daniela Gierschek, and Christoph Schommer. An annotation framework for lux-embourgish sentiment analysis. In *Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL)*, pages 172–176, 2020.
- Ingmar Steiner, Sébastien Le Maguer, Judith Manzoni, Peter Gilles, and Jürgen Trouvain. 28. Konferenz Elektronische Sprachsignalverarbeitung 2017, Saarbrücken. 2017.
- Alan M. Turing. Computing machinery and intelligence. *Mind*, LIX(236):433–460, 1950. ISSN 0026-4423. doi: 10.1093/mind/LIX.236.433. URL https://doi.org/10.1093/mind/LIX.236.433.
- Jean-Jacques Weber and Kristine Horner. The trilingual luxembourgish school system in historical perspective: Progress or regress? *Language, Culture and Curriculum*, 25(1):3–15, 2012.