

THE MALTESE IL-LINGWA  
LANGUAGE IN MALTJA FL-ERA  
THE DIGITAL DIGITALI  
AGE

Mike Rosner  
Jan Joachimsen





---

White Paper Series

Serie ta' White Papers

THE MALTESE  
LANGUAGE IN  
THE DIGITAL  
AGE

IL-LINGWA  
MALTIJA FL-ERA  
DIGITALI

Mike Rosner University of Malta

Jan Joachimsen University of Malta

---

Georg Rehm, Hans Uszkoreit  
(edituri, editors)



## DAĦLA

## PREFACE

Din il-White Paper hija għall-edukaturi, ġurnalisti, politikanti, komunitajiet ta' lingwi u oħrajn, li jixtiequ jstabilixxu Ewropa tabilhaqq multilingwali.

Din hija parti minn serje White Papers li jippromwovu għarfien dwar it-teknoloġija lingwistika u l-potenzjal tagħha. Id-disponibilità u l-użu tat-teknoloġija lingwistika fl-Ewropa tvarja bejn lingwa u oħra. Konsegwentement, l-azzjonijiet li huma meħtieġa sabiex jiġu appoġġjati r-riċerka u l-iżvilupp tat-teknoloġiji lingwistiċi ivarjaw ukoll f'kull lingwa. L-azzjonijiet meħtieġa jiddependu fuq bosta fatturi, bħal kumplessità ta' lingwa partikolari u d-daqs tal-komunità tagħha.

META-NET, Netwerk ta' Eċċellenza tal-Kummissjoni Ewropea, wettaq analiżi dwar ir-riżorsi u t-teknoloġiji lingwistiċi kurrenti f'din is-serje ta' *white papers* (p. 91). Din l-analiżi kienet ibbażata fuq 23 lingwa uffiċjali Ewropeja, kif ukoll lingwi reġjonali oħra importanti fl-Ewropa. Ir-riżultati ta' dan l-analiżi jissuggerixxu li hemm bosta nuqqasijiet fir-riċerka għal kull lingwa. Analizi aktar dettaljata u esperta u assessjar tas-sitwazzjoni kurrenti għandha tgħin sabiex timmasimizza l-impatt ta' riċerka addizzjonali.

Minn Novembru 2011 META-NET tikkonsisti f'54 centru ta' riċerka minn 33 pajjiż [1] (p. 87) li qed jaħdmu ma' partijiet interessati minn negozji kummerċjali, aġenziji governattivi, industriji, organizzazzjonijiet ta' riċerka, kumpaniji ta' software, fornituri ta' teknoloġija u universitajiet Ewropej. Flimkien, dawn qed johlqu vizjoni teknoloġika komuni filwaqt li jżviluppaw agenda ta' riċerka strateġika li turi kif applikazzjonijiet teknoloġiċi lingwistiċi jistgħu jindirizzaw xi nuqqasijiet ta' riċerka sal-2020.

This white paper is part of a series that promotes knowledge about language technology and its potential. It addresses journalists, politicians, language communities, educators and others. The availability and use of language technology in Europe varies between languages. Consequently, the actions that are required to further support research and development of language technologies also differs. The required actions depend on many factors, such as the complexity of a given language and the size of its community.

META-NET, a Network of Excellence funded by the European Commission, has conducted an analysis of current language resources and technologies in this white paper series (p. 91). The analysis focused on the 23 official European languages as well as other important national and regional languages in Europe. The results of this analysis suggest that there are tremendous deficits in technology support and significant research gaps for each language. The given detailed expert analysis and assessment of the current situation will help maximise the impact of additional research.

META-NET consists of 54 research centres from 33 European countries [1] (p. 87). META-NET is working with stakeholders from economy (software companies, technology providers, users), government agencies, research organisations, non-governmental organisations, language communities and European universities. Together with these communities, META-NET is creating a common technology vision and strategic research agenda for multilingual Europe 2020.

L-awturi ta' dan id-dokument huma grati lejn l-awturi tal-*White Paper* Ġermaniż għall-permess biex jużaw mill-ġdid materjali magħżulin mid-dokument tagħhom [2] li huma indipendenti mill-lingwa.

Grazzi lil Ritianne Stanyer u Roberta Abela għat-translazzjoni ta' dan id-dokument. L-awturi huma obbligati lejn il-Professur Ray Fabri, li l-artiklu [3] tiegħu kien l-ispirazzjoni għal hafna mill-kontenut u hafna mill-eżempji f'din id-dokument.

L-iżvilupp ta' din il-white paper ġie ffinanzjat mis-Seba' Programm ta' Qafas u l-Programm ta' Appoġġ għall-Politika dwar l-ICT tal-Kummissjoni Ewropea taħt il-kuntratti T4ME (Grant Agreement 249 119), CESAR (Grant Agreement 271 022), METANET4U (Grant Agreement 270 893) u META-NORD (Grant Agreement 270 899).

---

The authors of this document are grateful to the authors of the White Paper on German for permission to re-use selected language-independent materials from their document [2].

Thanks to Ritianne Stanyer and Roberta Abela for the translation of this document. The authors are indebted to Prof. Ray Fabri, whose article [3] was the inspiration for much of the content and many of the examples in the document.

The development of this white paper has been funded by the Seventh Framework Programme and the ICT Policy Support Programme of the European Commission under the contracts T4ME (Grant Agreement 249 119), CESAR (Grant Agreement 271 022), METANET4U (Grant Agreement 270 893) and META-NORD (Grant Agreement 270 899).



# WERREJ CONTENTS

## IL-LINGWA MALTIJA FL-ERA DIGITALI

<b>1</b>	<b>Sommarju Eżekuttiv</b>	<b>1</b>
<b>2</b>	<b>Riskju għal-Lingwi Tagħna u Sfida għat-Teknoloġija Lingwistika</b>	<b>4</b>
2.1	Il-Konfini Lingwistiċi jfixklu s-Socjetà Ewropea tal-Infommazzjoni . . . . .	5
2.2	Il-Lingwi Tagħna f'Riskju . . . . .	5
2.3	It-Teknoloġija Lingwistika hija Teknoloġija Katalizzanti Ewlenija . . . . .	6
2.4	Opportunitajiet għat-Teknoloġija Lingwistika . . . . .	6
2.5	Sfidi li t-Teknoloġija Lingwistika Taffaċċja . . . . .	7
2.6	Il-Ksib tal-Lingwi tal-bnedmin u tal-magni . . . . .	8
<b>3</b>	<b>Il-Malti fis-Socjetà tal-Infommazzjoni Ewropea</b>	<b>10</b>
3.1	Fatti Ġenerali . . . . .	10
3.2	Partikolaritajiet tal-Lingwa Maltija . . . . .	11
3.3	Żviluppi riċenti . . . . .	12
3.4	Il-Kultivazzjoni tal-Lingwa f'Malta . . . . .	13
3.5	Il-Lingwi fl-Edukazzjoni . . . . .	14
3.6	Aspetti internazzjonali . . . . .	16
3.7	Il-Malti fuq l-Internet . . . . .	17
<b>4</b>	<b>Appoġġ ta' Teknoloġija Lingwistika għall-Malti</b>	<b>22</b>
4.1	Arkitetturi ta' Applikazzjonijiet . . . . .	22
4.2	Oqsma ewlenin ta' applikazzjoni . . . . .	24
4.3	Oqsma oħra tal-applikazzjoni . . . . .	32
4.4	Programmi tal-Edukazzjoni . . . . .	33
4.5	Programmi u Sforzi Nazzjonali . . . . .	35
4.6	Disponibbiltà ta' Għodod u Riżorsi . . . . .	36
4.7	Tqabbil ta' Trans-Lingwi . . . . .	38
4.8	Konkluzjonijiet . . . . .	39
<b>5</b>	<b>Dwar META-NET</b>	<b>42</b>

# THE MALTESE LANGUAGE IN THE DIGITAL AGE

<b>1</b>	<b>Executive Summary</b>	<b>43</b>
<b>2</b>	<b>Languages at Risk: a Challenge for Language Technology</b>	<b>46</b>
2.1	Language Borders Hold back the European Information Society . . . . .	47
2.2	Our Languages at Risk . . . . .	47
2.3	Language Technology is a Key Enabling Technology . . . . .	48
2.4	Opportunities for Language Technology . . . . .	48
2.5	Challenges Facing Language Technology . . . . .	49
2.6	Language Acquisition in Humans and Machines . . . . .	49
<b>3</b>	<b>The Maltese Language in the European Information Society</b>	<b>51</b>
3.1	General Facts . . . . .	51
3.2	Particularities of the Maltese Language . . . . .	52
3.3	Recent Developments . . . . .	54
3.4	Language Cultivation in Malta . . . . .	55
3.5	Language in Education . . . . .	56
3.6	International Aspects . . . . .	57
3.7	Maltese on the Internet . . . . .	58
<b>4</b>	<b>Language Technology Support for Maltese</b>	<b>62</b>
4.1	Application Architectures . . . . .	62
4.2	Core Application Areas . . . . .	63
4.3	Other Application Areas . . . . .	71
4.4	Educational Programmes . . . . .	72
4.5	National Projects and Initiatives . . . . .	73
4.6	Availability of Tools and Resources . . . . .	75
4.7	Cross-language comparison . . . . .	76
4.8	Conclusions . . . . .	77
<b>5</b>	<b>About META-NET</b>	<b>80</b>
<b>A</b>	<b>Referenzi – References</b>	<b>81</b>
<b>B</b>	<b>Membri ta' META-NET – META-NET Members</b>	<b>87</b>
<b>C</b>	<b>Is-Serje ta' White Papers ta' META-NET – The META-NET White Paper Series</b>	<b>91</b>

## SOMMARJU EŻEKUTTIV

Matul dawn l-aħħar sittin sena, l-Ewropa saret struttura politika u ekonomika distinta, iżda kulturalment u lingwistikament għadha diversa ħafna. Dan ifisser li mill-Portugiz għall-Pollakk u t-Taljan għall-Islandiż, ta' kuljum il-komunikazzjoni bejn iċ-ċittadini tal-Ewropa kif ukoll il-komunikazzjoni fl-oqsma tan-negozju u l-politika hija inevitabbilment ikkonfrontata minn ostakli lingwistiċi. L-istituzzjonijiet tal-UE jonfqu madwar biljun euro fis-sena fuq iż-żamma tal-politika tagħhom tal-multilingwiżmu, jiġifieri, it-traduzzjoni ta' testi u l-interpretar tal-komunikazzjoni mitkellma. Madankollu, għandu dan ikun ta' daqsekk piż? It-teknoloġija moderna tal-lingwi u r-riċerka lingwistika jistgħu jkunu ta' kontribuzzjoni sinifikanti biex jitwaqqgħu dawn l-ostakli lingwistiċi. Meta kkombinati mal-apparati u l-applikazzjonijiet intelliġenti, it-teknoloġija lingwistika tkun fil-futur tista' tgħin lill-Ewropej jtkellmu ma' xulxin b'mod faċli u jagħmlu n-negozju ma' xulxin anke jekk ma jkunux jtkellmu lingwa komuni.

---

### It-teknoloġija lingwistika tibni rabtiet.

---

L-ostakli tal-lingwa jistgħu jwaqqfu n-negozju, b'mod speċjali għall-SMEs li m'għandhomx il-mezzi finanzjarji biex iregġgħu lura s-sitwazzjoni. L-unika alternattiva (inkonċepibbli) għal din it-tip ta' Ewropa multilingwali tkun li tippermetti lingwa waħda biex tiegħu pożizzjoni dominanti u tispicċa tissostitwixxi kull lingwa oħra. Mod klassiku ta' kif tista' tegħleb l-ostakli tal-lingwa huwa li titgħallem il-lingwi barranin. Iżda

mingħajr appoġġ teknoloġiku, li tkun taf it-23 lingwa uffiċjali tal-Istati Membri tal-Unjoni Ewropea u xi sit-tin lingwa Ewropea oħra huwa ostaklu insormontabbli għaċ-ċittadini tal-Ewropa u l-ekonomija, id-dibattitu politiku, u l-progress xjentifiku tagħha.

Is-soluzzjoni hija li jinbnew teknoloġiji ewlenin li jippermettu dan. Dawn ikunu joffru lill-atturi Ewropej vantaġġi tremendi, mhux biss fis-suq komuni Ewropew iżda wkoll f'relazzonijiet ta' kummerċ ma' pajjiżi terzi, b'mod speċjali l-ekonomiji emergenti. Biex jintlaħaq dan il-għan u tiġi preservata d-diversità kulturali u lingwistika tal-Ewropa, huwa neċessarju li l-ewwel issir analiżi sistematika tal-partikolaritajiet tal-lingwi Ewropej kollha u l-istat attwali tal-appoġġ tat-teknoloġija lingwistika li hawn għalihom. Soluzzjonijiet tat-teknoloġija lingwistiċi se jservu eventwalment bħala rabta unika bejn il-lingwi tal-Ewropa.

L-għodda ta' traduzzjoni awtomatizzata u tal-ipproċessar tad-diskors li huma attwalment disponibbli fis-suq għadhom pjuttost lura minn dan il-għan ambizzjuż. L-atturi dominanti f'dan il-qasam huma primarjament l-intrapriżi b'sidien privati għall-profit ibbażati fl-Amerka ta' Fuq. Diġà fl-aħħar tal-1970, l-UE rrealizzat ir-relevanza profonda tat-teknoloġija lingwistika bħala sewwieq lejn l-unità Ewropea, u bdiet tiffinanzja l-ewwel proġetti ta' riċerka tagħha, bħall-EUROTRA. Fl-istess żmien, proġetti nazzjonali ġew imwaqqfa u ġġeneraw riżultati ta' valur iżda qatt ma wasslu għal azzjoni Ewropea kkonċertata. F'kuntrast ma' dan l-isforz ta' finanzjament ferm selettiv, soċjetajiet multilingwali oħra bħall-Indja (22 lingwa uffiċjali) u l-Afrika t'Isfel



(l-il lingwa ufficjali) waqqfu reċentement programmi nazzjonali fit-tul għar-riċerka tal-lingwi u l-iżvilupp tat-teknoloġija.

L-atturi predominanti fit-teknoloġija lingwistika llum jiddependu fuq approċċi ta' statistika mhux preċiżi li ma jagħmlux użu minn metodi lingwistiċi u għarfien aktar fondi. Per eżempju, is-sentenzi jiġu awtomatikament tradotti billi tiġi kkumparata sentenza ġdida ma' eluf ta' sentenzi oħrajn li jkunu ġew tradotti qabel minn umani. Il-kwalità tar-risultat il-biċċa l-kbira tiddependi fuq l-ammont u l-kwalità tal-kampjun tal-korp disponibbli. Filwaqt li t-traduzzjoni awtomatika ta' sentenzi sempliċi fil-lingwi b'ammonti suffiċjenti ta' materjal ta' testi disponibbli tista' tikseb risultati utli, metodi ta' statistika vojta bħal dawn huma destinati li jfallu fil-każ ta' lingwi b'korp ta' kampjuni tal-materjal ferm iżgħar jew fil-każ ta' sentenzi bi strutturi kumplessi.

---

It-teknoloġija lingwistika bħala ċavetta għall-futur.

---

Għalhekk l-Unjoni Ewropea ddecidiet li tiffinanzja proġetti bħall-EuroMatrix u l-EuroMatrixPlus (mill-2006) u iTranslate4 (mill-2010) li jwettqu riċerka bażika u applikata u jiġġeneraw riżorsi biex jiġu stabbiliti soluzzjonijiet għat-teknoloġija lingwistika ta' kwalità għolja għall-lingwi Ewropej kollha. Li tanalizza l-proprjetajiet strutturali aktar fondi tal-lingwi huwa l-uniku pass 'il quddiem jekk irridu nibnu applikazzjonijiet li jaħdmu tajjeb tul il-firxa kollha tal-lingwi tal-Ewropa.

Ir-riċerka Ewropea f'dan il-qasam diġà kisbet numru ta' suċċessi. Per eżempju, is-servizzi ta' traduzzjoni tal-Unjoni Ewropea issa jużaw is-softwer ta' traduzzjoni bil-magni b'sors miftuħ MOSES li ġie prinċipalment żviluppat permezz ta' proġetti ta' riċerka Ewropej. F'Malta, l-oqsma tat-teknoloġija lingwistika l-aktar avanzati bħalissa huma daww tas-sintezi tat-taħdit u l-korpora tat-testi: fil-qasam tas-sintezi tat-taħdit bil-Malti, proġett appoġġjat mill-Gvern parzjalment iff-

nanzjat mill-fond ta' żvilupp reġjonali tal-UE qiegħed fil-proċess li jwassal it-teknoloġija tat-taħdit lill-persuni b'diżabilità. Il-konsorzju, li jikkonsisti f'SME (Crimson Wing Ltd), fondazzjoni (FITA, Fundazzjoni għall-Access tat-TI), u l-Università, wiegħed li dawn iriżorsi se jkunu disponibbli għal skopijiet ta' riċerka. Fil-qasam tal-korpora tat-testi, is-Server għar-Riżorsi Lingwistiċi bil-Malti (MLRS) qed jagħti frott u sforzi sinifikanti li għadhom għaddejjin fl-Università, permezz tal-Istitut tal-Lingwistika (A. Gatt, C. Borg, R. Fabri) u d-Dipartiment ta' Sistemi Intelligenti tal-Kompjuter (M. Rosner), li jsostnu u jiżviluppaw dan. Bħalissa il-korpus jinkludi madwar 100M kelma, u hemm aktar għodod ippjanati inkluż tagger għall-kategoriji tal-kliem u ċekkjatur ortografiku.

---

It-teknoloġija lingwistika tgħin biex tgħaqqad lill-Ewropa.

---

Jekk inħarsu lejn l-għarfien miksub s'issa, jidher li t-teknoloġija 'ibrida' tal-lingwi tal-lum li thallat metodi ta' pproċessar fondi ma' daww ta' statistika se tkun tista' timla' l-vojt ta' bejn il-lingwi Ewropej kollha u aktar. Bħalma din is-serje ta' *white papers* turi, hemm differenza drammatika fl-istat ta' pronteżza fir-rigward ta' soluzzjonijiet ta' lingwi u l-istat tar-riċerka bejn l-Istati Membri tal-Ewropa. Din il-*white paper* għall-lingwa Maltija turi li hemm il-potenzjal għal industrija tat-teknoloġija lingwistika u ambjent tar-riċerka f'Malta. Izda għalkemm numru ta' teknoloġiji u riżorsi jeżisti, hemm ħafna anqas minn lingwi Ewropej li huma "akbar" u ċertament mhux biżżejjed biex tiġi appoġġjata l-firxa kompleta ta' applikazzjonijiet sensittivi għall-lingwi li huma disponibbli għal daww il-lingwi l-oħra. Skont il-valutazzjoni ddetaljata f'dan ir-rapport, il-kisba ta' suċċess fit-teknoloġija tal-lingwa Maltija tirrekkjedi ċiklu sħiħ ta' bidliet li jkun jinvolvi fornituri ta' kontenut, żviluppaturi u utenti tat-teknoloġija lingwistika.

Xi bidliet fil-politika tal-lingwa nazzjonali jridu jiġu implimentati qabel ma xi suċċessi għall-Lingwa Maltija jkunu jistgħu jinkisbu.

L-għan fit-tul tal-META-NET huwa li tiġi introdotta teknoloġija lingwistika bi kwalità għolja għall-lingwi kollha sabiex tinkiseb l-unità politika u ekonomika permezz tad-diversità kulturali. It-teknoloġija tkun tgħin biex tkisser l-ostakli eżistenti u jinbnew ir-rabtiet bejn il-lingwi tal-Ewropa. Dan jirrikjedi l-partijiet interessati

kollha – fil-politika, ir-riċerka, in-negozju, u s-soċjetà – biex jingħaqdu l-isforzi tagħhom fil-futur.

Din is-serje ta' *white papers* tikkomplimenta azzjonijiet strateġiċi oħra mehuda minn META-NET (ara l-appendiċi għal harsa ġenerali). Informazzjoni aġġornata bħall-verżjoni attwali tal-*vision paper* [4] tal-META-NET jew l-Aġenda għar-Riċerka Strateġika (SRA) jistgħu jinstabu fuq il-websajt tal-META-NET <http://www.meta-net.eu>.

## RISKJU GĦAL-LINGWI TAGĦNA U SFIDA GĦAT-TEKNOLOĠIJA LINGWISTIKA

Qed naraw rivoluzzjoni diġitali li qed tħalli impatt b'mod drammatiku fuq il-komunikazzjoni u s-soċjetà. Żviluppi riċenti fit-teknoloġija tal-komunikazzjoni diġitizzata u tan-networks xi drabi jiġu mqabbla mal-invenzjoni tal-istampar ta' Gutenberg. X'tista' tgħidilna din l-analoġija dwar il-futur tas-soċjetà tal-informazzjoni Ewropea u b'mod partikolari l-lingwi tagħna?

---

Ir-revoluzzjoni diġitali hija komparabbli mal-invenzjoni tal-istamperija ta' Gutenberg.

---

Wara l-invenzjoni ta' Gutenberg, skoperti reali fil-komunikazzjoni u skambju ta' għarfien kienu mwettqa permezz ta' sforzi bħat-traduzzjoni tal-Bibbja ta' Luther għal-lingwa komuni. Fis-sekli sussegwenti, tekniki kulturali ġew żviluppati biex jimmaniġġjaw aħjar l-ipproċessar tal-lingwa u l-iskambju ta' għarfien:

- l-istandardizzazzjoni ortografika u grammatikali ta' lingwi kbar ippermettiet it-tixrid rapidu ta' ideat xjentifiċi u intellettuali godda;
- l-iżvilupp ta' lingwi uffiċjali għamel possibbli għaċ-ċittadini li jikkomunikaw bejn ċerti konfini (ta' spiss politici);
- it-tagħlim u t-traduzzjoni tal-lingwi ppermettew skambju bejn il-lingwi;
- il-ħolqien ta' linji gwida ġurnalistiċi u biblijografici żguraw il-kwalità u d-disponibilità ta' materjal ip-printjat;

- il-ħolqien ta' midja differenti bħal gazzetti, radju, televiżjoni, kotba, u formati oħra ssodisfaw htigijiet differenti ta' komunikazzjoni.

Fl-aħħar għoxrin sena, it-teknoloġija tal-informatika (TI) għenet biex hafna mill-proċessi jiġu awtomatizzati u ffacilitati:

- is-software ta' desktop publishing jissostitwixxi l-ittajpar u t-typesetting;
- il-Microsoft PowerPoint tissostitwixxi t-trasparenzi tal-projectors;
- il-posta elettronika tibgħat u tircievi dokumenti aktar malajr minn fax;
- Skype jagħmel telefonati bl-internet u jospita laqgħat virtwali;
- il-formati ta' kodifikazzjoni audio u video jagħmluha faċli għal skambju ta' kontenut multimidjali;
- il-magni ta' tiftix jipprovdu aċċess ibbażat fuq kelma ewlenija għall-paġni web;
- is-servizzi fuq l-internet bħal Google Translate jiproduċu traduzzjonijiet ta' malajr u approssimattivi;
- il-pjattaformi tal-midja soċjali jiffacilitaw il-kollaborazzjoni u l-qsim ta' informazzjoni.

Għalkemm għodod u applikazzjonijiet bħal dawn huma utli, bħalissa ma jistgħux jappoġġjaw b'mod suffiċjenti soċjetà tal-informazzjoni multilingwi Ewropea sostenibbli, soċjetà moderna u inklussiva fejn l-informazzjoni u l-merkanzija jistgħu jiċċirkolaw b'mod hieles.

## 2.1 IL-KONFINI LINGWISTIČI JFIXKLU S-SOĊJETÀ EWROPEA TAL-INFORMAZZJONI

Ma nistgħux inbassru b'mod preċiż kif se tkun is-soċjetà tal-informazzjoni tal-ġejjieni. Mandankollu, hemm probabbiltà li r-revoluzzjoni fit-teknoloġija tal-komunikazzjoni tgħaqqad lin-nies li jtkellmu lingwi differenti b'modi ġodda. Din titfa' pressjoni kemm fuq individwi biex jitgħallmu lingwi ġodda u speċjalment kemm għal iżviluppaturi ta' softwer biex joħolqu applikazzjonijiet ta' teknoloġija ġodda biex jassiguraw ftehim komuni u aċċess għal għarfien kondivizibbli. F'ekonomija globali u spazju ta' informazzjoni, aktar lingwi, kelliema u kontenut jikkonfrontawna u jirrikjeduna li ninteraġixxu malajr ma' tipi ġodda ta' midja. Il-popolarità kurrenti ta' midja soċjali (Wikipedia, Facebook, Twitter u YouTube) hija biss parti zghira minn stampa akbar.

---

L-ekonomija globali u s-spazju ta' informazzjoni jikkonfrontana ma' lingwi, kelliema u kontenut differenti.

---

Illum, aħna nistgħu nittrażmettu gigabytes ta' test madwar id-dinja fi ftit sekondi qabel nagħrfu li huwa b'lingwa li aħna ma nifhmux. Skont rapport riċenti mitlub mill-Kummissjoni Ewropea, 57% tal-utenti tal-internet fl-Ewropa jixtru oġġetti u servizzi b'lingwi li mhumiex il-lingwa nattivha tagħhom. (l-Ingliż huwa l-ilsien barrani l-aktar komuni segwit mill-Franċiż, il-Ġermaniż u l-Ispanjol.) 55% tal-utenti jaqraw kontenut f'lingwa barranija filwaqt li 35% biss jużaw lingwa oħra biex jiktbu ittri elettronici jew jibagħtu kummenti fuq il-web [5]. Ftit snin ilu, l-Ingliż seta' kien il-*lingua franca* tal-web – il-maġġoranza l-kbira tal-kontenut fuq il-web kien bl-Ingliż – iżda s-sitwazzjoni issa nbidlet drastikament. L-ammont ta' kontenut fuq l-internet b'lingwi

oħra (partikolarment lingwi tal-Asja u Għarab) kiber f'daqqa waħda.

Firda digitali li tinsab kullimkien u li hija kkawżata minn konfini lingwistiċi sorprendentement ma kisbitx ħafna attenzjoni fid-diskors pubbliku; madankollu, qajmet kwistjoni urġenti ħafna, “Liema lingwi Ewropej se jirnexxu u jippersistu fis-soċjetà tal-informazzjoni u l-għarfien, ibbażata fuq in-networks?”

## 2.2 IL-LINGWI TAGĦNA F'RISKJU

L-istampar ikkontribwixxa għal skambju imprezzabbli ta' informazzjoni fl-Ewropa, iżda wassal ukoll għall-estinzjoni ta' bosta lingwi Ewropej. Lingwi reġjonali u dawk f'minoranza rarament ġew stampati. Bħala riżultat ta' dan, ħafna lingwi bħal Cornish jew Dalmatian spiss kienu limitati għal forom orali ta' trasmissjoni, li llimita l-adozzjoni kontinwa, it-tixrid u l-użu tagħhom.

---

Il-varjetà wiesgħa tal-lingwi fl-Ewropa hija assi kulturali l-aktar sinjuri u importanti tagħha.

---

Il-lingwi tal-Ewropa, madwar 80, huma wieħed mill-assi l-aktar prezżjużi tagħha u l-aktar importanti f'dak li huwa assi kulturali. In-numru kbir ta' lingwi Ewropej huwa wkoll parti vitali mis-suċċess soċjali tagħha [6]. Filwaqt li l-lingwi popolari bħall-Ingliż jew l-Ispanjol żgur li se jzommu l-preżenza tagħhom fis-soċjetà digitali u s-suq li qed jitfaċċaw, bosta lingwi Ewropej jistgħu jinqatgħu mill-komunikazzjonijiet digitali u jsiru irrilevanti għas-soċjetà tal-internet. Żviluppi bħal dawn żgur li ma jkunux mixtieqa. Minn naħa, opportunità strateġika tista' tintilef li jista' jkun id-għajjef il-pożizzjoni globali tal-Ewropa. Min-naħa l-oħra, żviluppi bħal dawn jistgħu jmorru kontra l-għan ta' parteċipazzjoni ugwali għal kull ċittadin

Ewropew irrispettivament mil-lingwa. Skont rapport tal-UNESCO dwar il-multilingwiżmu, il-lingwi huma mezz essenzjali għat-tgawdija tad-drittijiet fundamentali, bhall-espressjoni politika, l-edukazzjoni u l-partecipazzjoni fis-socjeta [7].

## 2.3 IT-TEKNOLOGIJA LINGWISTIKA HIJA TEKNOLOGIJA KATALIZZANTI EWLENIJA

Fil-passat, l-isforzi tal-investment iffukaw fuq l-edukazzjoni u t-traduzzjoni tal-lingwi. Pereżempju, skont ċerti stimuli, is-suq Ewropew għat-traduzzjoni, l-interpretazzjoni, il-lokalizzazzjoni ta' software, u l-globalizzazzjoni ta' websajts kien ta' EUR 8.4 biljun fl-2008 u kien mistenni li jikber b'10% fis-sena [8]. Madankollu, din il-kapaċità eżistenti mhix biżżejjed biex tissodisfa l-htigijiet kurrenti u futuri.

It-teknologija lingwistika hija teknologija katalizzanti ewlenija li tista' thares u trawwem lingwi Ewropej. It-teknologija lingwistika tgħin lin-nies jikkollaboraw, iwettqu negozju, jaqsmu l-għarfien u jieħdu sehem f'dibattiti soċjali u politiċi irrispettivament mill-ostakoli lingwistiċi jew il-hiliet tal-kompjuter. It-teknologija lingwistika diġà qed tassisti kompiti ta' kuljum, bħal kitba ta' ittri elettronici, twettiq ta' tiftix fuq l-internet jew prenotazzjoni ta' titjiriet. Ahna nibbenefikaw minn teknologija lingwistika meta:

- nsibu informazzjoni permezz ta' magni ta' tiftix fuq l-internet;
- niċċekkjaw l-ortografija u l-grammatika fil-word processor;
- naraw rakkomandazzjonijiet dwar prodotti f'hanut online;
- nisimghu struzzjonijiet verbali ta' sistema ta' navigazzjoni;

- nitraduċu paġni web permezz ta' servizz fuq l-internet.

It-teknologija lingwistiċi deskritti f'dan id-dokument huma parti essenzjali minn applikazzjonijiet futuri innovattivi. It-teknologija lingwistika hija teknologija li tipikament tippermetti xogħol f'qafas ta' applikazzjoni akbar bħal sistema ta' navigazzjoni jew magna ta' tiftix. Dawn il-White Papers jiffokaw fuq il-prontezza ta' teknologiji ewlenin għal kull lingwa.

---

L-Ewropa teħtieġ teknologija robusta u affordabbli għal-lingwi kollha Ewropej.

---

Fil-futur qrib, inkunu neħtieġu teknologija lingwistika għal-lingwi kollha Ewropej li tkun disponibbli, bi prezz raġonevoli u integrata sewwa f'ambjenti ta' software akbar. Esperjenza interattiva, multimedjali u multilingwi tal-utent mhijiex possibbli mingħajr teknologija lingwistika.

## 2.4 OPPORTUNITAJIET GĦAT-TEKNOLOGIJA LINGWISTIKA

It-teknologija lingwistika tista' twettaq traduzzjoni awtomatika, produzzjoni ta' kontenut, ipproċessar ta' informazzjoni u ġestjoni ta' għarfien possibbli għal-lingwi Ewropej kollha. It-teknologija lingwistika tista' wkoll tkompli l-iżvilupp ta' interfaces intuittivi bbażati fuq il-lingwa għal elektronika tad-dar, makkinarju, vetturi, kompjuters u robots. Għalkemm hafna prototipi diġà jeżistu, l-applikazzjonijiet kummerċjali u industrijali għadhom fi stadji bikrin ta' żvilupp. Kisbiet riċenti fir-riċerka u l-iżvilupp ħolqu tieqa ġenwina ta' opportunità. Pereżempju, it-traduzzjoni awtomatika (TA) diġà qed tagħti ammont raġonevoli ta' preċiżjoni fi hdan dominji

speċifiċi, u applikazzjonijiet esperimentali jipprovdu informazzjoni multilingwi u ġestjoni ta' għarfien kif ukoll produzzjoni ta' kontenut f'ħafna ilsna Ewropej.

Applikazzjonijiet lingwistiċi, interfaces għall-utenti bbażati fuq il-vuċi u sistemi ta' djalogu jinsabu tradizzjonalment f'dominji speċjalizzati ħafna, u ħafna drabi jagħtu prestazzjoni limitata. Qasam wieħed attiv ta' riċerka huwa l-użu tat-teknoloġija lingwistika għal operazzjonijiet ta' salvataġġ f'zoni ta' diżastri. F'ambjenti bħal dawn ta' riskju għoli, l-eżattezza tat-traduzzjoni tista' tkun kwistjoni ta' ħajja jew mewt. L-istess raġunament japplika għall-użu tat-teknoloġija lingwistika fl-industrija tal-kura tas-saħħa. Robots intelligenti b'kapacitajiet lingwistiċi u bejn il-lingwi għandhom il-potenzjal li jsalvaw il-ħajjiet.

Hemm opportunitajiet tas-suq enormi fl-educazzjoni u l-industrija tad-divertiment għall-integrazzjoni ta' teknoloġiji lingwistiċi fil-logħob, offerti ta' edudivertiment, ambjenti ta' simulazzjoni jew programmi ta' taħriġ. Servizzi mobbli ta' informazzjoni, software għat-tagħlim tal-lingwi assistit mill-kompjuter, ambjenti tal-eLearning, għodod għal awtoevalwazzjoni u software għal sejbien ta' plagjarizmu huma biss ftit eżempji oħra fejn it-teknoloġija lingwistika jista' jkollha rwol importanti. Il-popolarità ta' applikazzjonijiet soċjali tal-midja bħal Twitter u Facebook jissuġġerixxu ħtieġa akbar għal teknoloġiji lingwistiċi sofistikati li jistgħu jissorveljaw postijiet, iwettqu sommarji ta' diskussjonijiet, jissuġġerixxu tendenzi ta' opinjonijiet, jiskopru reazzjonijiet emozzjonali, jidentifikaw ksur tad-drittijiet tal-awtur jew użu ħażin ta' sistemi ta' kompjuters.

---

It-teknoloġija lingwistika tgħin biex tegħleb id-"dizabilità" tad-diversità lingwistika.

---

It-teknoloġija lingwistika tirrappreżenta opportunità kbira għall-Unjoni Ewropea li tagħmel sens kemm ekonomikament kif ukoll kulturalment. Il-

multilingwizmu fl-Ewropa sar ir-regola. Negożzi, organizzazzjonijiet u skejjet Ewropej huma wkoll multinazzjonali u diversi. Iċ-ċittadini jridu jikkomunikaw bejn il-konfini lingwistiċi li għadhom jeżistu fis-Suq Komuni Ewropew. It-teknoloġija lingwistika tista' tgħin biex jingħelbu dawn l-ostakli li fadal waqt li tappoġġja l-użu ħieles u miftuħ tal-lingwi. Barra minn hekk, teknoloġija lingwistika innovattiva u multilingwi għall-Ewropej tista' wkoll tgħinna nikkomunikaw mal-imsieħba globali tagħna u l-komunitajiet multilingwi tagħhom. It-teknoloġiji lingwistiċi jappoġġjaw l-ammont kbir ta' opportunitajiet ekonomiċi internazzjonali.

## 2.5 SFIDI LI T-TEKNOLOĠIJA LINGWISTIKA TAFFAĊĊJA

---

Il-pass kurrent tal-progress teknoloġiku progress huwa bil-wisq.

---

Għalkemm it-teknoloġija lingwistika għamlet progress konsiderevoli fl-aħħar ftit snin, il-pass preżenti tal-progress teknoloġiku u l-innovazzjoni tal-prodotti miexi bil-mod wisq. Teknoloġiji lingwistiċi b'użu wiesa', bħal karatteristiċi ta' ortografija u grammatika fil-word processors, huma tipikament monolingwi, u huma disponibbli biss għal numru żgħir ta' lingwi. Servizzi ta' traduzzjoni awtomatika fuq l-internet huma eċċellenti fil-holqien ta' approssimazzjoni tajba ta' kontenut f'dokument, iżda huma mimlija diffikultajiet varji meta jkun meħtieġa traduzzjonijiet preċiżi ħafna u kompluti. Minħabba l-kumplessità tal-lingwa umana, nimmudellaw ilsiena b'software u nittestjawom fid-dinja vera huwa proċess twil u għoli li jeħtieġ impenn ta' fondi sostnuti. L-Ewropa għandha għalhekk tmantni l-irwol pijuniera tagħha fl-affaċċjar ta' sfidi teknoloġiċi ta' komunita multi lingwali billi tivvinta metodi godda sabiex



taċċellera l-iżvilupp dritt madwar il-mappa. Dawn jistgħu jinkludu kemm avvanzi ta' kompjutazzjoni u kemm dawk tekniċi bħal *crowdsourcing*.

## 2.6 IL-KSIB TAL-LINGWI TAL-BNEDMIN U TAL-MAGNI

Sabiex nuru kif il-kompjuters jimmaniġġjaw il-lingwa u għaliex il-ksib tal-lingwi huwa komputu diffiċli ħafna, nagħtu ħarsa fil-qosor lejn il-mod kif il-bnedmin jiksibu l-ewwel u t-tieni lingwa, imbagħad nagħmlu skeċċ kif sistemi tat-traduzzjoni awtomatika jahdmu – hemm raġuni għaliex il-qasam tat-teknoloġija lingwistika huwa marbut mill-qrib mal-qasam tal-intelligenza artifiċjali. Il-bnedmin jiksibu l-ħiliet lingwistiċi permezz ta' żewġ modi differenti. L-ewwel, it-tarbija titgħallem lingwa billi tisma' l-interazzjoni bejn il-kelliema tal-lingwa. Espozizzjoni għal eżempji konkreti lingwistiċi mill-utenti tal-lingwi, bħal ġenituri, aħwa u membri oħra tal-familja, jgħinu lit-trabi mill-età ta' madwar sentejn jew viċin jipproduċu l-ewwel kelmiet u frażijiet qosra tagħhom. Dan huwa possibbli biss minħabba d-dispożizzjoni ġenetika speċjali li għandhom il-bnedmin għat-tagħlim tal-lingwi.

It-tagħlim tat-tieni lingwa normalment jeħtieġ sforz ferm aktar meta tifel jew tifla ma jkollhiex immersjoni fil-komunità lingwistika ta' kelliema nattivi. Fl-età skolastika, lingwi barranin jinkisbu ġeneralment permezz ta' tagħlim tal-istrutturi grammatikali, il-vokabularju u l-ortografija tagħhom minn kotba u materjal edukattiv li jiddeskrivu l-għarfien lingwistiku f'termini ta' regoli as-tratti, tabelli u testi bħala eżempji.

---

Il-bnedmin jiksibu l-ħiliet lingwistiċi permezz ta' żewġ modi differenti: talli jtagħlmu eżempji u talli jtagħlmu ir-regoli bażiċi tal-lingwa.

---

Iż-żewġ tipi ewlenin ta' sistemi ta' teknoloġija lingwistika jakkwistaw kapacitajiet lingwistiċi b'mod simili bħall-bnedmin. Metodi statistiċi jiksibu għarfien lingwistiku minn ġbir vast ta' testi konkreti bħala eżempji f'lingwa waħda jew fl-hekk imsejha testi paralleli li huma disponibbli f'żewġ lingwi jew aktar. L-algoritmi tat-tagħlim awtomatiku jiffurmaw ċerta fakultà lingwistika li tista' tikseb mudelli ta' kif kliem, frażijiet qosra u sentenzi sħaħ jintużaw b'mod korrett f'lingwa waħda jew jiġu tradotti minn lingwa għal oħra. In-numru kbir ta' sentenzi li metodi statistiċi jeħtieġu huwa enormi. Il-kwalità tal-prestazzjoni tiżdied hekk kif in-numru ta' testi analizzati jiżdiedu. Huwa komuni li sistemi bħal dawn jitharrġu fuq testi li jinkludu miljuni ta' sentenzi. Din hija waħda mir-raġunijiet għaliex fornituri ta' magni ta' tiftix huma herqana li jiġbru kemm jista' jkun materjal bil-miktub. Il-korrezzjoni ortografika f'word processors, l-informazzjoni disponibbli fuq l-internet, u s-servizzi ta' traduzzjoni bħal Google Search u Google Translate jiddependu fuq metodu statistiku (mmexxi minn dejta).

---

Iż-żewġ tipi prinċipali tas-sistemi tat-teknoloġija langwistika jakkwistaw lingwi b'mod simili.

---

Sistemi bbażati fuq regoli huma t-tieni tip ewleni ta' teknoloġija lingwistika. Esperti mil-lingwistika, lingwistika kompjutazzjonali u x-xjenza tal-kompjuter jikkodifikaw analiżi grammatikali (regoli tat-traduzzjoni) u jikkumpilaw listi ta' vokabularju (diz-zjunarji). It-twaqqif ta' sistema bbażata fuq regoli jieħu ħafna ħin u jinvolvi xogħol intensiv. Sistemi bbażati fuq regoli jeħtieġu wkoll esperti speċjalizzati sew. Uħud mis-sistemi ta' traduzzjoni awtomatika bbażati fuq regoli kienu taħt żvilupp kostanti għal aktar minn għoxrin sena. Il-vantaġġ tas-sistemi bbażati fuq regoli huwa li l-esperti jistgħu jikkontrollaw b'mod aktar dettaljat l-ipproċessar tal-lingwa. Dan jagħmel possibbli li

l-izbalji jiġu kkoreġuti b'mod sistematiku fis-software u tingħata informazzjoni dettaljata lill-utent, speċjalment meta s-sistemi bbażati fuq regoli jintużaw għat-tagħlim tal-lingwi. Minhabba restrizzjonijiet finanz-

jarji, teknoloġija lingwistika b'sistemi bbażati fuq regoli hija possibbli għal-lingwi ewlenin biss.



# IL-MALTI FIS-SOĊJETÀ TAL-INFORMAZZJONI EWROPEA

## 3.1 FATTI ĠENERALI

Il-Malti huwa l-lingwa nazzjonali tal-arċipelagu Malti, li jikkonsisti fil-gżejjer ta' Malta, Għawdex u Kemmuna. Flimkien mal-Ingliż, il-Malti huwa wkoll l-ilsien uffiċjali ta' Malta. Skont id-Demographic Review 2009 mill-Uffiċċju Nazzjonali tal-Istatistika ta' Maltavi, l-istima tal-popolazzjoni Maltija (minbarra l-barranin) fl-aħħar tas-sena 2009 kienet 396,278. Huwa stmat li llum, minhabba l-fażijiet tal-emigrazzjoni minn Malta l-aktar fil-hamsinijiet u s-sittinijiet, bejn wieħed u iehor l-istess numru ta' kelliema nattivi espatrijati jgħixu barra mill-pajjiż (l-aktar fir-Renju Unit, l-Awstralja, l-Istati Uniti u l-Kanada).

Għalkemm il-Malti jappartjeni għall-fergħa Għarbija tan-Nofsinhar tal-familja lingwistika Semitika, huwa pjuttost differenti mill-ilsna neo-Għarbin l-oħra. L-istruttura tiegħu huwa r-rizultat ta' sitwazzjonijiet ta' kuntatt lingwistiċi differenti li ffurmaw taħt mexxejja differenti tal-gżejjer fil-kors ta' millennju. Filwaqt li l-qalba tal-Malti hija Semitika, fiha wkoll superstrat Rumanz u adstrat Ingliż. Barra minn hekk, il-Malti huwa l-uniku lsien Semitiku miktub bl-alfabett Latin (modifikat).

Il-qalba Semitika tal-Malti tniisset mill-konkwista Għarbija fit-870 AD u l-popolazzjoni mill-ġdid sussegwenti tagħha permezz ta' nies li ġew jgħixu f'Malta li jtkellmu bl-Għarbi. L-ewwel kuntatt dirett ma' lingwi Rumanzi ġie stabbilit fl-1090 meta Malta nħakmet minn Normanni, li ġabu l-Isqalli magħhom, filwaqt li

l-popolazzjoni kienet għadha qed tuża l-Għarbi vernakulari tagħha fil-ħajja ta' kuljum. Malta kienet aktar u aktar maqtugħa politikament, kulturalment u lingwistikament mid-dinja Għarbija. Fis-sekli ta' wara, taħt l-influwenza tal-ilsna Rumanz tal-mexxejja, aktar u aktar kliem Rumanz misluf daħal fid-djalett Għarbi. Meta Malta kienet taħt il-hakma Ingliża fl-1800, il-lingwa uffiċjali nbidlet mit-Taljan għall-Ingliż, li ġab miegħu numru dejjem akbar ta' kliem Ingliż misluf fil-Maltivii. Il-sentenza li ġejja meħuda minn artiklu ta' gazzetta (*l-Orizzont* mis-7 ta' Settembru, 1995; riprodott f' [9, p. 135]) tista' turi l-influwenzi differenti tal-ilsna f'kuntatt (kliem Rumanz misluf huwa b'tipa grassa, kliem mill-Ingliż sottolinjat):

- (1) Il-hold-up sar minn zaghżugh li kien liebes nuċċali skur tax-xemx.

Wieħed mill-fatti notevoli dwar il-Malti huwa li minkejja n-numru relattivament żgħir tal-kelliema tiegħu u z-zona żgħira fejn hu mitkellem, hemm numru pjuttost rikk ta' varjanti jew djaletti. B'mod ġenerali, distinzjoni ewlenija tista' ssir bejn il-varjetà standard mitkellma fiz-zoni urbani bħall-Belt Valletta u tas-Sliema u varjetajiet mhux standard mitkellma fiz-zoni rurali. Barra minn Malta, il-Malti mitkellem fl-Awstralja żviluppa f'etnolett uniku msejjaħ *Maltraljan* [10]. Dan huwa differenti mill-Malti Standard prinċipalment f'dak li huwa l-lessiku tiegħu (jigifieri, il-vokabularju) li huwa r-rizultat ta' kliem misluf b'mod estensiv mill-Ingliż (Aw-

straljan) u bidla sussegwenti fit-tifsira. Minhabba li l-Ingliż huwa t-tieni ilsien uffċjali f’Malta, hafna Maltin huma bilingwi. Bejn il-pilastri tal-monolingwiżmu u l-bilingwiżmu sħiħ, hemm sekwenza ta’ taħlit ta’ lingwi u codeswitching. Fid-dar u bejniethom, hafna Maltin jikkellmu biss bil-Malti. L-Ingliż, min-naħa l-oħra, huwa l-lingwa użata fil-kuntest ta’ kitba f’edukazzjoni oġġla u fil-komunikazzjoni mal-barranin.

## 3.2 PARTIKOLARITAJIET TAL-LINGWA MALTIJA

Il-Malti huwa l-uniku lsien Semitiku fl-Unjoni Ewropea u l-uniku lsien Semitiku miktub bl-alfabett Latin. L-alfabett Malti jagħmel użu minn xi grafemi speċjali li jvarjaw minn alfabetti oħra Latini (il-valuri tal-hoss huma mogħtija fl-Alfabett Fonetiku Internazzjonali): ċ [tʃ], ġ [dʒ], għ (il-biċċa l-kbira siektra), ħ [h], ż [z] [3, 11]. Xi karatteristiċi partikolari tal-Malti huma:

- ordni tal-kliem ħielsa
- morfologija Semitika
- sistema temporali bbażata fuq l-aspett
- nuqqas ta’ infinitiv morfologiku

---

L-ordni tal-kliem huwa relattivament ħieles fis-sentenzi Maltin.

---

Anki jekk ma fiħx trufijiet tal-kazi, il-Malti għandu ordni ta’ kliem ħielsa hafna. Is-sentenza *Il-kelb gidem il-qattusa lbierah* għandha l-ordni tal-kliem S(uġġett) V(erb) O(ġġett) iżda tista’ wkoll tiġi espressa bħala:

- (2) *Ilbierah il-kelb gidem il-qattusa.*  
 yesterday the-dog (m) he.bit the-cat (f)  
 (SVO)  
 ‘Yesterday, the dog bit the cat.’

- (3) *Gidem il-qattusa l-kelb lbierah.*  
 he.bit the-cat (f) the-dog (m) yesterday  
 (VOS)

‘The dog bit the cat yesterday.’

- (4) *Il-qattusa gidimha l-kelb lbierah.*  
 the-cat (f) he.bit.her the-dog (m) yesterday  
 (OVS)

‘The cat, it was bitten by the dog yesterday.’

Kif it-traduzzjonijiet bl-Ingliż jippruvaw juru, l-ordnijiet tal-kliem differenti għandhom enfasi differenti fit-tifsira. Fl-ewwel zewġ eżempji, l-ordni tal-kliem mhux immarka, bl-oġġett wara l-verb. Fl-aħħar eżempju, l-oġġett *il-qattusa* jippreċedi l-verb. Kif jissemma f’ [12, p. 140], din l-ordni tal-kliem hija mmarkata u tenfasizza l-oġġett għal kuntrast. Bl-oġġett quddiem, kelliema nattivi jippreferu jimmarkaw *il-qattusa* bl-enklitika tal-oġġett *-ha* mal-verb. Barra minn hekk, fid-diskors, dan il-kuntrast huwa espress b’intonazzjoni differenti. L-ordni tal-kliem fit-tieni eżempju (VOS) tista’ tintuża biex tesprimi tifsira kuntrastiva kif ukoll, b’intonazzjoni xierqa, tqiegħed l-enfasi fuq *gidem il-qattusa*. Mingħajr din it-tifsira kuntrastiva (u mingħajr l-intonazzjoni kuntrastiva) l-enfasi tkun fuq il-fatt innifsu bħal: ‘Ma smajtx dak li ġara lbierah? Il-kelb gidem il-qattusa lbierah!’ (Fabri, konverżazzjoni personali).

---

Kliem Maltin jistgħu jinbidlu internament matul inflessjoni u derivazzjoni.

---

Bħala lsien Semitiku, il-Malti juri morfologija mhux konkatenattiva, jiġifieri inflettiva u l-forom ta’ kliem imnisslin jinbidlu internament:

F’lingwi bħall-Ingliż, il-forom tal-kliem huma magħmula minn zkuk u affissi, jiġifieri b’mod konkatenattiv. Il-verb *shoot* jista’ jkun ikkonjugat fit-terza persuna biż-żieda tal-affiss *-s* maz-zokk bħal f’ (*he*) *shoot-s*. Barra minn hekk, miz-zokk verbali jista’ jitnissel nom billi

jizdied l-affiss *-er* bħal f' *shoot-er*. Għaldaqstant kemm l-inflessjoni kif ukoll id-derivazzjoni jsehħu mingħajr tib-dil fl-istruttura interna, jiġifieri b' mod konkatenattiv.

Fil-Malti, hemm tahlita ta' morfologija bbażata fuq iz-zokk morfemiku u morfologija bbażata fuq l-għerq u l-mudell. Fil-komponent Semitiku, l-“unità” bażika f'kelma ta' spiss ma tkunx iz-zokk iżda l-għerq magħmul minn tliet (xi kultant erba') konsonanti f'ordni fissa li ġgħorr magħha tifsira generali. Zkuk tal-kliem bit-tifsira speċifika tagħhom huma fformati billi l-konsonanti jiġu organizzati skont ċertu mudell. Pereżempju, l-għerq *k-t-b* ihaddan it-tifsira ta' dak kollu marbut mal-“kitba”. F'dan li ġej, il-mudelli huma rappreżentati bħala numri 1,2,3 għall-konsonanti tal-għerq u v għall-vokali bejniethom, pereżempju 1v2v3. Meta wiehed japplika l-mudell 1v2v3 u jimla l-pożizzjonijiet tal-vokali bejn il-konsonanti tal-għerq 1,2 u 3 bis-sekwenza tal-vokali *i-e*, wiehed jifforma l-verb *kiteb*. L-inflessjoni ta' dan verb għall-plural issir bit-twaħħil tal-affiss tal-plural *-u*, li tagħti l-forma *kitbu*. L-applikazzjoni tal-1v22v:3 mal-għerq tagħti n-nom aġent *kittieb*. L-inflessjoni tan-nom biz-żieda tal-affiss *-a* tagħti l-plural *kittieba*. Wiehed jinnota li s-suffiss tal-plural *-a* jixbah lil markatur għall-femminili *-a* sabiex *kittieba* tista' wkoll tirreferi għal kittieb femminili. Is-suffissi l-oħra Semitiċi Maltin tal-plural huma *-in* bħal fi *mħallef*, *imħallfin*; *-at/ -iet* bħal f' *kittieba*, *kittiebat*; *-ijiet* bħal fi *żmien*, *żminijiet*. Nomi fil-plural fil-Malti jistgħu jiġu fformati wkoll b' mod mhux konkatenattiv (l-hekk imsejja forom ta' plural miksur), jiġifieri l-ebda affiss ma jizdied, iżda n-nom jindidel internament, pereżempju *kiteb* vs. *kitba*.

Verbi mislufa illum huma importati l-aktar permezz ta' klassi ta' verbi speċjali li tista' takkomoda zkuk mhux maħduma [13]. Pereżempju, iz-zokk *park-* bl-Ingliż sar il-baži tal-forom tal-verbi Maltin *pparkjajt*, *pparkjajt*, *pparkja*. Illum, din il-klassi ta' verbi speċjali li qabel kienet klassi marginali Semitika żdiedet fid-daqs minhabba l-influss ta' verbi mislufa mill-Ingliż. Dan

huwa produttiv ferm, ħafna drabi iwassal għal self *ad-hoc* ta' verbi bl-Ingliż li diġà għandhom kontroparti Semitika bil-Malti. Pereżempju 'to download (a file)' tista' tiġi espress bl-użu tal-verb Semitiku *nizzel* (originarjament tfsisser 'he caused to come down'). Meta wiehed jieħu z-zokk Ingliż *download* u jimportah permezz tal-klassi tal-verb speċjali dan minflok jagħti forom bħal *ddawnlowdjajt*, *ddawnlowdjat*, *ddawnlowdja*. Din l-istrateġija tiġi ta' spiss ikkritikata li qed tikkorrompi l-lingwa [3].

---

### Is-sistema temporali tal-Malti hija bbażata fuq l-aspett.

---

Il-verbi bil-Malti huma aċcentwati għall-aspett, jiġifieri jekk azzjoni tkunx kompluta (perfettiva) jew mhux kompluta (mhux perfettiva) – għal rapport shiħ dwar tempi u aspetti fil-Malti, ara [14, 15]. Fin-nuqqas ta' kwalunkwe markaturi grammatikali oħra, verbi perfettivi huma interpretati bħala “it-temp tal-passat” u verbi li mhumiex perfettivi bħala “it-temp tal-preżent”: *Andrew kiteb*; *Andrew jikteb*. Il-kumbinazzjoni tal-verb li mhuwiex perfettiv ma' kien, tesprimi passat abitwali: *Andrew kien jikteb*. Iż-żieda tal-kelma *qed* ‘progressiva’ (bħall-forma bl-Ingliż *-ing*) tagħti *Andrew kien qed jikteb* eċċ. Il-verbi Maltin ma għandhomx infinittivi morfologiċi. Għaldaqstant, fi predikati kumplessi bħal fis-sentenza bl-Ingliż ‘Andrew wants to write’, iż-żewġ verbi huma morfologikament finiti: *Andrew jrid jikteb* (literalment: ‘Andrew he wants he writes’) anki jekk minna semantika, *jikteb* mhuwiex finit.

## 3.3 ŻVILUPPI RIĊENTI

Bl-avanz tal-Ingliż għal-lingwa internazzjonali u l-lingwa tat-teknoloġija wara t-Tieni Gwerra Dinjija, l-ammont ta' kliem misluf mill-Ingliż fil-Malti kiber b' mod sostanzjali. Ħafna minnhom saru “nattivi”, jiġifieri ġew adottati fl-użu regolari tant li anki kliem meħud

mis-Semitiku ma jistax jieħu posthom. Pereżempju, minflok il-kelma użata ta' spiss *emphajruport* (mill-Ingliż *airport*), il-kelma Semitika *mitjar* kienet proposta (imnissla minn *tar* 'he flew'). Madankollu, din qatt ma ġiet aċċettata mill-komunità lingwistika. Min-naħa l-oħra, kliem misluf jidhol fil-lingwa pjuttost malajr, jiġi impurtat b'mod spontanju, anki jekk diġà hemm kliem Malti tajjeb għalihom (pereżempju *ddawnlowdja* vs *nizzel* 'he downloaded'). Dan iqabbad biżżejjat fost xi wħud li l-lingwa tista' ssir "korrotta" [3].

Żvilupp ieħor riċenti għall-Malti huwa l-istatus tiegħu bħala lingwa uffiċjali tal-Unjoni Ewropea. Dan għandu kemm vantaġġi, kif ukoll żvantaġġi [3]. Minn naħa waħda, il-Malti finalment sar ilsien rikonoxxut internazzjonalment, status li ma kellux għal żmien twil, kien imwarrab bħala l-"lingwa tal-kċina" fis-sekli ta' qabel. Min-naħa l-oħra, it-tradutturi tal-UE Maltin qed jaffaċċjaw ċerti sfidi: bosta termini tekniċi u legali għad iridu jiġu "ivvintati" bil-Malti. Dan jirriżulta eventwalment f'espansjoni lessikali tal-lingwa (tabilhaqq aspett pozittiv), li, madankollu, għandu jiġi kkoordinat minn korpus ċentrali sabiex tradutturi individwali ma johlqux termini differenti għall-istess kunċetti indipendentement minn xulxin (li hija problema serja). Il-korp ċentrali biex jitratta din l-isfida huwa l-Kunsill Nazzjonali għall-Ilsien Malti.

Żviluppi oħra fis-snin riċenti jikkonċernaw l-ortografija Maltija. Il-Malti (flimkien mal-Ingliż) sar l-ilsien uffiċjali ta' Malta fl-1 ta' Jannar, 1934 – bl-ortografija maħruġa mill-Għaqda tal-Kittieba tal-Malti fl-1924. Minn dakinhar, l-ortografija għaddiet minn tliet revizjonijiet (1984, 1992 u 2008).

L-aħħar riforma ġiet rilaxxata fl-2008. L-għan tagħha kien li jitnaqqsu l-inċertezzi tal-kittieba li jirriżultaw minn numru konsiderevoli ta' varjanti ortografiċi għal ċerti kliem. Kif id-dokument *Deċiżjonijiet 1* [16] tal-Kunsill jirrimarka, ammont kbir ta' varjanti jista' jitnaqqas billi jinstab bilanċ konsistenti bejn l-ortografija

grammatikali u fonetika. Għaldaqstant l-erba' varjanti *zobtu*, *zoptu*, *sobtu* u *soptu* jistgħu jitnaqqsu għal żewġ varjanti għal *zoptu* ['zɔp.tɔ] u *soptu* ['sɔp.tɔ]. Għal raġuni simili, il-kelma *skond* [skɔnt] 'according to' nbidlet għal *skont* minħabba li l-forom grammatikali tagħha l-oħra ma jiġġustifikawx ortografija b' *d* (imnissla minn *secondo* Taljan), bħal pereżempju *skontok* ['skɔn.tɔk] 'according to you'.

Għat-tielet qasam (tal-kliem misluf), il-prinċipju jibqa' li l-kliem misluf jinkiteb skont l-ortografija Maltija jekk dawn jitqiesu bħala "nattivi" u jekk dan ma johloqx kunflitti fil-pronunzja jew ma' regoli oħra tal-kitba Maltija. Madankollu, ħafna Maltin jippreferu li jiktbu kliem Ingliż misluf bl-ortografija oriġinali tagħhom, minħabba li drawhom. Fil-fatt, waqt seminar pubbliku dwar l-użu ta' kliem Ingliż misluf f'April 2008, kien hemm diskussjoni-jiet emozzjonali fost l-udjenza fil-każ ta' kliem bħal *email* u l-ortografija l-ġdida tagħha propost bħala *imejl*. Fatturi bħal drawwiet tal-komunità lingwistika jagħmlu l-istandardizzazzjoni tal-ortografija saħansitra aktar diffiċli milli li jinstab bilanċ bejn il-prinċipji grammatikali u fonetiċi [17]. Dawn l-eżempji jagħtu biss idea żgħira tal-ħidma iebes li l-Kunsill Nazzjonali għall-Ilsien Malti qed iwettaq bħala parti mill-kultivazzjoni tal-lingwa f'Malta. Is-sezzjoni li jmiss se tagħti ħarsa lejn l-istorja tal-kultivazzjoni tal-lingwa f'Malta.

### 3.4 IL-KULTIVAZZJONI TAL-LINGWA F'MALTA

Meta mqabbel ma' lingwi oħra tal-Ewropa, l-istatus tal-Malti bħala lsien uffiċjali (mill-1934) fih innifsu huwa żvilupp riċenti. Għaldaqstant il-kultivazzjoni tal-lingwa wkoll bdiet tard. Għal sekli sħaħ, il-Malti kien biss il-mezz mitkellem tal-popolazzjoni Maltija u kien imwarrab meta mqabbel mal-lingwa uffiċjali rispettiva tal-mexxejja ta' Malta. Dan beda jinbidel mal-moviment tal-lingwa ta' nofs-/tmieni is-seklu 18 meta

l-ewwel studji sistematiki lingwistiċi tmexxew minn Agius de Soldanis (1750) u Mikiel Anton Vassalli (1797). Speċjalment Vassalli ppromwova l-ilsien Malti billi nkorraġixxa l-użu tiegħu f'kull qasam tal-ħajja ta' kuljum. It-traduzzjonijiet tal-bibbja ta' Fortunato Panzavecchia f'nofs is-seklu 19 ikkontribwixxew għal aktar standardizzazzjoni tal-lingwa [18]. Barra minn hekk, permezz tal-pass lejn ortografija standardizzata fil-bidu tas-seklu 20, ittiegħed pass importanti mill-fondazzjoni tal-Għaqda tal-Kittieba tal-Malti fl-1920. Is-sistema ortografika, li għet żviluppata minn din l-organizzazzjoni, saret l-ortografija uffiċjali ta' Malta fl-1934 u, b'xi bidliet u żidiet, ilha tintuża minn dak iż-żmien.

Fl-1964, wara li nkisbet l-indipendenza mill-Gran Britanja, l-istatus tal-Malti bħala lingwa nazzjonali u bħala lingwa uffiċjali flimkien mal-Ingliż inkiteb fil-kostituzzjoni. Meta Malta ssiehbet fl-UE fl-2004, il-Malti sar ilsien uffiċjali tal-UE. Kif issemma fit-taqsimat t'hawn fuq, dan wassal għal ċerti sfidi, li jistgħu jissolvew biss minn korpus li jikkoordina l-istandardizzazzjoni u l-prassi komuni fix-xogħol tat-traduzzjoni.

Il-korp f'Malta biex jagħmel dan ix-xogħol huwa l-Kunsill Nazzjonali għall-Ilsien Malti. Dan twaqqaf fl-2005 bħala l-ewwel organizzazzjoni tal-gvern sabiex tittratta uffiċjalment kwistjonijiet lingwistiċi u ppjanar lingwistiku għal-lingwa Maltija. Il-kompiti tal-Kunsill huma, kif imnizzlin fl-Att tal-Ilsien Malti (ATT Nru V tal-2004): il-jippromwovi l-ilsien Malti, "jadotta politika, pjan u strateġija lingwistika xierqa" u jwettaq dan fil-prattika. Xogħol ieħor importanti tal-Kunsill huwa li jagġorna l-ortografija Maltija u jiddeċiedi fuq ortografija korretta (jieħu f'idejha l-inkarigu mill-Akkademja tal-Malti u b'hekk ikun prinċipalment responsabbli għar-riforma għall-ortografija Maltija tal-2008). Fuq is-sitweb tiegħu, il-Kunsill joffri wkoll korsijiet ta' taħriġ għall-qarrejja tal-provi u korsijiet tal-lingwa Maltija għall-barranin [19].

Qabel twaqqaf il-Kunsill, l-istandardizzazzjoni tal-ortografija kienet il-kompitu tal-Akkademja tal-Malti. Din originat fl-1964 mill-Għaqda tal-Kittieba tal-Malti, li kienet il-korp li waqqaf l-ewwel ortografija uffiċjali fl-1924/1932. Illum l-għan prinċipali tal-Akkademja huwa li tippromwovi studji akkademiki fil-lingwa u l-letteratura Maltija, tippromwovi l-użu tal-Malti f'kull qasam tal-ħajja ta' kuljum u tibni kuntatti mal-persuni li huma hbieb tal-lingwa u li jużawha barra minn Malta [20]. L-Akkademja taħdem mill-qrib flimkien mal-Kunsill Nazzjonali għall-Ilsien Malti.

Il-motivazzjoni wara l-Att tal-Ilsien Malti kienet l-idea li lingwa nazzjonali waħda li hija kondiviza mill-individwi kollha fi hdan dak in-nazzjon tiffirma l-bażi għall-identità kulturali u nazzjonali. Dan naturalment jehtieg standardizzazzjoni tal-lingwa. Fil-fatt, mill-moviment tal-kultivazzjoni tal-lingwa mis-seklu 19 sal-lum, il-Malti avvanza minn ilsien imwarrab u vernakulari kif kien qabel għal ilsien nazzjonali ta' prestiġju għoli. Dan jidher ukoll fl-ammont dejjem jikber ta' xogħlijiet letterarji bil-Malti matul l-istess perjodu ta' żmien u fin-numru kbir ta' organizzazzjonijiet influwenti u l-korpi għal-lingwa u l-letteratura Maltija (ara [3]).

### 3.5 IL-LINGWI FL-EDUKAZZJONI

Partikolarment f'soċjetà bilingwali bħal dik ta' Malta, diversi aspetti għandhom rwol meta niġu għal-lingwa fl-edukazzjoni. Aspett wieħed huwa l-lingwa ta' istruzzjoni, jiġifieri l-lingwa li tintuża uffiċjalment mill-għalliema matul il-lezzjonijiet fl-iskola jew fis-seminars fl-università. Fattur ieħor huwa l-lingwa użata f'ċerti kotba tal-iskola. Bl-Ingliż bħala l-lingwa tax-xjenzi teknoloġiċi u naturali, hafna mill-kotba tal-iskola dwar dawn is-suġġetti huma bl-Ingliż. Fil-fatt, l-isforzi biex jiġu tradotti termini tekniċi u xjentifiċi għall-Malti ltaqgħu ma' bosta problemi, waħda minnhom hija l-aċċettazzjoni mill-komunità tal-lingwa. Għaldaqstant is-suġġetti skolastiċi, ukoll, possibbilment jiddeter-



minaw il-lingwa ta' istruzzjoni għal ċerti lezzjonijiet, għalkemm jista' jkun ukoll li l-kotba tal-iskola bl-Ingliż (u t-terminologija bl-Ingliż li tinsab fihom) jintużaw waqt li l-lingwa tat-tagħlim tkun il-Malti.

Madankollu aspett ieħor huwa l-lingwa użata mill-individwi. Kelliema bilingwi mhux biss jużaw lingwi differenti f'kuntesti soċjali differenti ("dominji"), eż. il-Malti ma' tal-familja fid-dar, l-Ingliż mal-barranin, il-Malti jew l-Ingliż matul il-lezzjonijiet tal-iskola eċċ. Dawn għandhom tendenza wkoll li jużaw iż-żewġ lingwi flimkien, jew iħalltu ż-żewġ lingwi (eż. kliem bl-Ingliż jithalltu f'konverżazzjoni li tkun qed issir bil-Malti) jew permezz tal-codeswitching (eż. konverżazzjoni bil-Malti tinqaleb għall-Ingliż u lura għall-Malti, bil-partijiet tal-Ingliż ikunu akbar minn kliem biss waħedhom, iżda spiss jikkonsistu minn diversi sentenzi). Għaldaqstant anki matul il-lezzjonijiet tal-iskola li jiġu mgħallma b'lingwa waħda, il-konverżazzjonijiet bejn l-għalliema u l-istudenti jistgħu jaqilbu bejn il-lingwi [21].

Meta wiehed iżomm dawn it-tliet fatturi f'moħhu, wiehed jinduna li l-espożizzjoni attwali tal-istudenti għal-lingwa rispettiva fl-iskejjel jew fl-università hija xi haġa differenti mil-lingwa magħżula ta' istruzzjoni. Rigward il-lingwa uffċjali ta' istruzzjoni fl-educazzjoni, kemm il-Malti kif ukoll l-Ingliż jintużaw fl-iskejjel u fl-università, minhabba li l-Malti u l-Ingliż jaqsmu l-istatus bħala lingwi uffċjali ta' Malta. Fl-iskejjel, it-tnejn li huma jiġu mgħallma bħala suġġetti minn età bikrija. Liema lingwa tintuża bħala l-lingwa ta' istruzzjoni jiddependi mit-tip ta' skola. Skejjel privati għandhom it-tendenza li jużaw l-Ingliż aktar mill-Malti (xi kultant b'mod aktar estensiv), filwaqt li fl-iskejjel Maltin tal-istat il-Malti huwa kemxejn preferut mill-Ingliż. L-iskejjel tal-knisja għandhom il-preferenzi individwali tagħhom jiġifieri li xi whud tradizzjonalment jippreferu lingwa waħda minn oħra.

Kif issemma qabel, il-biċċa l-kbira tal-kotba tax-xjenza li jintużaw fl-iskola huma bl-Ingliż. Għaldaqstant, bl-introduzzjoni ta' aktar u aktar suġġetti xjentifiċi aktar 'il quddiem fl-iskola u iżjed fl-università, l-istudenti huma esposti għal żewġ lingwi fl-istess ħin, li jintużaw f'sitwazzjonijiet differenti: jista' jkollhom il-lezzjonijiet tagħhom mgħallma bil-Malti, iżda jaqraw il-kotba tagħhom u jiktbu l-essays tagħhom bl-Ingliż. Speċjalment għal studenti tal-università, il-konverżazzjonijiet bejniethom, mal-ħbieb u l-lecturers spiss isefħu bil-Malti, xi kultant jużaw il-code-switching/iħalltu bejn il-Malti jew ikunu saħansitra bl-Ingliż biss (tal-aħħar pereżempju ma' studenti internazzjonali jew lecturers). Madankollu, fid-dar mal-familja tagħhom u l-ħbieb, hafna Maltin jitekellmu bil-Malti, xi whud iħalltu l-lingwi u f'it familji biss jitekellmu bl-Ingliż biss.

Kif jidher mill-eżempji ta' hawn fuq, minkejja l-fatt li kemm il-Malti kif ukoll l-Ingliż jintużaw bħala lingwi fl-educazzjoni, hemm distribuzzjoni ċara meta niġu għall-użu tagħhom fis-soċjetà. Sciriha u Vassallo (2001, p. 29, iċċitati f' [3]) isemmu li "70% ta' dawk li wiegħbu qalu li jużaw il-Malti fuq ix-xogħol, filwaqt li 90% qalu li jikkomunikaw mal-membri tal-familja tagħhom fid-dar bil-Malti. ... il-persentaġġi għall-Malti mitkellem huma għolja hafna iżda jonqsu f'hiliet oħra bħall-qari u l-kitba." Din id-distribuzzjoni tal-Malti li qed jintuża prinċipalment bħala l-mezz mitkellem u l-Ingliż bħala l-mezz tal-kitba toħloq ċertu riskju, minhabba li jista' jkollha impatt fuq il-hiliet differenti tal-kelliema nativa tagħha f'dak li għandu x'jaqsam ma' taħdit, qari jew kitba. Sabiex wiehed jagħti r-raġunijiet għal dan il-fatt, wiehed għandu jħares lejn il-karatteristiċi bażiċi tal-lingwa mithaddta u dik miktuba.

B'mod ġenerali, it-testi miktubin jiddistingwu ruħhom mid-diskors f'numru ta' modi. Li għandhom komuni huwa li t-tnejn huma modi ta' trasferiment ta' informazzjoni bejn iż-żewġ partijiet, jiġifieri l-kelliem u min qed jisma, u l-kittieb u l-qarrej, rispettivament. Madankollu,

huma differenti fil-mod kif l-informazzjoni tgħaddi bejniethom. Fi kliem sempliċi, test miktub, kuntrarju għal diskors, isehh barra minn sitwazzjoni komunikattiva, interattiva u konkreta. Minn naha, id-diskors jiddependi fuq l-interazzjoni bejn il-kelliem u min qed jisma'. Il-kelliem irid jibni struttura tal-informazzjoni b'ċertu mod. Dan huwa importanti minhabba l-memorja limitata u qasira tal-bniedem: min qed jisma' fil-konverżazzjoni jista' jassorbi ċertu ammont ta' informazzjoni biss qabel ma jkollu jinterrompi u jsaqsi lill-kelliem biex jiżgura li fehmem.

Test miktub, min-naha l-oħra, mhuwiex interattiv safejn il-qarrej ma' jistax jitlob għal aktar informazzjoni speċifika. Madankollu huwa jista' jara x'hemm 'il quddiem u lura fit-test (xi haġa li min qed jisma' ma jistax jagħmilha fid-diskors). B'dan il-mod, it-test innifisu miktub iservi bhala memorja fit-tul għall-qarrej. Għaldaqstant, test miktub jistruttura l-informazzjoni b'mod differenti minn kif isir f'konverżazzjoni mithaddta. Pereżempju, test għandu jipprovdi aktar informazzjoni ta' sfond sabiex jagħti bażi komuni lill-qarrej qabel tibda għaddejja l-informazzjoni attwali. Din ma tkunx problema, jekk it-test jista' jservi bhala memorja fit-tul għall-qarrej. Fil-fatt, dan jippermetti struttura aktar elaborata mid-diskors, jiġifieri normalment ikun fih sentenzi itwal u ammont oġġla ta' propożizzjonijiet subordinati.

Din id-distinzjoni tar-reġistru (jiġifieri "l-istil tal-lingwa") hija dik li fil-letteratura, eż. [22], issejthet strutturi ta' testi *orali* versus *letterati*. Tabilhaqq, test jista' jkun miktub f'reġistru orali li jixbah konverżazzjonijiet mithaddta (eż. f'forum ta' iċċettjar jew posta elettronika informali). Izda dan mhuwiex ir-reġistru normalment użat pereżempju f'essays. Idealment, il-kelliema nattivi jiksbu r-reġistru letterat digà minn età żgħira, eż. permezz tal-ġenituri tagħhom li jaqrawhom l-istejjer. Aktar tard fl-iskola, dan l-għarfien jissahha minn, pereżempju, eżercizzju attiv tal-kitba tal-essays.

Reġistru letterat jizviluppa maż-żmien f'lingwa bi tradizzjoni letterarja. Il-Malti, meta mqabbel mal-istorja qasira tiegħu bhala lsien uffċjali miktub (mill-1934) għandu storja letterarja twila u rikka. Anki jekk l-eqdem letteratura skoperta hija skarsa hafna (*Il Cantilena* minn Pietro Caxaro, li tmur lura għal madwar l-1450), tradizzjoni letterarja bdiet tiffirma madwar l-erbghinijiet fis-seklu 17. Fis-seklu 19, l-ammont ta' letteratura bil-Malti kienet qed tikber [3], u flimkien magħha, il-Malti kien qed jespandi. Illum huwa lsien li għandu reġistru letterat komplut.

Madankollu, dan ir-reġistru, jehtieg li jiġi pprattikat sabiex jinżamm l-istatus tal-lingwa bhala lingwa kemm konverżazzjonali kif ukoll letterarja. It-tendenza fl-educazzjoni oġġla biex jinkitbu essays aktar bl-Ingliż milli bil-Malti, mill-anqas teoretikament, tohloq ir-riskju li l-Malti jibqa' jintuza f'reġistru orali biss. Ammont oġġla ta' websajts Maltin tal-ġeneri kollha huwa mixtieq biex wiehed ikopri ż-żewġ reġistri u s-sottotipi tagħhom u jiġi żgurat status stabbli tal-lingwa fir-rikkezza kollha tagħha.

### 3.6 ASPETTI INTERNAZZJONALI

Meta wiehed iżomm f'moħħu t-taqsimiet precedenti, issa għandu jkun mifhum li l-aspetti internazzjonali tal-Malti huma pjuttost differenti minn lingwi oħra. B'anqas minn miljun kelliema nattivi madwar id-dinja, il-Malti huwa kkunsidrat bhala lingwa "mitkellma anqas". Fl-istorja tiegħu, il-Malti ma kienx l-ilsien tal-okkupanti izda wiehed ta' dawk li qed jokkupaw il-post. Bhala rizultat ta' dan, il-Malti qatt ma kien meqjus bhala lingwa internazzjonali jew lingua franca kif kien il-każ eż. tal-Latin, l-Ispanjol, il-Portugiz jew l-Ingliż, li kollha huma l-lingwi tal-konkwistaturi. Il-Malti tabilhaqq infirex lejn pajjiżi oħra, fejn għadu mitkellem sal-lum (l-Awstralja, il-Kanada, l-Istati Uniti u r-Renju Unit), izda bhala lingwa tal-komunità biss. Kien jehtieg kważi 200 sena mill-ewwel interess tal-

grammatici Maltin fil-lingwa tagħhom sakemm eventwalment kiseb l-istatus ta' lingwa uffiċjali. Saħansitra dakinhar, il-lingwa uffiċjali l-oħra, l-Ingliż, serviet bhala l-lingwa għar-relazzjonijiet internazzjonali. Il-bidla biex il-Malti jsir ilsien internazzjonalment viżibbli seħhet mas-šhubija ta' Malta fl-UE fl-2004. Minn dakinhar, il-Malti huwa lsien uffiċjali fl-Unjoni Ewropea, flimkien bil-benefiċċji u l-isfidi kollha li huma marbutin ma' dan l-istatus.

Akkademikament, l-interess fil-Malti bhala sugġett tax-xjenza jmur lura sal-1603 meta Hieronymus Megiser ippubblika t-*Thesaurus Polyglottus* tiegħu, li kien jinkludi lista ta' kliem bil-Malti. L-ewwel studjuż li b'mod sistematiku esplora u ppromwova l-lingwa Maltija kien Mikiel Anton Vassalli. Huwa ppubblika grammatika (1790), dizżjunarju (1797) u alfabetti diversi (1788 u 1790) għall-Malti u llum huwa msejjaħ "il-missier tal-Ilsien Malti" [23]. Fis-seklu 20, kien ippubblikat il-*Grammar of the Maltese Language* (1936) ta' Sutcliffe. Mis-sittinijiet tas-seklu 20, il-Lingwistika tal-Ilsien Malti kisbet għarfien akkademiku internazzjonali permezz tal-pubblikazzjonijiet ta' Joseph Aquilina (eż. *Papers in Maltese Linguistics* (1961) u *Maltese-English Dictionary*, żewġ volumi (1987 and 1990)). Minn dak iż-żmien, aktar u aktar studjużi barra minn Malta wrew interess fil-Malti. L-2007 rat it-twaqqif tal-Għaqda Internazzjonali tal-Lingwistika Maltija [24], assoċjazzjoni ta' lingwisti li huma interessati fil-lingwa Maltija. L-għan ewlieni tal-GHILM, kif jidher fuq is-sitweb tagħha, huwa li ttiprovdi "konnessjoni bejn studjużi interessi fil li gejjin mid-dixxiplina kollha tal-Lingwistika", b'hekk tiffacilita r-riċerka dwar il-Malti. Din l-għaqda trid ukoll li tgħaqqad flimkien nies minn sfondi differenti li jaħdmu bl-ilsien Malti (lingwisti, tradutturi, studenti u oħrajn).

## 3.7 IL-MALTI FUQ L-INTERNET

Stharriġ tal-Uffiċċju Nazzjonali tal-Istatistika ta' Malta fit-tieni kwart tal-2009 [25] juri li fost il-popolazzjoni ta' madwar 400,000 ruħ, 67 fil-mija kellhom aċċess għall-kompjuter u 64 fil-mija kellhom aċċess għall-internet. Stharriġ riċenti tal-Ewrobarometru (ippubblikat f'Mejju tal-2011) [26] dwar id-drawwiet tat-tiftix fuq l-internet fost l-utenti Ewropej wera li huma biss 6.5 fil-mija tal-utenti tal-internet Maltin li jużaw eskusivament il-Malti fuq l-internet meta jaqraw, jikkunsmaw il-kontenut jew jikkomunikaw. Minflok, 90.6 fil-mija jagħzlu li jibbrawzjaw is-websajts bl-Ingliż u 20.1 fil-mija bit-Taljan, rispettivament. Dawn iċ-ċifri ifurmaw il-bażi tal-artiklu fil-gazzetta Maltija li toħroġ kuljum *The Times of Malta*, li holoq diskussjoni interessanti l-aktar fost il-qarrejja Maltin tal-edizzjoni fuq l-internet [27]. Madankollu, ir-rizultati eżatti tal-istharriġ, iwasslu għall-konkluzjoni li din l-abitudni mhix għażla maħsuba: Meta mistoqsija liema lingwa l-Maltin jikkunsidraw bhala l-lingwa materna tagħhom, 89.5 fil-mija ta' dawk li wiegħbu qalu li l-Malti kien l-ilsien nattiv tagħhom (meta mqabbel ma' 7.6 fil-mija biss għall-Ingliż u 0.2 fil-mija għat-Taljan).

Il-lingwi l-oħra barra dik użata minn dawk li wiegħbu biex jaqraw jew jaraw kontenut fuq l-internet kienu l-Ingliż (90.6 fil-mija) u t-Taljan (20.1 fil-mija). 6.5 fil-mija biss wiegħbu li jużaw il-lingwa tagħhom, li mhux fatt sorprendenti, minħabba li hafna Maltin huma bil-lingwali bil-Malti u bl-Ingliż u numru konsiderevoli jtkellmu bit-Taljan ukoll. F'dak li għandu x'jaqsam ma' kitba fuq l-internet, in-numri favur il-Malti huma oghla meta l-utenti jaqraw jew jaraw kontenut: 87 fil-mija qalu li jużaw il-Malti, 85 fil-mija l-Ingliż u 8 fil-mija t-Taljan. Ir-raġuni li l-maġġoranza jużaw l-Ingliż bhala l-lingwa biex jikkunsmaw kontenut fuq l-internet tista' tkun sempliciment in-numru limitat ta' websajts bil-Malti minflok il-preferenza għall-Ingliż fiha nnifisha. Niftakru li hafna minn dawk li wiegħbu ma jikkunsidrawx l-Ingliż



bhala l-lingwa tagħhom u li l-użu tal-Malti żdied meta gie prodott kontenut fuq il-web, anki jekk dan l-użu tal-Malti fil-maġġoranza tal-każijiet iseħħ f'forums ta' iċċettjar u pjattaformi soċjali, għalhekk fi stil ta' lingwaġġ tat-taħdit, jiġifieri fir-registru orali.

Karatteristika partikolari dwar il-Malti użat mill-generazzjoni żagħżugħa fi pjattaformi soċjali u forums ta' iċċettjar hija l-ortografija fonetika tagħha, mingħajr il-karattri bħal *gh* siekta u *l-h*. Għaldaqstant *ghax* tinkiteb *ax*, *tieghi* *tiei* eċċ. Ir-raġuni għal dan tista' tkun l-introduzzjoni tard tal-karattri speċjali tal-Malti fid-dinja tal-PC. Minkejja li l-Malti gie implimentat fil-qafas tal-Unicode mill-bidu tiegħu, kompjuters u sistemi ta' operazzjoni segwew ħafna aktar tard. L-Awtorita Maltija tal-Istandards ħarġet forma standardizzata ta' tastiera Maltija fl-2002, u s-sistema ta' operazzjoni Windows tal-Microsoft kienet disponibbli fil-verżjoni tal-lingwa Maltija fl-2006 biss (mal-Windows XP). Fil-każ ta' telefonijiet ċellulari, l-ittri speċjali Maltin għadhom ma ġewx implimentati. Għaldaqstant naraw jekk l-ortografija *ad hoc* tal-forums tal-iċċettjar hix se twitti t-triq għal ortografija b'karattri speċjali ladarba dawn ikunu disponibbli fuq it-telefonijiet ċellulari jew jekk din l-ortografija fonetika se tkompli teżisti bhala "soċjolekt" tal-generazzjoni żagħżugħa [28].

Fir-rigward tal-ammont ta' Malti fuq l-internet b'mod generali, huwa diffiċli biex toħroġ b'numri eżatti, mhux l-anqas minħabba l-għadd ta' websajts qed jinbidel kontinwament. Izda hemm fatturi oħra li jaġhtu idea dwar l-ammont ta' Malti fuq l-internet meta mqabbel ma' lingwi oħra. L-ewwel harsa lejn l-ammont ta' dahliet fil-Wikipedia (fl-1 ta' Ġunju, 2011) wera li kien hemm madwar 2,820 dahla bil-Malti b'kuntrast ma' aktar minn 3,640,000 dahla bl-Ingliż u aktar minn 1,238,000 dahla bil-Ġermaniż. Meta wiehed iqabbel in-numru tad-Dominju tal-Ogħla Livell (TLD), it-TLD .mt jokkupa l-pożizzjoni 213 (minn 358) b'numru mhux speċifikat ta' dominji .mt rreġistrati (membru taċ-Ċentru Infor-

mazzjoni tan-Network ta' Malta ta stima ta' madwar 5,000), imqabbel ma' 21,336,063 dominju irreġistrati għal .com (kummerċjali, klassifikazzjoni 1) u 5,459,604 dominji għad- .de (il-Ġermanja, klassifikazzjoni 2). Naturalment, in-numru ta' dominji rreġistrati ma jgħid xejn dwar il-lingwa li l-paġni taht ċertu dominju huma miktubin biha.

Xi numri approssimattivi tal-ammont tal-lingwa Maltija fuq l-internet jistgħu jiġu kkalkulati permezz ta' procedura proposta minn [29] (L-awturi huma obbligati lejn Dr Albert Gatt (L-Istitut tal-Lingwistika, L-Università ta' Malta) talli ġibed l-attenzjoni tagħhom għal dan id-dokument.). L-idea bażika hija li kliem funzjonali (eż. izda, għal, dan eċċ.) huma aktar frekwenti minn kliem ta' kontenut (eż. nomi, verbi, aġġettivi) u jiffurma sett finit fil-lingwa. Barra minn hekk, il-persentaġġ ta' kliem funzjonali f'lingwa jkun stabbli f'kampjun ta' test meta d-daqs tal-kampjun jiżdied (il-Ligi Zipf). Għaldaqstant, wiehed jista' jikkalkula l-ammont ta' kliem għal kull lingwa fuq l-internet kif ġej:

L-ewwel nett, wiehed jikkalkula l-ammont ta' kliem funzjonali magħzula bil-Malti f'korpus (jiġifieri gabra ta' testi) li d-daqs ikun magħruf. It-tieni nett, wiehed juża magna ta' tiftix (eż. Google) biex isib il-frekwenza għall-istess kliem funzjonali fuq il-web. Fit-tielet pass, il-frekwenza min-numru tal-korpus tiġi estrapolata għall-Google Search u mbagħad medja tiġi kkalkolata għall-frekwenza tal-kliem funzjonali fir-rizultati ta' tiftix.

Xi restrizzjonijiet ta' dan il-metodu għandhom jissemmew: L-ewwel nett, in-numri miksuba b'dan il-metodu huma biss paġni mtellgħin. Perezempju, 94,300 paġna tal-Google għall-kelma *għal* mhumieq 94,300 każ tal-kelma fuq l-internet, izda 94,300 paġni fuq il-web li fihom il-kelma *għal* mill-anqas darba. It-tieni, it-tiftixija ssib biss paġni fuq il-web li għandhom URL individwali [29]. Paġni li huma aċċessibbli biss permezz ta' interfaċċ tal-web mhumieq miksuba bit-tiftix bl-internet. It-tielet, magna ta' tiftix tiftix biss għal sensiela ta' ittri

Kelma	f/m	Google (.mt biss, Reġjun=mt)	Estrapolazzjoni
għal	3730.96	94,300	25,274,996
qed	4770.79	118,000	24,733,849
minn	4833.58	173,000	35,791,276
kien	4073.83	93,800	23,025,015
biex	5276.78	179,000	33,922,202
dan	6412.28	434,000	67,682,634
kienet	1452.42	116,000	79,866,705
kienu	1465.56	135,000	92,114,959
kont	521.43	34,200	65,588,861
konna	301.39	19,400	64,368,426
jekk	2776.8	72,100	25,965,140
mhux	2101.32	79,500	37,833,362
Medja			48,013,952

1: Tfittxija bil-Google, ristretta għad-dominju .mt u r-reġjun ta' Malta

Kelma	f/m	Google (.mt biss)	Estrapolazzjoni
għal	3730.96	1,340,000	359,156,892
qed	4770.79	966,00	202,482,188
minn	4833.58	1,240,000	256,538,632
kien	4073.83	3,100,000	760,954,679
biex	5276.78	6,530,000	1,237,497,110
dan	6412.28	3,980,000	620,684,062
kienet	1452.42	665,000	457,856,543
kienu	1465.56	436,000	297,497,202
kont	521.43	450,000	863,011,334
konna	301.39	81,600	270,745,546
jekk	2776.8	1,120,000	403,341,976
mhux	2101.32	1,040,000	494,926,998
Medja			518,724,430

2: Tfittxija bil-Google, ristretta għad-dominju .mt biss

irrispettivament mill-ambjent tagħha fuq il-paġna tal-web. Din ma tagħmel l-ebda ġudizzju dwar jekk ċerta sensiela ta' ittri hijiex tabilhaqq kelma ta' lingwa.

Il-metodu deskritt aktar kmieni, applikat għall-kliem funzjonali Maltin, jiġġenera stimi differenti għall-Malti. Għas-websajts mad-dominju .mt li jinsabu f'Malta, id-daqs stmat huwa ta' 50 miljun kelma, filwaqt li għas-websajts bid-dominju .mt fir-reġjuni kollha id-daqs huwa 500 miljun kelma. Ir-raġuni għal din id-differenza hija li ħafna dominji .mt huma riżervati għal servers barra minn Malta.

Ir-riżultati eżatti tat-tfittxijiet fil-Google (imwettqa fit-8 ta' Lulju, 2011) u l-estrapolazzjoni tagħhom tista' tiġi ttraċċata lura fil-Figura 1 u l-Figura 2 hawn taht. Il-kolonna f/m (jiġifieri frekwenza għal kull miljun) tidentifika kemm il-kelma rispettiva ssehh ta' spiss f' miljun, fil-korpus MLRS. Pereżempju, fil-Figura 1, il-kelma *għal* 'for' tidher kważi 3731 darba f' miljun kelma. It-tiftix tal-Google għall-kelma *għal* tirriżulta f'94,300 paġna b'mill-anqas okkażjoni waħda ta' *għal* fuq paġna web taht id-dominju .mt f'Malta. Multiplikazzjoni ta' miljun u diviżjoni b'3730.96 jagħti ammont stmat ta' 25,274,996 każ ta' kull kelma Maltija fuq il-paġni taht id-dominju .mt ġewwa Malta. Jekk wiehed jagħmel dan il-kalkolu għall-kliem l-ieħor fil-figura u jsib il-medja tar-riżultati, wiehed jasal għal numru fit anqas minn 50 miljun kelma. Għall-paġni web madwar id-dinja kollha elenkati taht id-dominju .mt, ir-riżultati huma għaxar darbiet oghla.

Naturalment, għal studju serju, din it-tfittxija u l-estrapolazzjoni jkollhom jinkludu aktar kliem biex jaslu għal numri aktar affidabbli għall-ammont ta' Malti fuq l-internet. Iżda meta wiehed iqabbel ir-riżultati mat-Tabella 3 f' [29], wiehed jista' jgħid li ż-żewġ numri huma baxxi ħafna: għall-paġni web f'Malta biss, l-ammont huwa aktar mil-Latvjan u anqas mill-Iżlandiż għaxar snin ilu (in-numri fit-Tabella 3 ġew ikkalkulati f'Marzu 2001). Għall-paġni web dinjin, l-ammont tal-

Malti huwa aktar mill-Ungeriz u anqas miċ-Ċek għaxar snin ilu. Minħabba li "l-proporzjon ta' testi mhux bl-Ingliż għall-Ingliż qed jikber" [29], il-Malti jista' jkun saħansitra rappreżentat anqas fuq l-internet illum mil-lingwi li għadhom kemm issemmeu.

Apparti minn paġni ewlenin privati u weblogs, hemm numru ta' websajts uffiċjali bil-Malti. L-ewwel nett, hemm il-paġna ewlenija tal-gvern Malti [30], li hija disponibbli kemm bil-Malti kif ukoll bl-Ingliż. Barra minn hekk, hemm l-edizzjonijiet tal-internet tal-gazzetti ta' kuljum u ta' kull ġimgħa bil-lingwa Maltija: *In-Nazzjon*, *L-Orizzont* (kuljum), *Illum*, *Il-ĠENSillum*, *KullHadd*, *Lehen is-Sewwa*, *It-Torċa* (kull ġimgħa).

Is-websajts tat-TV Malti u l-istazzjonijiet tar-radju juru taħlita tal-Ingliż u l-Malti fi gradi differenti. Pereżempju, is-websajts tal-istazzjonijiet tat-televiżjoni NET TV [31] u One TV [32] għandhom qafas bl-Ingliż, flimkien ma' xi artikli bil-Malti, minkejja li l-programm tagħhom jinkludi titli kemm bil-Malti kif ukoll bl-Ingliż. L-istazzjon tar-radju tal-knisja RTK [33] (Malti u Ingliż) jippermetti lill-utent jagħzel bejn iż-żewġ lingwi. Is-sitweb tal-Public Broadcasting Services (PBS) [34] fih sezzjonijiet bl-Ingliż u taqsimiet bil-Malti kif għandu s-sitweb tar-Radju 101 [35]. Din it-taħlita bejn Ingliż u Malti tirrifletti l-użu tal-lingwa fil-ħajja ta' kuljum. Madankollu, fil-programmi, is-sitwazzjoni hija aktar ċara, minħabba li l-*Awtorità tax-Xandir ta' Malta* ħarġet linji gwida stretti għall-użu tal-Malti fuq it-TV u r-radju. Skont dawn, il-preżentaturi għandhom jikkellmu bil-Malti jew bl-Ingliż u mhux jaqilbu bejn iż-żewġ lingwi [3]. Għaldaqstant il-programmi tal-istazzjonijiet jinkludu xandiriet bil-Malti biss u oħrajn bl-Ingliż biss. Dawn ikunu wkoll spiss disponibbli fuq l-internet, jew bħala live stream inkella podcasts.

Barra minn Malta, kollezzjoni kbira ta' testi bil-Malti tinsab fi ħdan il-EUR-Lex [36] li tospita l-liġi uffiċjali u dokumenti oħra tal-Unjoni Ewropea mill-1951 fit-23

lingwa uffiċjali tagħha. Hafna jekk mhux id-dokumenti kollha tal-web disponibbli b'mod miftuħ jintużaw fi proġetti tal-korpus, eż. il-*JRC-Acquis Multilingual Parallel Corpus* [37], li huwa korpus parallel li fih testi sħaħ

tal-Liġi tal-Unjoni Ewropea fi 22 lingwa. Korpus ieħor li fih numru dejjem jikber ta' dokumenti tal-web vizibbli bil-Malti huwa l-korpus tal-MLRS (Server għar-Riżorsi Lingwistiċi bil-Malti) [38].

# APPOĠĠ TA' TEKNOLOĠIJA LINGWISTIKA GĦALL-MALTI

Teknoloġiji lingwistiċi huma teknoloġiji ta' informazzjoni li huma speċjalizzati biex jitrattaw il-lingwa umana. Għalhekk dawn it-teknoloġiji huma wkoll ta' spiss ikorporati taħt it-terminu “Teknoloġija Lingwistika Umana”. Il-lingwa umana sseħh fil-forma mitkellma u miktuba. Filwaqt li t-taħdit huwa l-eqdem u l-aktar mod naturali ta' komunikazzjoni lingwistika, l-informazzjoni kumplessa u l-biċċa l-kbira tal-għarfien uman jinżammu u jiġu trażmessi f'testi miktubin. Teknoloġiji ta' taħdit u testi jipproċessaw jew jipproduċu dawn iż-żewġ modi ta' lingwa bl-użu ta' dizżjunarji, regoli grammatikali u semantika. Dan ifisser li t-teknoloġija lingwistika (LT) jorbot il-lingwa mal-forom diversi tal-għarfien, indipendentement mill-midja (tat-taħdit inkella tat-testi) li fihom huma espressi. Figura 3 turi l-pajsaġġ grafiku tat-Teknoloġija Lingwistika.

Fil-komunikazzjoni tagħna aħna nhalltu l-lingwa ma' modi oħra ta' komunikazzjoni u mezzi oħra ta' informazzjoni – per eżempju, aħna norbtu t-taħdit ma' gesti u espressjonijiet tal-wiċċ. Testi diġitali jingħaqdu ma' stampi u hsejjes. Il-films jistgħu jinkludu lingwa f'forma mitkellma u miktuba. Għaldaqstant teknoloġiji ta' taħdit u testi jikkoinċidu u jinteraġixxu ma' hafna teknoloġiji oħra ta' komunikazzjoni multimodali u ta' multimidja. F'din it-taqsim, se niddiskutu l-oqsma ta' applikazzjoni tat-teknoloġija lingwistika, jiġifieri ċekk-jatur lingwistiku, tfittxija ta' web, interazzjoni tat-taħdit u traduzzjoni awtomatika. Dawn il-applikazzjonijiet u teknoloġiji baġiċi jinkludu

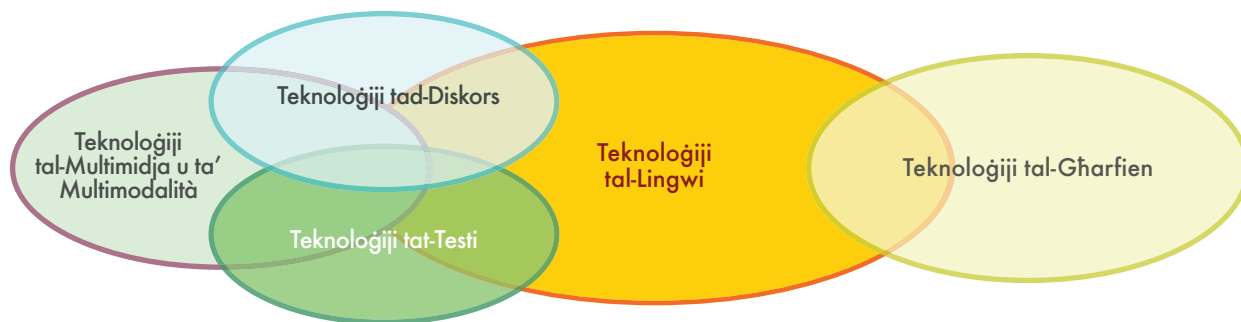
- korrezzjoni ortografika
- appoġġ għal min jikteb
- taġħlim tal-lingwi assistita mill-kompjuter
- irkupru ta' informazzjoni
- estrazzjoni ta' informazzjoni
- qosor ta' testi
- twegib ta' mistoqsijiet
- rikonossiment ta' taħdit
- sintezi ta' taħdit

It-teknoloġija lingwistika hija qasam stabbilit ta' riċerka li għalih hemm ammont estensiv ta' letteratura introdutturja. Il-qarrej interessat għandu jirreferi għar-referenzi li ġejjin: [39, 40, 41, 42, 43].

Qabel ma niddiskutu l-oqsma tal-applikazzjoni referruti hawn fuq, se niddeskrivu fil-qosor l-arkitettura ta' sistema LT tipika.

## 4.1 ARKITETTURI TA' APPLIKAZZJONIJET

Applikazzjonijiet ta' software tipiċi għall-ipproċessar tal-lingwa jikkonsistu f'diversi komponenti li jirriflettu aspetti differenti tal-lingwa u tal-kompitu li jkunu qed jimplimentaw. Ir-rappreżentazzjoni 4 turi arkitettura ferm simplifikata li tista' tinsab f'sistema ta' pproċessar ta' testi. L-ewwel tliet moduli jitrattaw l-istruttura u t-tifsira tat-test imdahħal fis-sistema:

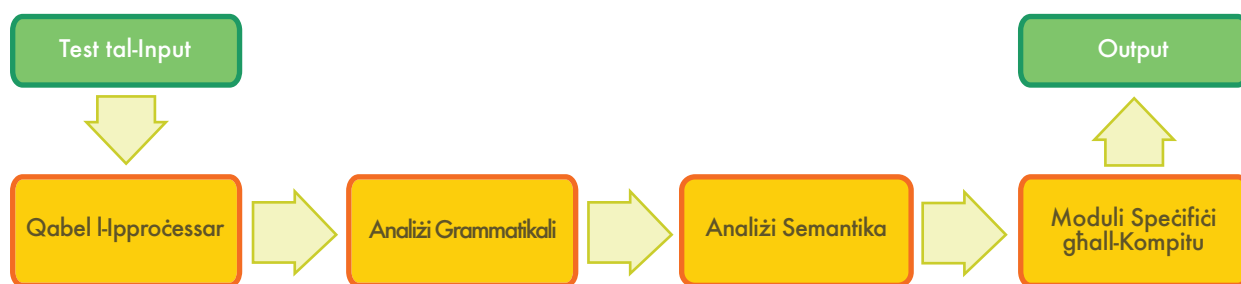


### 3: Teknoloġġija lingwistika fil-kuntest

1. Qabel l-ipproċessar: tindif tad-dejta, tnehhija tal-ifformattjar fejn hu xieraq, sejbien tal-lingwa mdaħhla fis-sistema, standardizzazzjoni tar-rappreżentazzjoni ta' simboli speċjali bhas-sing fil-Malti.
2. **Analizi grammatikali:** tiftix tal-verb u l-oġġetti tiegħu, il-modifikaturi, eċċ; sejbien tal-istruttura tas-sentenza.
3. **Analizi semantika:** tnehhja tal-ambigwià (Liema tifsira ta' "bank" hija t-tajba f'kuntest partikolari?), soluzzjoni ta' anafori u espressjonijiet ta' referenza bħal "hi", "il-karozza", eċċ; rappreżentazzjoni tat-tifsira tas-sentenza b'mod li tinqara minn magna.

Moduli ta' kompiti speċifiċi mbagħad iwettqu hafna operazzjonijiet differenti bħal sommarju awtomatiku ta' test imdaħhal fis-sistema, tfittxija ta' bażi ta' dejta u hafna oħrajn. Hawn taht, se nagħtu eżempji ta' oqsma ewlenin ta' applikazzjoni u niġbdu l-attenzjoni għal xi wħud mill-moduli ta' arkitetturi differenti f'kull sezzjoni. Għal darb'ohra, l-arkitetturi huma simplifikati ferm u idealizzati, sabiex iservu biex wiehed jispjega l-kumplessità tal-applikazzjonijiet tat-teknoloġija lingwistika b'mod ġenerali li jinftiehem.

Wara l-introduzzjoni tal-oqsma ewlenin ta' applikazzjoni, se nagħtu harsa ġenerali fil-qosor lejn is-sitwazzjoni fir-riċerka u l-edukazzjoni tat-TL, u nikkon-



4: Arkitettura tipika għall-ipproċessar ta' testi

kludu b'deskrizzjoni ta' programmi ta' ffinanzjar (tal-passat). Fl-aħħar ta' din it-taqsim, se nipprezentaw stima esperta dwar is-sitwazzjoni fir-rigward ta' għodod prinċipali u riżorsi tat-TL f'numru ta' dimensjonijiet bħal disponibbiltà, maturità, jew kwalità. Is-sitwazzjoni generali tat-TL għall-Malti hija mqassra fil-figura 10 (p. 37) fl-aħħar ta' dan il-kapitlu. Din it-tabella tinnota r-riżorsi kollha li huma b'tipa grassa fit-test. L-appoġġ tat-TL għall-Malti jiġi mqabbel ma' lingwi oħra li huma parti ta' din is-serje ta' *white papers*.

## 4.2 OQSMA EWLENIN TA' APPLIKAZZJONI

F'din it-taqsim, niffokaw fuq l-għodod u r-riżorsi l-aktar importanti u nipprovdu harsa generali dwar attivitajiet ta' TL f'Malta.

### 4.2.1 L-lċekkjar tal-lingwa

Kull min juża għodda ta' pprocessar tal-kliem bħal Microsoft Word iltaqa' ma' komponent li jiċċekkja l-ortografija, jindika żbalji ortografiċi u jipproponi korrezzjonijiet. 40 sena wara l-ewwel programm ta' korrezzjoni tal-ortografija minn Ralph Gorin, ċekkjaturi lingwistiċi llum ma jqabblux biss il-lista ta' kliem estratta ma' dizżjunarju ta' kliem spellut b'mod korrett, iżda saru dejjem aktar sofistikati. Minbarra l-algoritmi li jiddependu

fuq il-lingwa għall-immaniġġjar tal-morfoloġija (eż. formazzjoni tal-plural), xi wħud issa huma kapaci jagħrfu żbalji marbuta ma' sintassi, bħal verb nieqes jew verb li ma jaqbilx mas-sugġett tiegħu fil-persuna u n-numru, eż. f' *'She \*write a letter.'* Madankollu, iċ-ċekkjaturi l-aktar disponibbli (inkluż Microsoft Word) ma jsibu l-ebda żbalji fl-ewwel vers ta' poezija [44] li tidher hawn taht:

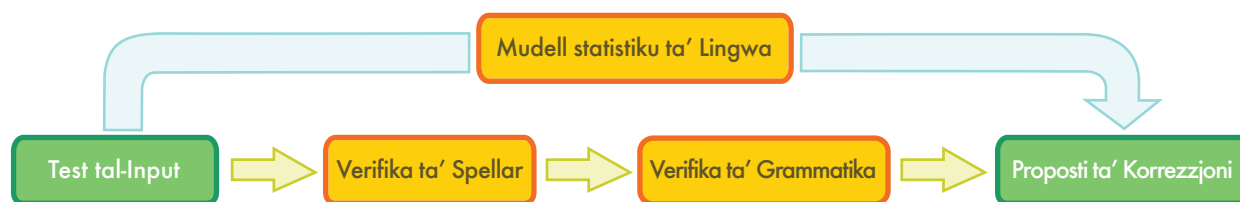
I have a spelling checker,  
It came with my PC.  
It plane lee marks four my revue  
Miss steaks aye can knot sea.

Għall-immaniġġjar ta' dan it-tip ta' żbalji, l-analiżi tal-kuntest huwa meħtieġ f'ħafna każijiet, eż. sabiex jiġi deċiż f'liema pożizzjoni l-*għ* siekta għandha tinkiteb f'verb Malti, bħal f'dan l-eżempju:

1. ... *in-negozjati li kien għamel il-Gvern ...*
2. *Pawlu, aghmel l-eżamijiet!*
3. \*... *in-negozjati li kien aghmel il-Gvern ...*

Iż-żewġ verbi *għamel* u *aghmel* jiġu ppronunzjati [e:.mɛl].

Dan jew jeħtieġ il-formulazzjoni ta' regoli tal-grammatika li jkunu speċifiċi, jiġifieri livell għoli ta' kompetenza u xogħol manwali, jew l-użu tal-hekk im-sejjaħ mudell lingwistiku tal-istatistika. Mudelli bħal



5: Verifika tal-Lingwa (fuq: ibbażat fuq statistika, isfel: ibbażat fuq regoli)



dawn jikkalkulaw il-probabbiltà ta' kelma partikolari li tidher f'ambjent speċifiku (jiġifieri, il-kliem ta' qabel u ta' wara). Pereżempju, kien għamel hija sekwenza ta' kliem hafna aktar probabbli milli kien aghmel. Mudell lingwistiku tal-statistika jista' jtnissel awtomatikament minn ammont kbir ta' dejta lingwistika (xierqa) (jiġifieri **korpus**).

Sa issa, dawn il-metodi ġew l-aktar żviluppanti u evalwati fuq dejta lingwistika bl-Ingliż. Madankollu, dawn mhux neċessarjament jittrasferixxu sew għal-lingwi li jkollhom inflessjoni għolja bħall-Malti, fejn tip ta' kelma partikolari, bħal verb, tista' tagħti numru kbir ta' forom ortografici.

Bħal lingwi oħra, mezz biex jiddetermina jekk sensiela partikolari hijiex kelma valida mhux kundizzjoni suffiċjenti għal sejbien ta' żbalji ortografici, iżda huwa kundizzjoni neċessarja. S'issa, għalkemm saru diversi tentattivi, l-ebda mezz bħal dan ma jeżisti għall-Malti.

Wieħed minn tal-ewwel kien ta' [45] li juża forom rudimentali ta' analiżi morfologiċi bbażati fuq regoli. Kelma kienet essenzjalment ikkunsidrata valida jekk tistax tkun derivata permezz ta' zokk misjub f'dizzjunarju. Il-problema b' dan il-metodu huwa li jeħtieġ lista kompluta ta' kull zokk, u naturalment, ir-regoli għandhom ikunu preċiżi hafna. Ir-riżultati kienu kemxejn limitati mil-lista ta' zokk, li ma kinitx kompluta, u n-natura imperfetta tar-regoli.

Metodu ieħor haress lejn l-istatistiċi għal soluzzjoni. L-idea intuittiva hija li għal lingwa partikolari, ċerti sekwenzi ta' karattri huma improbabbli hafna. Bl-Ingliż, pereżempju, qatt ma nsibu s-sekwenza *kk*, għalhekk jekk isseħħ taht l-istess sekwenza f'kelma miktuba, nistgħu nbassru, bi grad għoli ta' kunfidenza, li l-kelma mhijiex valida. B'mod aktar ġenerali, nistgħu nikkalkulaw il-probabbiltà ta' xi sekwenza bħala funzjoni tal-probabbiltajiet tas-sekwenzi ta' taħtha kollha, bl-adozzjoni tal-prinċipju li sabiex il-kelma tiġi kkunsidrata valida, il-probabbiltà trid taqbeż ċertu limitu.

Ċekkjatur ortografiku statistiku li jagħmel użu minn prinċipju bħal dan kien żviluppant minn [46]. Dan ma kienx jeħtieġ dizzjunarju, iżda minflok kien ibbażat fuq id-distribuzzjoni ta' n-grammi ta' karattri misjuba f'korpus ta' gazzetta. Deher ċar li biex dan il-metodu jirnexxi kien jeħtieġ (i) mudell lingwistiku aktar preċiż li jirrikjedi aktar dejta lingwistika minn dik disponibbli dak iż-żmien, u (ii) li l-probabbiltà tas-sekwenza waħedha ma kinitx biżżejjed sabiex kelma ortografika tiġi kklassifikata bħala żball. Kif issuggerit hawn fuq, informazzjoni oħra hija meħtieġa, bħal kategorija tal-kliem mill-kuntest tal-madwar.

Tentattivi oħra biex jiġi żviluppant ċekkjatur ortografiku għall-Malti jinkludu ċekkjatur fuq l-internet li ġie żviluppant minn Ramon Casha tal-Linux User Group [47]. Dan huwa bbażat fuq lista ta' kliem b'madwar miljun tip ta' kelma oriġinarjament miġbura minn korpus li jvarja, u sussegwentement estiż permezz ta' regoli differenti għall-immanigġjar ta' inflessjonijiet. L-eżattezza tiegħu ma gietx stabbilita ufficjalment. Microsoft ukoll kienu qed jaħdmu fuq ċekkjatur ortografiku biex jinkluduh mal-pakkett ta' interface tagħhom għal-lingwa Maltija għalkemm mhux magħruf meta dan se jiġi rilaxxat.

L-użu ta' ċekkjatur lingwistiku mhuwiex limitat għal għodod ta' pproċessar tal-kliem. L-iċċekkjar tal-lingwa jiġi wkoll applikat biex jiġu kkoreġuti awtomatikament mistoqsijiet mibgħuta lil magni ta' tiftix, eż. suggerimenti tat-tip "Ridit tfisser ..." lil Google.

Ir-riżultat taż-żieda mgħaġġla fid-domanda għall-prodotti tekniċi huwa li bosta kumpaniji bdew biex jiffokaw dejjem aktar fuq il-kwalità tad-dokumentazzjoni teknika quddiem il-ilmenti potenzjali tal-klijenti dwar l-użu hazin tal-lingwa u l-esiġi tal-ħsara li tirriżulta minn instruzzjonijiet ħziena jew li ġew mifhuma hazin. Software biex jappoggja lil min jikteb jista' jgħin lill-awtur tad-dokumentazzjoni teknika sabiex juża vokabularju u strutturi tas-sentenzi li huma konsistenti ma' ċerti regoli



espremiti formalment u ristrizzjonijiet (korporattivi) tat-terminologija.

Softwer biex jappoġġja lil min jikteb ma jeżistix bhalissa, iżda jista jkun hemm iskop konsiderevoli għall-użu ta' tali softwer fin-naħa tal-produzzjoni tal-Malti. Waħda mir-raġunijiet għall-iskarsezza ta' kontenut miktub bil-Malti, per eżempju fil-korrispondenza tan-negozju, hija li l-produzzjoni ta' kitba bil-Malti korretta hija diffiċli. Bosta kelliema nattivi kompetenti huma inklinati li jagħmlu zbalji meta jiġu għal-lingwa miktuba, u allura jippreferu jiktbu bl-Ingliż.

Id-disponibbiltà ta' għodod sempliċi u tajbin għall-appoġġjar tal-kitba jistgħu jtaffu din il-problema.

Barra minn CD ta' dizżjunarju interattiv bl-istampi [48], sal-lum l-ebda applikazzjoni bħal din ma giet żviluppata għall-Malti.

#### 4.2.2 Tiftix fuq il-web

Tiftix fuq il-web, f'intranets, jew libreriji diġitali huwa probabbilment l-aktar teknologija lingwistika uzata illum, iżda l-anqas żviluppata. Il-magna ta' tiftix Google, li bdiet fl-1998, illum hija uzata għal madwar 80% tat-tiftix kollu fid-dinja kollha [49]. Mill-2004, il-verb *google* huwa saħansitra msemmi fid-dizżjunarju *Cambridge Advanced Learner's Dictionary*. La l-interface tat-tiftix u lanqas il-preżentazzjoni tar-rizultati miksuba ma nbidlu b'mod sinifikanti mill-ewwel verżjoni. Fil-verżjoni kurrenti, Google tipproponi korrezzjoni tal-ortografija għall-kliem spellut hażin u, fl-2009, inkorporat kapacitajiet bażiċi ta' tiftix semantiku fit-taħlita algoritmika tagħha [50], li tista' ttejjeb il-preċiżjoni tat-tiftix billi tanalizza t-tifsira tat-termini mistoqsija f'kuntest. Is-suċċess tal-istorja tal-Google turi li b'ammont sostanzjali ta' dejta disponibbli u tekniki effiċjenti għall-indiċjar ta' din id-dejta, metodu bbażat il-biċċa l-kbira fuq statistika, jista' jwassal għal rizultati sodisfaċenti.

Madankollu, għal talba aktar sofistikata għall-informazzjoni, l-integrazzjoni ta' għarfien aktar profond tal-lingwistika hija essenzjali. Fil-laboratorji ta' riċerka, esperimenti bl-użu ta' **rizersi lessikali** bħal teżawri li jinqraw minn magni u rizersi ontoloġiċi tal-lingwa bħal WordNet urew titjib billi jippermettu s-sejbien ta' pagna fuq il-baži ta' sinonimi tat-termini ta' tiftix, eż. *enerġija atomika*, *enerġija nukleari* jew saħansitra termini relatati aktar mill-bogħod.

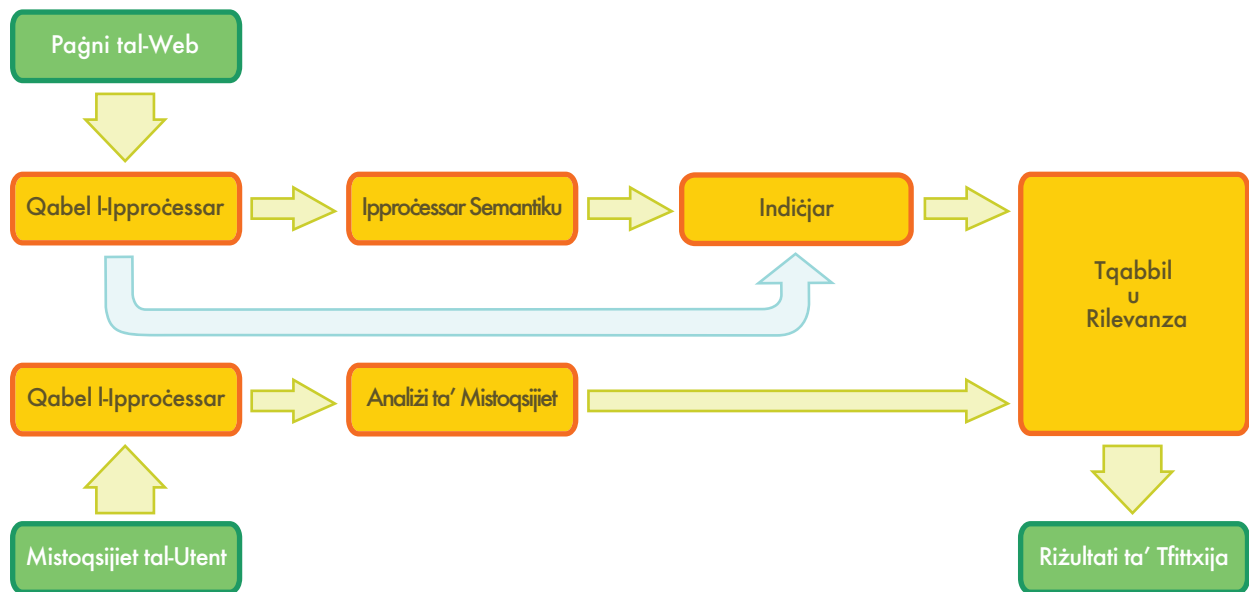
---

Il-ġenerazzjoni li jmiss tal-magni ta' tiftix trid tkun tinkludi iktar teknologija lingwistika sofistikata.

---

Il-ġenerazzjoni li jmiss ta' magni ta' tiftix se jkollhom jinkludu teknologija lingwistika ferm aktar sofistikata. Jekk mistoqsija ta' tiftix tikkonsisti f'domanda jew tip ieħor ta' sentenza minflok lista ta' kliem ewlieni, il-ksib ta' tweġibiet rilevanti għal din il-mistoqsija tehtieg **analizi semantiku** u sintattiku ta' din is-sentenza kif ukoll id-disponibbiltà ta' indiċi li jippermetti rkupru mgħaġġel tad-dokumenti rilevanti. Perezempju, immagina utent idahhal il-mistoqsija "Tini lista tal-kumpaniji kollha li ttiehdu minn kumpaniji oħra fl-aħħar hames snin". Għal tweġiba sodisfaċenti, parsing sintattika jehtieg li tiġi applikata biex tiġi analizzata l-istruttura grammatikali tas-sentenza u jiġi determinat li l-utent qed ifittex kumpaniji li ttiehdu minn kumpaniji oħra. Barra minn hekk, l-espressjoni *l-aħħar hames snin* jehtieg li tiġi pproċessata biex jinstab liema snin qed tirreferi għalihom.

Fl-aħħar nett, il-mistoqsija pproċessata tehtieg li tiġi mqabbla ma' ammont enormi ta' dejta mhux strutturata sabiex jinstabu l-biċċa jew biċċiet ta' informazzjoni li l-utent qed ifittex. Dan huwa ġeneralment imsejjaħ rkupru ta' informazzjoni (RI) u jinvolvi t-ftitxija għal u l-klassifikazzjoni ta' dokumenti rilevanti. Barra minn hekk, meta niġġeneraw lista ta' kumpaniji, jenhtieg li niġbru l-informazzjoni ta' sekwenza partikolari ta' kliem



6: Tiftix fuq il-web

f'dokument li tirreferi għal isem ta' kumpanija. Din it-tip ta' informazzjoni tkun disponibbli mill-hekk imsejja identifikazzjoni ta' entità bl-isem.

Sahansitra aktar diffiċli huwa t-tentattiv biex inqabblu mistoqsija ma' dokumenti miktuba f'lingwa differenti. Għall-irkupru ta' informazzjoni bejn il-lingwi, għandna nitraduċu awtomatikament il-mistoqsija fil-lingwi sorsi kollha possibbli u nittrasferixxu l-informazzjoni miksuba lura għal-lingwa fil-mira. Il-perċentwal dejjem jiżdied ta' dejta disponibbli f'formati mhux testwali imexxi d-domanda għal servizzi li jippermettu rkupru ta' informazzjoni multimidjali, jiġifieri, it-tiftix ta' informazzjoni li jinkludu immaġini, audio u dejta bil-videos. Għal files audio u video, dan jinvolvi modulu ta' **identifikazzjoni ta' taħdit** biex ibiddel il-kontenut tat-taħdit f'test jew rappreżentazzjoni fonetika, li magħhom jistgħu jitqabblu l-mistoqsijiet tal-utent.

F'Malta, hemm numru ta' websajts ta' tiftix li huma speċifikament immirati għal Malta [51]. Barra minn

hekk, hemm numru żgħir ta' SMEs ibbażati f'Malta li jinkorporaw tekniki relattivament sofistikati tal-ipproċessar tal-lingwa fl-ambitu ta' applikazzjonijiet ta' tiftix. Charonite [52], pereżempju, hija SME lokali li tittratta l-aħjar użu ta' magni ta' tiftix. Madankollu, bħalissa m'hemm l-ebda magni ta' tiftix kummerċjalment disponibbli li huma speċifikament immirati lejn l-ilsien Malti, apparti minn prototip għall-irkupru ta' informazzjoni bejn il-lingwi żviluppat għal skopijiet tal-LT4eL [53], proġett ta' riċerka Ewropew tal-FP6 li uża għodod ta' teknoloġija lingwistika multilingwi u tekniki ta' kodifikazzjoni ta' semantika għal titjib fil-ksib ta' materjal għat-tagħlim.

#### 4.2.3 L-Interazzjoni tat-taħdit

L-interazzjoni tat-taħdit hija l-bażi għall-ħolqien ta' interfaces li jippermettu lill-utent biex jinteraġixxi ma' magni billi juża l-lingwa mitkellma aktar milli, pereżempju, stampa grafika, tastiera, u mouse. Il-

lum, l-interfaces għall-vuċi (VUIs) huma normalment użati għal offerti ta' servizzi ta' awtomatizzazzjoni parzjali jew kompluti pprovduti mill-kumpaniji lill-klijenti tagħhom, l-impjegati jew l-imsieħba permezz tat-telefon. Dominji ta' negozju li jiddependu ħafna fuq VUIs huma l-banek, il-logistika, it-trasport pubbliku, u t-telekommunikazzjonijiet. Uzi oħra tat-Teknoloġija għall-Interazzjoni tat-Taħdit huma interfaces għal apparat partikolari, eż. sistemi ta' navigazzjoni fil-karozza, u l-użu tal-lingwa mitkellma bħala alternattiva għall-modalitajiet ta' input/output ta' interfaces grafiċi għall-utent, eż. fl-smartphones.

---

L-interazzjoni tat-taħdit hija l-bażi għall-ħolqien ta' interfaces li jippermettu lill-utent biex jinteraġixxi billi juża l-lingwa mitkellma aktar milli stampa grafika, tastiera, u mouse.

---

Fil-qalba tagħha, l-Interazzjoni tat-Taħdit tinkludi l-erba' teknoloġiji differenti li ġejjin:

1. **L-Identifikazzjoni awtomatika tat-taħdit (ASR)** hija responsabbli biex tiddetermina liema kliem kien attwalment mitkellem f'sekwenza partikolari ta' hsejjes imlissna minn utent.
2. L-analizi sintattika u l-interpretazzjoni semantika janalizzaw l-istruttura sintattika tat-tliissin tal-utent u jinterpretaw tal-aħħar skont l-għan tas-sistema rispettiva.
3. Il-ġestjoni tad-djalogu hija meħtieġa sabiex jiġi determinata ma' liema parti tas-sistema l-utent jinteraġixxi, liema azzjoni għandha tittiehed skont l-input tal-utent u l-funzjonalità tas-sistema.
4. It-teknoloġija tas-sintezi **tat-taħdit** (Text-to-Speech, TTS) hija użata biex tittrasforma l-kliem ta' dik l-espressjoni fi hsejjes li jkunu prodotti għall-utent.

Waħda mill-isfidi ewlenin hija li jkollok sistema ASR li tidentifika l-kliem imlissin minn utent bl-aktar mod

preċiż possibbli. Dan jeħtieġ jew restrizzjoni tal-firxa ta' espressjonijiet possibbli tal-utent għal sett limitat ta' kliem ewlieni, inkella l-ħolqien manwali ta' mudelli lingwistiċi li jkopru firxa kbira ta' espressjonijiet ta' lingwa naturali tal-utent. Bl-użu ta' tekniki tat-tagħlim tal-magni, mudelli tal-lingwa jistgħu wkoll jiġu ġġenerati awtomatikament minn **korpora tat-taħdit**, jiġifieri kollezzjonijiet kbar ta' files audio ta' diskors u transkrizzjonijiet. Ir-restrizzjoni ta' espressjonijiet normalment jirriżulta f'użu pjuttost riġidu ta' *voice user interface* (VUI) u jikkawża aċċettazzjoni baxxa mill-utenti; iżda l-ħolqien, l-irfinar u l-manutenzjoni tal-mudelli lingwistiċi jistgħu jżidu l-ispejjeż b'mod sinifikanti. VUIs li jużaw mudelli lingwistiċi u fil-bidu jippermettu lill-utent biex jesprimi l-intenzjoni tiegħu b'mod flessibbli – evokati, pereżempju, minn tislama bħal *Kif nista' nghinek?* – għandha aktar aċċettazzjoni mill-utenti.

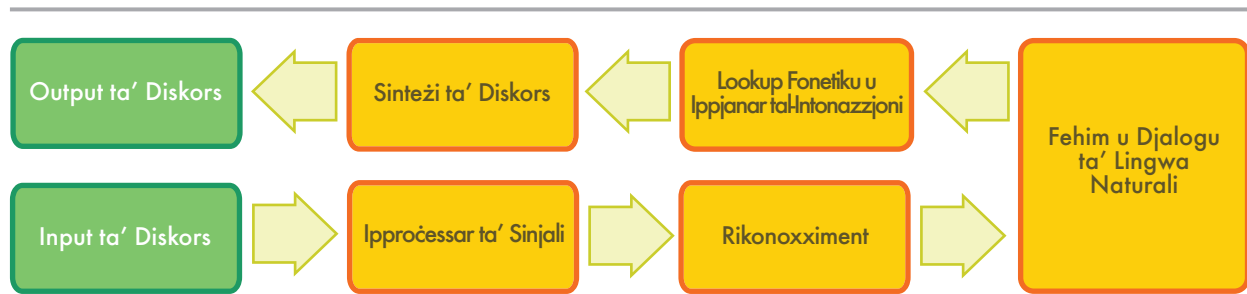
Il-kumpaniji għandhom tendenza li jużaw ħafna espressjonijiet irrekordjati minn qabel ta' kelliema professjonali – idealment korporattivi. Għal espressjonijiet statiči, li fihom il-kliem ma jiddependix fuq il-kuntesti partikolari tal-użu jew id-dejta personali ta' utent partikolari, dan iwassal għal esperjenza rikka tal-utent. Madankollu, aktar ma l-espressjoni jkollha tikkunsidra kontenut dinamiku, aktar l-esperjenza tal-utent tista' tbatu aktar minn prosodija fqira li tirriżulta minn files audio individwali marbutin ma' xulxin. Bl-ottimizzazzjoni, is-sistemi TTS tal-lum qegħdin jitjiebu fil-produzzjoni tan-naturalizza prosodika ta' espressjonijiet dinamiċi.

---

L-interazzjoni tat-taħdit hija l-bażi għall-ħolqien ta' interfaces li jippermettu lill-utenti sabiex jinteraġixxu fid-diskors minflok jużaw stampa grafika, tastiera, u mouse.

---

Rigward is-suq tat-teknoloġija għall-Interazzjoni tat-Taħdit, l-aħħar għaxar snin għaddew minn standard-



7: Sistema tal-interazzjoni tat-taħdit

izzazzjoni qawwija tal-interfaces bejn il-komponenti ta' teknoloġiji differenti, kif ukoll bi standards għal holqien ta' prodotti ta' software partikolari għal ċerta applikazzjoni. F'dawn l-aħħar għaxar snin kien hemm ukoll konsolidazzjoni soda tas-suq. Is-swieq nazzjonali fil-pajjiżi tal-G20 (jiġifieri pajjiżi ekonomikament b'saħħithom b'popolazzjoni konsiderevoli) huma dominati minn anqas minn 5 parteċipanti dinjija, b'Nuance (Istati Uniti) u Loquendo (Italja) dawk l-aktar prominenti fl-Ewropa. Fl-2011, Nuance ħabbret l-akkwist ta' Loquendo li jirriprezenta pass 'il quddiem fil-konsolidazzjoni tas-suq. Ħafna mill-iżvilupp tat-teknoloġija tat-taħdit f'Malta kkonċentra fuq 'mit-test għat-taħdit' (TTS). Xi xogħol pijunier fil-bidu kien imwettaq minn [54] u dan kien segwit minn numru ta' teżijiet f'livell ta' Masters [55]. Xi xogħol fuq sistema TTS ibbażata fuq il-web beda minn [56].

Żvilupp sinifikanti fis-sintezi tat-taħdit għall-Malti kien il-kisba ta' offerta mingħand il-gvern għall-iżvilupp ta' sintetizzatur tat-taħdit mill-kumpanija lokali Crimson Wing Ltd. Malta. Dan ix-xogħol huwa parzjalment if-finanzjat mill-Fond Ewropew għall-Iżvilupp Reġjonali u kkummissjonat mill-Fondazzjoni għall-Aċċessibilità għat-Teknoloġija tal-Infommazzjoni (FITA). Il-prototip se jkun konformi ma' SAPI u se jinkludi tliet vuċijiet (tal-irġiel, tan-nisa, u tat-tfal). Skont preżentazzjoni riċenti [57] ix-xogħol qed javvanza sew u prototip, mis-

teni fl-2012, se jkun disponibbli biex jitniżżel mingħajr ħlas.

Ix-xogħol fuq l-identifikazzjoni tat-taħdit huwa anqas avvanzat. Prototip biex jidentifika n-numri nholoq minn [58] f'dominji sempliċi. Fir-rigward tat-taħdit, il-problema fundamentali tibqa' nuqqas ta' dejta annotata b'mod xieraq minħabba li dan jehtieġ sforz manwali sinifikanti. Xi tentattivi ta' dħul awtomatiku saru minn [59]. Il-ħolqien ta' korpus u qafas deskrittiv għall-istudju tal-intonazzjoni Maltija bdiet mill-Istitut tal-Lingwistika u twestqet minn Vella u Farrugia [60]. Huwa mistenni li l-korpus li qed jiġi żviluppat minn Crimson Wing se jkun disponibbli għar-riċerka.

Ħarsa lil hinn mill-istat tat-teknoloġija tal-lum, turi li se jkun hemm bidliet sinifikanti minħabba l-firxa ta' smartphones bħala pjattaforma ġdida għall-ġestjoni ta' relazzjonijiet mal-klijenti – minbarra t-telefon, l-internet u mezzi għall-posta elettronika. Din it-tendenza se taffettwa wkoll l-użu tat-teknoloġija għall-Interazzjoni tat-Taħdit. Minn naħa, id-domanda għal telefonija bbażata fuq VUIs se tonqos, fuq medda ta' tul ta' żmien. Min-naħa l-oħra, l-użu tal-lingwa mitkellma bħala modalità ta' input faċli għall-utent għall-ismartphones se jikseb importanza sinifikanti. Din it-tendenza hija appoġġjata minn titjib osservabbli tal-eżattezza tal-identifikazzjoni tad-diskors li hija indipendenti mill-kelliem għal dettatura ta' taħdit li diġà hija offruta

bħala servizzi ċentralizzati għall-utenti tal-ismartphone. B'din l-esternalizzazzjoni tal-kompitu ta' identifikazzjoni tal-infrastruttura tal-applikazzjonijiet, l-użu ta' applikazzjoni speċifika ta' teknoloġiji ewlenin lingwistiċi għandha tikber fl-importanza meta mqabbla mas-sitwazzjoni preżenti.

#### 4.2.4 Traduzzjoni awtomatika

L-idea tal-użu ta' kompjuters diġitali għat-traduzzjoni ta' lingwi naturali bdiet fl-1946 minn A. D. Booth u kienet segwita minn fondi sostanzjali għar-riċerka f'dan il-qasam fil-hamsinijiet u ssoktat mill-ġdid fit-tmeninijiet. Madankollu, it-**Traduzzjoni Awtomatika** (TA) tibqa' tonqos li tissodisfa l-aspettattivi għolja li holqot fis-snin bikrin tagħha.

Fil-livell bażiku tagħha, TA sempliċiment tissostitwixxi l-kliem ta' lingwa naturali waħda b'ta' oħra. Dan jista' jkun utli f'dominji ta' suġġetti b'lingwa ristretta u konvenzjonali ħafna, pereżempju, rapporti tat-temp. Madankollu, għal traduzzjoni tajba ta' testi anqas standardizzati, unitajiet ta' testi akbar (frazzjiet, sentenzi, jew anki siltiet sħaħ) jeħtieġ li jkunu mqabbla mal-eqreb kontropartijiet tagħhom fil-lingwa fil-mira. Id-diffikultà prinċipali hawnhekk tinsab fil-fatt li l-lingwa umana hija ambigwa, u tohloq sfidi fuq livelli multipli, eż. tneħħija ta' ambigwiżità mis-sens tal-kelma fuq livell lessikali ('Jaguar' tista' tisser karozza jew animal) jew it-twaħħil ta' frazzjiet prepożizzjonali fuq livell sintattiku bħal fl-eżempju li ġejjin:

1. *Il-Kuntistabbli osserva lir-raġel bit-teleskopju.*
2. *Il-Kuntistabbli osserva lir-raġel bir-rivolver.*

Mod wieħed kif wieħed jitratta l-kompitu huwa bbażat fuq regoli lingwistiċi. Għat-traduzzjonijiet bejn lingwi marbutin flimkien mill-qrib, traduzzjoni diretta tista' tkun possibbli f'kazijiet bħall-eżempju t'hawn fuq. Izda sistemi bbażati fuq regoli (jew immexxija minn għarfien) ta' spiss janalizzaw it-test imdahħal fis-sistema

u joħolqu rappreżentazzjoni intermedjarja u simbolika, li minnhom it-test fil-lingwa fil-mira jiġi ġġenerat. Is-suċċess ta' dawn il-metodi jiddipendi ħafna fuq id-disponibbiltà ta' dizżjunarji estensivi b' informazzjoni morfologika, sintattika, u semantika, u settijiet kbar ta' regoli tal-**grammatika** mfasla bir-reqqa minn lingwista tas-sengħa.

---

Fil-livell bażiku tagħha, TA sempliċiment  
tissostitwixxi l-kliem ta'  
lingwa naturali waħda b'ta' oħra.

---

Lejn it-tmiem tat-tmeninijiet, kif l-enerġija kompjutazzjonali żdiedet u saret anqas għolja, kien hemm aktar interess fil-mudelli statistiċi għat-TA. Il-parametri ta' dawn il-mudelli statistiċi jittieħdu mill-analizi ta' **korpora ta' testijiet** bilingwi, bħalma hu l-**korpus parallel** tal-Europarl, li fih il-proċedimenti tal-Parlament Ewropew fi 21-il lingwa Ewropea. B'dejta suffiċjenti, TA statistika jaħdem tajjeb biżżejjed li jikseb tifsira approssimattiva għal test ta' lingwa barranija. Madankollu, bid-differenza ta' sistemi mmexxija mill-għarfien, it-TA statistika (jew immexxija minn dejta) spiss tiġġenera produzzjoni mhux grammatikali. Min-naħa l-oħra, minbarra l-vantaġġ li anqas sforz uman huwa meħtieġ għall-kitba grammatikali, TA mmexxija minn dejta tista' wkoll tkopri partikolaritajiet tal-lingwa li jintilfu f'sistemi mmexxija minn għarfien, pereżempju espressjonijiet idjomatiċi.

Minħabba li l-punti tajbin u dgħajfin tat-TA mmexxija minn għarfien u dejta huma kumplimentari, ir-riċerkaturi llum unanimament jimmiraw għal metodi ibridi li jgħaqqdu l-metodoloġiji tat-tnejn li huma. Dan jista' jsir b'modi diversi. Wieħed jinkludi l-użu ta' kemm sistemi mmexxija minn għarfien kif ukoll dejta u jkollu modulu ta' għażla li jiddeċiedi dwar l-aħjar output għal kull sentenza. Madankollu, għal sentenzi itwal, l-ebda riżultat mhu se jkun perfett. Soluzzjoni aħjar hija li



8: Traduzzjoni awtomatika (xellug: ibbażata fuq statistika; lemin: ibbażata fuq regoli)

tgħaqqad l-aħjar partijiet ta' kull sentenza minn outputs multipli, li jistgħu jkunu pjuttost kumplessi, minhabba li partijiet korrispondenti ta' alternattivi multipli mhux dejjem huma ovgji u jridu jkunu allinjati.

Il-kwalità tas-sistemi tat-TA għall-Malti hija meqjusa li għad għandha potenzjal kbir ta' titjib.

F'Malta l-hidma mwettqa fit-Traduzzjoni Awtomatika kienet ristretta għal ftit teżijiet tal-grad ta' Baċcellerat u Masters. Sistema ta' trasferiment ibbażata fuq l-LFG kienet żviluppata għall-Ingliż/Malti minn [61] u kienet tittraduċi b'suċċess it-tbassir tat-temp. Aktar tard J. Bajada [62, 63] hadem fuq TA statistika (SMT) bl-enfasi fuq tekniki għall-produzzjoni ta' mudelli lingwistiċi u tat-traduzzjoni. Ix-xogħol precedenti kien jikkonċerna mudelli bbażati fuq kliem, filwaqt li tal-aħħar żviluppa tekniki għal ġbir ta' dejta bilingwali ta' frażijiet minn korpus limitat.

Bħal f'bosta oqsma oħra, il-problema bażika hija n-nuqqas ta' kwantitajiet kbar ta' dejta bilingwali annotata b'mod xieraq. Minhabba din ir-raġuni, forsi, is-sistema bil-punt ta' referenza biex jiġu ġgudikati avvanzi tibqa' Google Translate.

Il-kwalità tas-sistemi tat-TA hija meqjusa li għad għandha potenzjal kbir ta' titjib. L-isfidi jinkludu

l-adattabilità tar-rizorsi lingwistiċi għal dominju ta' suġġett partikolari jew qasam tal-utent u l-integrazzjoni tax-xogħol għaddej eżistenti b'bażijiet ta' termini u memorji tat-traduzzjoni. Barra minn hekk, hafna mis-sistemi kurrenti huma bbażati fuq l-Ingliż u jappoġġjaw lil ftit biss mil-lingwi minn u għall-Ġermaniż, li jwassal għal tensjonijiet fix-xogħol għaddej totali tat-traduzzjoni, u eż. jisforza utenti tat-TA biex jitgħallmu għodod ta' kodiċi ta' lessiku differenti għal sistemi differenti.

Kampanji ta' evalwazzjoni jippermettu tqabbil tal-kwalità tas-sistemi tat-TA, il-metodi diversi u l-istatus ta' sistemi TA għal-lingwi differenti. Il-figura 9, ipprezentata fi hdan il-proġett Euromatrix+ tal-KE, turi l-prestazzjonijiet par par miksuba għal 22 lingwa uffiċjali (il-Gaelic Irlandiż huwa nieqes) f'termini ta' punteġġ BLEU [64]. Aktar ma jkun għoli l-punteġġ, aktar tkun tajba t-traduzzjoni. Traduttur uman igib madwar 80 [64].

L-aħjar riżultati (muriġa bl-aħdar u l-blu) nkisbu minn lingwi li jibbenefikaw minn sforzi konsiderevoli ta' riċerka, fi hdan programmi kkoordinati, u mill-eżistenza ta' **korpora paralleli** u numerużi (eż. Ingliż, Franciż, Olandiż, Spanjol), l-aġħar (bl-aħmar) minn lingwi li ma bbenefikawx minn sforzi simili, jew li huma differenti hafna minn lingwi oħra (eż. Ungeriz, Malti, Finlandiż).



Lingwa tal-mira — Target language																						
	EN	BG	DE	CS	DA	EL	ES	ET	FI	FR	HU	IT	LT	LV	MT	NL	PL	PT	RO	SK	SL	SV
EN	-	40.5	46.8	52.6	50.0	41.0	55.2	34.8	38.6	50.1	37.2	50.4	39.6	43.4	39.8	52.3	49.2	55.0	49.0	44.7	50.7	52.0
BG	61.3	-	38.7	39.4	39.6	34.5	46.9	25.5	26.7	42.4	22.0	43.5	29.3	29.1	25.9	44.9	35.1	45.9	36.8	34.1	34.1	39.9
DE	53.6	26.3	-	35.4	43.1	32.8	47.1	26.7	29.5	39.4	27.6	42.7	27.6	30.3	19.8	50.2	30.2	44.1	30.7	29.4	31.4	41.2
CS	58.4	32.0	42.6	-	43.6	34.6	48.9	30.7	30.5	41.6	27.4	44.3	34.5	35.8	26.3	46.5	39.2	45.7	36.5	43.6	41.3	42.9
DA	57.6	28.7	44.1	35.7	-	34.3	47.5	27.8	31.6	41.3	24.2	43.8	29.7	32.9	21.1	48.5	34.3	45.4	33.9	33.0	36.2	47.2
EL	59.5	32.4	43.1	37.7	44.5	-	54.0	26.5	29.0	48.3	23.7	49.6	29.0	32.6	23.8	48.9	34.2	52.5	37.2	33.1	36.3	43.3
ES	60.0	31.1	42.7	37.5	44.4	39.4	-	25.4	28.5	51.3	24.0	51.7	26.8	30.5	24.6	48.8	33.9	57.3	38.1	31.7	33.9	43.7
ET	52.0	24.6	37.3	35.2	37.8	28.2	40.4	-	37.7	33.4	30.9	37.0	35.0	36.9	20.5	41.3	32.0	37.8	28.0	30.6	32.9	37.3
FI	49.3	23.2	36.0	32.0	37.9	27.2	39.7	34.9	-	29.5	27.2	36.6	30.5	32.5	19.4	40.6	28.8	37.5	26.5	27.3	28.2	37.6
FR	64.0	34.5	45.1	39.5	47.4	42.8	60.9	26.7	30.0	-	25.5	56.1	28.3	31.9	25.3	51.6	35.7	61.0	43.8	33.1	35.6	45.8
HU	48.0	24.7	34.3	30.0	33.0	25.5	34.1	29.6	29.4	30.7	-	33.5	29.6	31.9	18.1	36.1	29.8	34.2	25.7	25.6	28.2	30.5
IT	61.0	32.1	44.3	38.9	45.8	40.6	26.9	25.0	29.7	52.7	24.2	-	29.4	32.6	24.6	50.5	35.2	56.5	39.3	32.5	34.7	44.3
LT	51.8	27.6	33.9	37.0	36.8	26.5	21.1	34.2	32.0	34.4	28.5	36.8	-	40.1	22.2	38.1	31.6	31.6	29.3	31.8	35.3	35.3
LV	54.0	29.1	35.0	37.8	38.5	29.7	8.0	34.2	32.4	35.6	29.3	38.9	38.4	-	23.3	41.5	34.4	39.6	31.0	33.3	37.1	38.0
MT	72.1	32.2	37.2	37.9	38.9	33.7	48.7	26.9	25.8	42.4	22.4	43.7	30.2	33.2	-	44.0	37.1	45.9	38.9	35.8	40.0	41.6
NL	56.9	29.3	46.9	37.0	45.4	35.3	49.7	27.5	29.8	43.4	25.3	44.5	28.6	31.7	22.0	-	32.0	47.7	33.0	30.1	34.6	43.6
PL	60.8	31.5	40.2	44.2	42.1	34.2	46.2	29.2	29.0	40.0	24.5	43.2	33.2	35.6	27.9	44.8	-	44.1	38.2	38.2	39.8	42.1
PT	60.7	31.4	42.9	38.4	42.8	40.2	60.7	26.4	29.2	53.2	23.8	52.8	28.0	31.5	24.8	49.3	34.5	-	39.4	32.1	34.4	43.9
RO	60.8	33.1	38.5	37.8	40.3	35.6	50.4	24.6	26.2	46.5	25.0	44.8	28.4	29.9	28.7	43.0	35.8	48.5	-	31.5	35.1	39.4
SK	60.8	32.6	39.4	48.1	41.0	33.3	46.2	29.8	28.4	39.4	27.4	41.8	33.8	36.7	28.5	44.4	39.0	43.3	35.3	-	42.6	41.8
SL	61.0	33.1	37.9	43.5	42.6	34.0	47.0	31.1	28.8	38.2	25.7	42.3	34.6	37.3	30.0	45.9	38.2	44.1	35.8	38.9	-	42.7
SV	58.5	26.9	41.0	35.6	46.6	33.3	46.6	27.4	30.9	38.9	22.7	42.0	28.2	31.0	23.7	45.6	32.2	44.2	32.7	31.3	33.5	-

9: Traduzzjoni awtomatika bejn 22 lingwa uffiċjali tal-UE - Machine translation between 22 EU-languages [65]

### 4.3 OQSMA OĦRA TAL-APPLIKAZZJONI

Il-bini ta' applikazzjonijiet tat-teknoloġija lingwistika jinvolvi firxa ta' kompiti sekondarji li mhux dejjem jidhru fuq livell ta' interazzjoni mal-utent, iżda jipprovdu funzjonalitajiet ta' servizz sinifikanti "wara l-kwinti" tas-sistema. Għaldaqstant, dawn jikkostitwixxu kwistjonijiet importanti ta' riċerka li saru dixxiplini sekondarji individwali tal-Lingwistika Kompjutazzjonali fl-akkademja.

---

Applikazzjonijiet tat-teknoloġija lingwistika spiss jipprovdu funzjonalitajiet ta' servizz sinifikanti "wara l-kwinti" tas-sistema ta' softwer ikbar.

---

It-tweġib ta' mistoqsijiet sar qasam attiv ta' riċerka, li għalih inbnew il-korpora annotati u bdew kompetizzjonijiet xjentifiċi. L-idea hija li wiehed jimxi minn tfittxija bbażata fuq kelma ewlenija (li għalih il-magna tirrispondi b'gabra shiha ta' dokumenti potenzjalment rilevanti) għal xenarju fejn l-utent jistaqsi mistoqsija konkreta u s-sistema tipprovdi tweġiba waħda:

*Mistoqsija: F' liema età Neil Armstrong għamel l-ewwel pass fuq il-qamar?*

*Tweġiba: 38.*

Filwaqt li dan huwa ovvjament relatat ma' Tiftix fuq il-web tal-qasam ewlieni msemmi qabel, it-tweġib tal-mistoqsijiet, illum huwa primarjament terminu ġenerali għal mistoqsijiet ta' riċerka bħal liema tipi ta' mistoqsijiet għandhom ikunu distinti u kif dawn għand-

hom ikunu trattati, kif sett ta' dokumenti li potenzjalment fih ir-risposta jista' jiġi analizzat u mqabbel (dawn jagħtu twegibiet konfliġġenti?), u kif tista' l-informazzjoni speċifika – it-twegiba – tittiehed b'mod affidabbli minn dokument, mingħajr ma jiġi injorat il-kuntest.

Dan huwa min-naħa l-oħra marbut mal-kompitu tal-estrazzjoni ta' informazzjoni (EI), qasam li kien ferm popolari u influwenti fil-mument tal-'bidla statistika' għal-Lingwistika Kompjutazzjonali, fil-bidu tad-disgħinijiet. EI timmira li tidentifika biċċiet speċifiċi ta' informazzjoni fi klassijiet speċifiċi ta' dokumenti; dan jista' jkun eż. s-sejbien tal-atturi ewlenin fit-teħid ta' kontroll ta' kumpaniji kif irrappurtat fi stejjer fil-gazzetti. Xenarju ieħor li nħadem fuqu huwa rapporti dwar incidenti terroristiċi, fejn il-problema hija li tqabbel it-test ma' mudell li jispeċifika min hu l-awtur, il-mira, il-hin u l-post, u r-riżultati tal-incident. Il-mili tal-mudell permezz ta' dominju speċifiku hija l-karakteristika ċentrali tal-EI, li għal din ir-raġuni huwa eżempju ieħor ta' teknoloġija “wara l-kwinti” li tikkositwixxi qasam ta' riċerka demarkata sew, iżda għal skopijiet prattiċi mbagħad jeħtieġ li tkun inkorporata f'ambjent ta' applikazzjoni xierqa.

Żewġ oqsma “borderline”, li xi drabi jkollhom ir-rwol ta' applikazzjoni waħedha u xi kultant dak ta' komponent appoġġjat, “taħt il-kappa” huma sommarju ta' testi u **ġenerazzjoni ta' testi**. Sommarju, ovyament, jirreferi għall-kompitu ta' taqsis ta' test twil, u jiġi offrut pereżempju bħala funzjonalità fi hdan l-MS Word. Dan jaħdem aktar fuq bażi ta' statistika, billi l-ewwel jidentifika l-kliem “importanti” f'test (jiġifieri, pereżempju, kliem li huma frekwenti ferm f'dan it-test iżda sostanzjalment anqas frekwenti fl-użu ġenerali tal-lingwa) u mbagħad jiddetermina dawk is-sentenzi li jinkludu ħafna kliem importanti. Dawn is-sentenzi mbagħad jiġu mmarkati fid-dokument, jew estratti minnu, u jiġu meħuda biex isir is-sommarju. F'dan ix-xenarju, li huwa

bil-bosta l-aktar wieħed popolari, sommarju huwa ugwali għal estrazzjoni ta' sentenzi: it-test jiġi trattat bħala sett sekondarju tas-sentenzi tiegħu. Is-sistemi ta' sommarji kummerċjali kollha jagħmlu użu minn din l-idea. Metodu alternattiv, li għalih hija ddedikata ċerta riċerka, huwa li sentenzi *godda* jiġu attwalment sintetizzati, jiġifieri, jinbena sommarju ta' sentenzi li m'hemmx għalfejn jidhru f'dik il-forma fit-test sors. Dan jeħtieġ ċertu ammont ta' fehim aktar profund tat-test u għaldaqstant huwa ħafna anqas b'saħħtu. Kollox ma' kollox, ġeneratur ta' test huwa f'ħafna każijiet mhux applikazzjoni li tista' toqgħod waħedha iżda huwa inkorporat f'ambjent ta' software akbar, bħal fis-sistema ta' informazzjoni kliniku fejn dejta tal-pazjent hija miġbura, maħżuna u pproċessata, u l-ġenerazzjoni ta' rapporti hija biss waħda minn ħafna funzjonalitajiet.

## 4.4 PROGRAMMI TAL-EDUKAZZJONI

It-teknoloġija lingwistika huwa qasam ferm interdixiplinarju, li jinvolvi l-kompetenza ta' lingwisti, xjenzjati tal-kompjuter, matematiċi, filosofi, psikolingwisti, u newroxjentisti, fost oħrajn.

F'Malta l-maġġoranza l-kbira ta' riċerka u edukazzjoni fit-TL twettqet fl-Università ta' Malta. Madankollu, din kienet stabbilita pjuttost tard. Raġuni waħda għal dan kienet id-dehra tard tax-Xjenza tal-Kompjuter bħala suġġett kurrikulari fl-Università. It-tmexxija politika turbulenti tal-pajjiż matul l-1970 u l-1980 ma pprevedietx ir-rivoluzzjoni fl-informatika li kellha sseħħ u kien biss fil-bidu tad-disgħinijiet li għet offruta għażla ta' kors universitarju permezz tal-Fakultà tax-Xjenza tal-Kompjuter mal-Matematika.

L-għeruw tal-istess bidla seħħew fl-1994, meta twettqet inizzjattiva strateġika nazzjonali li tirrikonoxxi u ssahħaħ ir-rwol tal-TI fis-setturi kummerċjali, politiċi, u fuq kollox, dawk edukattivi. Waħda mill-konsegwenzi



immedjati ta' dan kienet l-introduzzjoni ta' programm sostanzjali ta' erba' snin ta' Baċċelerat – il-BSc. IT (Hons) – fl-Università kif ukoll it-twaqqif ta' Dipartiment ġdid tax-Xjenza tal-Kompjuter u Intelligenza Artifiċjali (CSAI, ingħata isem mill-ġdid “Department of Intelligent Computer Systems (ICS)” fl-2009). Kors fl-NLP kien inkluz bħala għażla avvanzata, u dan wasal, erba' snin wara, għal serje ta' proġetti bħala parti mill-kors għall-ewwel grad fl-aħħar sena tiegħu li trattaw kwistjonijiet tal-ipproċessar tal-lingwa inkluzi metodi ta' kompjutazzjoni għall-Malti [66, 45, 67, 61, 46, 62, 68, 69, 70, 71]. Id-Dipartiment tal-Inġinerija tal-Komunikazzjonijiet u l-Kompjuters ha sehem ukoll fil-programm, u dan wasal għal sett ieħor ta' proġetti għall-ewwel grad fit-teknoloġija tat-taħdit.

Influenza oħra importanti fuq ir-riċerka hija L-Istitut tal-Lingwistika tal-Università (IOL), imwaqqaf fl-1988 bil-għan li jgħalliem kif ukoll jippromwovi u jikkordinar r-riċerka kemm fil-Lingwistika Ġenerali kif ukoll Applikata, imexxi 'l quddiem ir-riċerka li tinvolvi d-deskrizzjoni ta' lingwi partikolari, mhux l-anqas il-Malti, irawwem l-istudju ta' oqsma sekondarji diversi tal-lingwistika, u jippromwovi riċerka interdixiplinarja li tinvolvi l-akkademiċi f'kooperazzjoni Prattika li tgħaddi bejn konfini dipartimentali u fakultajiet barra mill-pajjiż. L-Istitut tal-Lingwistika jmexxi żewġ programmi għall-ewwel grad: B.A. fil-Lingwistika Ġenerali u l-B.Sc. l-ġdid fit-Teknoloġija Lingwistika Umana li se jkun offrut minn Ottubru 2011. Huwa wkoll possibbli li wieħed jagħmel Masters u Dottorat fil-Lingwistika mal-Istitut.

Fl-1997, grupp interdixiplinarju ta' xjenzjati tal-kompjuter u lingwisti (M. Rosner, R. Fabri, J. Caruana, M. Montebello u oħrajn) bdew jaħdmu fuq il-Maltilex, proġett biex jinholoq dizżjunarju kompjutazzjonali, li kien sostnut minn għotja zghira mill-Università appoġġjata mill-Mid-Med Bank. Interface sempliċi fuq l-internet kien żviluppat biex jippermetti l-ħolqien u z-

żamma ta' dahliet, kif irrappurtat f' [72] fl-ewwel Grupp ta' Hidma tal-ACL dwar l-Approċċi Kompjuttazzjonali għal Lingwi Semitiċi [72]. Eluf ta' dahliet bħal dawn saru b'mod manwali, iżda l-proġett iltaqa' ma' problemi legali, billi l-kompilazzjoni tad-dahliet kienet fil-biċċa l-kbira mnebbha mid-dizżjunarju f'forma ta' ktieb ta' Joseph Aquilina [73, 74].

L-isforz imbagħad gie trasferit minn dizżjunarji f'forma ta' ktieb għal teħid ta' dahliet lessikali minn sorsi oħra. Żewġ teżijiet [75, 68] użaw teknika bbażata fuq allinjament meħuda mill-bijoinformatika sabiex jiġbru flimkien dahliet lessikali u dan kien użat bħala mezz ta' strutturar tal-lessiku b'mod awtomatiku.

Minkejja n-nuqqas ta' finanzjament, l-isforz tal-Maltilex issokta b'mod kemxejn frammentat, appoġġjat mill-istaff tal-IOL u d-Dipartiment tas-CSAI. Ma kienx qabel l-2005 li l-Kunsill Malti għax-Xjenza u t-Teknoloġija (MCST) nieda l-ewwel Inizjattiva tar-Riċerka u l-Iżvilupp Teknoloġiku tal-pajjiż u proposta kongunta għas-Server għar-Riżorsi Lingwistiċi bil-Malti (MLRS) kienet aċċettata, sakemm ikun hemm appoġġ finanzjarju suffiċjenti biex jimpjegaw riċerkatur full time bejn l-2006 u l-2008. Il-proġett kellu żewġ miri kemm li johloq dizżjunarju kif ukoll korpus [76], u stabbilixxa l-pedamenti għas-server tal-MLRS preżenti.

Ir-riċerka msemmija hawn fuq tittratta prinċipalment mal-lingwa miktuba. Żewġ ferġat tax-xogħol relatat mat-taħdit ukoll qeġħdin jitwettqu.

L-ewwel wahda, mibdija minn tradizzjoni ta' pproċessar tas-sinjali fi hdan il-Fakultà tal-Inġinerija, ħolqot prototip ta' sintetizzatur għat-taħdit [54]. Ix-xogħol tiegħu influwenza diversi proġetti oħra mmirati biex itejbu s-sintezi tat-taħdit minn perspettiva baxxa ta' riżorsi inkluzi [58, 55, 77, 57].

It-tieni, tittratta l-kwistjoni tal-intonazzjoni [78] minn perspettiva lingwistika. Xi xogħol pijunier biex jinholoq korpus u qafas deskrittiv għall-istudju tal-intonazzjoni tal-Malti sar minn Vella u Farrugia [60].

Barra minn Malta, żewġ gruppi ta' riċerka li qegħdin f'kollaborazzjoni attiva ma' sforzi lokali mmirati lejn it-TL jistħoqqilhom aċċenn speċjali. Fl-Università ta' Arizona, grupp immexxi mil-lingwista Adam Ussishkin huwa partikolarment interessat fil-kwistjonijiet psikolingwistiċi li jappartjenu għal-lingwi semitiċi inkluż il-Malti. Sabiex jiġu studjati dawn il-kwistjonijiet sar disponibbli korpus online [79]. Fl-Università ta' Bremen, il-Professor Thomas Stolz kien involut b'mod attiv fl-istudju akkademiku tal-Malti iżda huwa partikolarment magħruf talli ospita l-ewwel konferenza dwar il-Lingwistika Maltija fi Bremen [80], waqqaf ġurnal [81] u l-Għaqda Internazzjonali tal-Lingwistika Maltija, ibbażata wkoll fi Bremen, li taħdem flimkien mal-Kunsill għall-Ilsien Malti ibbażat f'Malta.

Kif diġà ssemma, il-komunitajiet sensittivi għat-TL li hemm fl-Università ta' Malta jinsabu prinċipalment fi hdan il-Fakultà tal-ICT, l-Istitut tal-Lingwistika. Hemm ukoll interess potenzjali fil-Fakultà tal-Arti (Dipartiment tal-Malti) u suġġetti oħra Umanistiċi għalkemm sa issa hemm it-tendenza li l-lingwistika kompjutazzjonali titqies bħala suġġett eżotiku li jinsab f'fakultajiet aktar xjentifiċi bħax-xjenza tal-kompjuter jew l-istudji umanistiċi u, għalhekk, it-temi ta' riċerka li ġew trattati jikkoinċidu b'mod parzjali biss.

Haġa kurjuża, Malta mhijiex nieqsa minn avvenimenti internazzjonali relatati mat-TL. L-LREC 2010 saret fil-Belt Valletta, u ġibdet 1200 parteċipant. Il-konferenza annwali tal-EAMT sehhet ukoll f'Malta fl-1994, u saru wkoll numru ta' workshops iżgħar matul dawn l-aħhar 10 snin.

## 4.5 PROGRAMMI U SFORZI NAZZJONALI

Malta ssiehbet fl-UE fl-2004 u dan l-avveniment immedjatament ta lill-Malti l-istatus ta' lingwa uffiċjali tal-UE. Flimkien ma' dan l-istatus inħolqu obbligi

godda – b'mod partikolari biex jiġu tradotti kwantitajiet kbar ta' dokumenti uffiċjali, u barra minn hekk, ir-rikonoxximent, fuq livell Ewropew, li bħala lingwa nazzjonali, għandu jkollu status tal-“ewwel klassi” minn perspettiva teknoloġika kif ukoll soċjali, u jingħata d-drittijiet u privileġġi kollha li jgawdu l-“akbar” lingwi Ewropej (jiġifieri li għandhom numri akbar ta' kelliema nattivi).

L-Istrateġija Nazzjonali tat-TI 2008-10 tal-gvern inkludiet numru ta' għanijiet marbutin mal-Lingwa Maltija inkluż (i) l-iżvilupp tal-gvern fuq l-internet bil-Malti, (ii) il-holqien ta' għodod għal-lingwa Maltija, b'kollaborazzjoni mal-Università, u (iii) appoġġ għal komunitajiet fuq l-internet bil-Malti. Waqt li qed jinkiteb dan fl-2011, mhux l-għanijiet kollha ntlahqu. Madankollu l-effetti fit-tul ta' din l-istrateġija qed jibdedew jiehdu forma.

Bħalissa x-xena tat-teknoloġija lingwistika f'Malta tinsab taht l-influwenza ta' erba' inizjattivi prinċipali:

1. L-ewwel nett, proġett appoġġjat mill-gvern parzjalment iffinanzjat mill-fond ta' żvilupp reġjonali tal-UE qiegħed fil-proċess li jwassal li t-teknoloġija tat-taħdit tkun tista' tintlaħaq minn persuni b'dizabilità. Il-proġett bħalissa qed jiffoka fuq sintezi tat-taħdit bil-Malti, u f'dan il-punt il-mudelli tal-lingwa rilevanti qegħdin fil-proċess li jiġu żviluppjati. Il-konsorzju, li jikkonsisti f'SME (Crimson Wing Ltd), fondazzjoni (FITA, Fundazzjoni għall-Aċċess tat-TI), u l-Università, wiegħed li dawn ir-riżorsi se jkun disponibbli għal skopijiet ta' riċerka. Wieħed għad irid jara jekk il-komponenti tas-sintetizzatur tat-taħdit se jkun disponibbli għan-networks li jqassmu r-riżorsi ispirati minn CLARIN u META.
2. It-tieni nett, kif jirriżulta mir-rapport kurrenti, Malta qed tipparteċipa fil-METANET4U u għalhekk tirċievi finanzjament sinifikanti mill-UE mmirati lejn it-tishih u d-distribuzzjoni ta' riżorsi u għodod li huma speċifikament għall-Malti. L-

Università ta' Malta hija membru tal-META-NET u l-hsieb huwa li twestaq l-obbligi tagħha lejn l-għanijiet tal-META, partikolarment rigward l-identifikazzjoni tal-partijiet interessati, attwali u potenzjali.

3. It-tielet, is-Server għar-Riżorsi Lingwistiċi bil-Malti (MLRS) [82, 38] qed jagħti l-frott u sforzi sinifikanti għadhom għaddejnin fl-Università, permezz tal-Istitut tal-Lingwistika (A. Gatt, C. Borg, R. Fabri) u d-Dipartiment ta' Sistemi Intelligenti tal-Kompjuter (M. Rosner), li jsostnu u jiżviluppaw dan. Bhalissa l-MLRS huwa online fuq <http://mlrs.research.um.edu.mt>. Il-korpus jinkludi madwar 100M kelma, u s-sistema tinkludi xi servizzi bażiċi li jinkludu tiftix KWIC u stampi, tftixxija skont il-mudelli, diversi tipi ta' analiżi statistika eċċ. Bhalissa hemm aktar għodod ippjanati inkluż tagger għall-kategoriji tal-kliem u ċekkjur ortografiku.
4. Fl-aħħarnett, programm ġdid għall-ewwel grad fit-Teknoloġija tal-Lingwa Umana għandu jiġi mniedi mill-Istitut tal-Lingwistika f'Ottubru 2011. Dan se jkopri firxa sħiħa ta' suġġetti u inevitabilmente se jhalli impatt pożittiv fit-tul fuq l-istudju tal-Malti minn perspettiva kompjutazzjonali.

Minbarra dawn, proġett biex tiġi żviluppata verżjoni elettronika tad-dizzjunarju ta' Aquilina [73, 74] qed tittejjja bhalissa. Dan huwa sforz kollaborattiv bejn l-Università ta' Malta li qed tforni l-kompetenza lingwistika, l-Università ta' Arizona, li diġà ddiġitalizzaw id-dizzjunarju f'forma li tinqara minn kompjuter, u l-pubblikaturi Midsea Books Valletta. L-għanijiet doppji tal-proġett huma li jiġi aġġornat il-kontenut, u biex jagħtu lir-riċerkaturi l-flessibilità biex jaċċessaw it-test malajr. Għaddej sforz lokalment, sabiex jiġi organizzat livell tajjeb ta' kompetenza lessikografika meħtieġa għall-aġġornament tal-kontenut.

Għandna wkoll insemmu r-relazzjoni ta' Malta mal-CLARIN, proposta ta' infrastruttura ta' riċerka tal-UE

li tindirizza l-provvista ta' riżorsi lingwistiċi għax-Xjenzi Umanistiċi u Soċjali. Matul il-fażi ta' speċifikazzjoni, l-Università setgħet tipparteċipa bis-saħħa ta' għotja żgħira ta' appoġġ mill-Kunsill lokali għax-Xjenza u t-Teknoloġija. Madankollu, l-isfida biex jiġi żgurat il-finanzjament fit-tul meħtieġ għall-fażi ta' kostruzzjoni tal-CLARIN kienet akbar. L-identifikazzjoni ta' entità tal-gvern xierqa biex tiehu r-responsabbiltà għall-programm s'issa kienet mingħajr suċċess. Konsegwentement, il-partecipazzjoni futura ta' Malta fil-fażi ta' kostruzzjoni s'issa għadha mhix deċiża.

## 4.6 DISPONIBBILTÀ TA' GħODOD U RIŻORSI

Il-figura 10 tipprovdi ħarsa ġenerali lejn is-sitwazzjoni kurrenti tal-appoġġ tat-teknoloġija lingwistika għall-Malti. Il-klassifikazzjoni ta' teknoloġiji eżistenti u riżorsi hija bbażata fuq stimi studjati minn esperti ewlenin diversi skont seba' kriterji, kull waħda tvarja minn 0 (baxxa hafna) sa 6 (għolja hafna).

Għall-Malti, il-karatteristiċi l-aktar evidenti li hargħu mill-figura huma li

- l-biċċa l-kbira tad-dahliet huma vojta, u
- l-ogħla grad li ntlahaq huwa 3.2.

Il-fatt li d-dahliet huma kwazi kollha vojta jirrifletti l-istat immatur ta' riċerka u żvilupp marbut mat-TL f'Malta. Għalkemm hemm sinjali li s-sitwazzjoni qed tittejjeb, l-investment fit-teknoloġija lingwistika jibqa' fuq livell baxx, u bhala riżultat, minkejja l-kisbiet lokali modesti, l-isforz huwa frammentat, kemm f'termini ta' kopertura ta' oqsma differenti, kif ukoll f'termini ta' sostenibbiltà ta' riċerka: kien hemm wisq proġetti li jinvolvu qasam wieħed, riċerkatur wieħed biss, u sena jew sentejn biss. L-isforzi kollettivi ma jammontawx għal dak li huwa mixtieq.

	Kwantità	Disponibilità	Kwalità	Kopertura	Maturità	Sostenibilità	Adattabilità
<b>Teknoloġija Lingwistika (Għodod, Teknoloġiji, Applikazzjonijiet)</b>							
Identifikazzjoni ta' taħdit	0.8	0.8	0.8	0.8	0.8	0.8	0.8
Sintezi ta' taħdit	2.4	0.8	3.2	3.2	2.4	2.4	2.4
Analizi grammatikali	0.8	0.8	0.8	0.8	0.8	0.8	0.8
Analizi semantika	0	0	0	0	0	0	0
Ġenerazzjoni ta' testi	0	0	0	0	0	0	0
Traduzzjoni awtomatika	1.6	1.6	1.6	1.6	1.6	1.6	1.6
<b>Riżorsi Lingwistiċi (Riżorsi, Dejta, Bazijiet ta' Għarfien)</b>							
Korpora ta' testi	3.2	3.2	2.4	2.4	2.4	3.2	3.2
Korpora ta' taħdit	2.4	0.8	2.4	1.6	2.4	2.4	2.4
Korpora Paralleli	3.2	3.2	2.4	1.6	1.6	1.6	1.6
Riżorsi lessikali	2.4	2.4	1.6	2.4	2.4	2.4	2.4
Grammatiċi	0	0	0	0	0	0	0

10: Listat tal-appoġġ tat-teknoloġija tal-lingwa għall-Malti

Allura x'inkiseb? Nistgħu naraw billi nharsu lejn id-daħliet mhux vojta, li l-medja tal-punteġġ tagħhom jagħti l-ordni li ġeja:

- Ghodod:
  1. Sistema ta' tokens, Sintezi ta' Taħdit
  2. Identifikazzjoni ta' Taħdit
- Riżorsi:
  1. Korpora ta' Referenzi
  2. Korpora Paralleli
  3. Dizzjunarji, Terminoloġiji (huwa mifhum li dawn jinkludu listi ta' kliem)
  4. Mudelli lingwistiċi

Fir-rigward tal-ghodod:

Estrazzjoni ta' testi f'livell baxx u ghodod ta' pproċessar huma disponibbli, inkluż fornitur ta' tokens. POS-tagger qed jiġi żviluppat, izda l-prestazzjoni tiegħu mhi-jjex l-aktar stat modern, sakemm ikun hemm aktar taħriġ b'dejta annotata ahjar.

Għodod ta' livell ogħla (analizi sintattika jew semantika, ghodod ta' klassifikazzjoni, estrazzjoni ta' informazzjoni eċċ.) huma kompletament neqsin. Il-konsegwenza hija li, pereżempju, m'hemm l-ebda tree-banks disponibbli għall-Malti.

Prototip ta' ghodod ta' identifikazzjoni ta' taħdit ġew żviluppati fl-Università izda mhumiex faċilment disponibbli fiż-żmien meta qed jinkiteb dan. Madankollu, il-magna tat-taħdit iffinanzjata mill-gvern imsemmija qabel għandha tippovdi sintetizzatur tat-

taħdit jiffunzjona sal-2013. Filwaqt li dan huwa żvilupp pożittiv ħafna, huwa ffukat ħafna fuq in-naħa tas-sintezi tat-taħdit. Kważi l-ebda xogħol fuq l-**identifikazzjoni tat-taħdit** ma huwa pjanat f'dan l-istadju.

Fir-rigward ta' riżorsi, is-sitwazzjoni hija xi ftit aktar strutturata, minħabba li diġà jeżisti l-MLRS, infrastruttura kompjutazzjoni estensiva fil-forma ta' server li tiprovdni funzjonalità bażika li tippermetti aċċess fuq il-web għall-korpora disponibbli, xi servizzi, u sistema rudimentali li tiffacilita l-prezentazzjoni ta' kontribuzzjonijiet. L-MLRS bħalissa qed jipprovdi xi servizzi bażiċi ħafna għall-estrazzjoni, rappreżentazzjoni, tiftix u analiżi ta' testi.

Il-korpus tal-MLRS eżistenti bħalissa għandu kobor ta' madwar 100 miljun tokens. Dan huwa prinċipalment testwali u monolingwali. Huwa wkoll kemxejn mhux rappreżentattiv: hemm abbondanza ta' materjal legalistiku, iżda nuqqas ta' testi akkademiċi u xogħlijiet fittizji. F'dan l-istadju, dan il-materjal jista' biss jiġi mfittex u analizzat permezz tas-server u ma jistax jiġi aċċessat direttament. Ir-raġunijiet huma legalistiċi. B'aċċess ristrett b'dan il-mod, il-kumplikazzjonijiet tal-IPR u d-drittijiet tal-awtur ġew evitati bil-pulit. Il-prezz huwa li dawn il-kumplikazzjonijiet eventwalment se jkollhom jiġu kkonfrontati fil-futur, u fil-fatt META tinsab fil-proċess ta' formulazzjoni ta' sett ta' ftehim ta' liċenzjar li jgħodd għad-distribuzzjoni tar-riżorsi, bħall-MLRS.

## 4.7 TQABBIL TA' TRANS-LINGWI

L-istat attwali tat-teknoloġija lingwistika jvarja b'mod konsiderabbli minn komunità ta' lingwa waħda għal oħra. Sabiex titqabbell is-sitwazzjoni ta' bejn il-lingwi, din it-taqsimha se tippreżenta evalwazzjoni bbażata fuq żewġ oqsma ta' kampjuni tal-applikazzjoni (traduzzjoni bil-magni u pproċessar tad-diskors) u teknoloġija sottostanti (analiżi ta' testi), kif ukoll riżorsi

bażiċi meħtieġa biex jinbnew applikazzjonijiet għat-teknoloġija lingwistika. Il-lingwi kienu kategorizzati skont skala b'ħames punti:

1. Appoġġ eċċellenti
2. Appoġġ tajjeb
3. Appoġġ medju
4. Appoġġ parzjali
5. Appoġġ baxx għal kważi xejn

L-appoġġ tat-teknoloġija lingwistika kien imkejjel skond il-kriterji li ġejjin:

**Ipproċessar tad-Diskors:** il-kwalità tat-teknoloġiji eżistenti tal-identifikazzjoni tat-taħdit, il-kwalità tat-teknoloġiji eżistenti tas-sintezi tat-taħdit, il-kopertura ta' dominji, in-numru u d-daqs tal-korpora eżistenti tad-diskors, il-ammont u l-varjetà tal-applikazzjonijiet bbażati fuq it-taħdit li huma disponibbli.

**Traduzzjoni Awtomatika:** il-kwalità tat-teknoloġiji eżistenti tat-Traduzzjoni Awtomatika, in-numru tal-pari tal-lingwi koperti, il-kopertura ta' fenomeni u dominji lingwistiċi, il-kwalità u d-daqs tal-korpora paralleli eżistenti, il-ammont u l-varjetà tal-applikazzjonijiet tat-Traduzzjoni Awtomatika li huma disponibbli.

**Analiżi ta' Testi:** il-kwalità u l-kopertura tat-teknoloġiji eżistenti għall-analiżi ta' testi (morfoloġija, sintassi, semantika), il-kopertura ta' fenomeni u dominji lingwistiċi, il-ammont u l-varjetà tal-applikazzjonijiet li huma disponibbli, il-kwalità u d-daqs tal-korpora (anotati) eżistenti ta' testi, il-kwalità u l-kopertura tar-riżorsi lessikali (eż. WordNet) u tal-grammatiċi li jeżistu.

**Riżorsi:** il-kwalità u d-daqs tal-korpora ta' testi, ta' taħdit u korpora paralleli li jeżistu, il-kwalità u l-kopertura tar-riżorsi lessikali u tal-grammatiċi li jeżistu. Il-figuri 11 sa 14 juru li l-lingwa Maltija għandha biss appoġġ ta' Teknoloġija ta' Lingwi baxx għal medju u għalhekk titqabbell tajjeb ma' lingwi oħrajn li huma mitkellma anqas fl-Ewropa. Jidher b'mod ċar li r-riżorsi



u l-ghodda tat-teknoloġija lingwistika għall-Malti għadhom ma jilhqax il-kwalità u l-kopertura ta' riżorsi paragonabbli u tal-ghodda għal lingwi 'maġġuri' bħall-Ġermaniż, u ċertament mhux dik ta' dawg għal-lingwa Inġliża, li qiegħda fil-vantaġġ fi kważi l-oqsma kollha tat-teknoloġija lingwistika. U għad hemm hafna aktar vojt fir-riżorsi tal-lingwa Inġliża fir-rigward ta' applikazzjonijiet ta' kwalità għolja.

## 4.8 KONKLUŻJONIJIET

*F'din is-serje ta' white papers, għamilna sforz inizjali importanti sabiex nivvalutaw l-appoġġ tat-teknoloġija lingwistika għal 30 lingwa Ewropea, u biex nipprovdnu tqabbil ta' livell għoli madwar dawn il-lingwi. Billi jiġi identifikat dan il-vojt, il-ħtiġijiet u d-defiċits, il-komunità Ewropea tat-teknoloġija lingwistika u l-partijiet interessati relatati qeġħdin issa f' pużizzjoni biex ifasslu riċerka fuq skala kbira u programm ta' żvilupp immirat lejn il-bini ta' Ewropa tassew multilingwali u bbażata fuq it-teknoloġija.*

Rajna li hemm differenzi kbar bejn il-lingwi tal-Ewropa. Filwaqt li hemm softwer u riżorsi ta' kwalità tajba disponibbli għal xi lingwi u oqsma ta' applikazzjoni, oħrajn (is-soltu lingwi 'izgħar') għandhom vojt sostanzjali. Hafna lingwi għandhom nuqqas ta' teknoloġiji bażiċi u r-riżorsi essenzjali biex jiġu żviluppati dawn it-teknoloġiji. Oħrajn għandhom għodda u riżorsi bażiċi imma s'issa għadhom mhux kapaċi jinvestu fl-

ipproċessar semantiku. Għalhekk ahna għad neħtieġu li nagħmlu sforz fuq skala kbira sabiex niksbu l-għan ambizzjuż li tiġi pprovduta traduzzjoni bil-magni ta' kwalità għolja bejn il-lingwi Ewropej kollha.

F'dan ir-rapport, ahna ppruvajna nwasslu l-istat paradossali tat-Teknoloġija tal-Lingwa Maltija. Il-paradoss iquom għaliex hemm sforzi sinifikanti li saru minn numru żgħir ta' nies kwalifikati sew tul spettru ta' attivitajiet relatati mat-teknoloġija lingwistika biex jittejjeb l-istat tal-arti, kemm jekk dan ikun f'termini ta' għodda, jew riżorsi, jew it-tnejn. Huwa wkoll ċar li għol-kuntest aktar wiesgħa ta' attivitajiet edukattivi, kummerċjali u kulturali fil-pajjiż, hemm post għat-teknoloġija lingwistika biex tagħmel kontribuzzjoni importanti. Il-problema hi li l-isforzi li saru mhumiex koordinati, huma ta' żmien qasir, u frammentarji, għalhekk il-progress huwa aktar bil-mod milli għandu jkun.

Koordinazzjoni sostnuta u diretta tal-isforzi hija, fl-opinjoni tagħna, l-unika mod li fihom il-benefiċċji tat-teknoloġija lingwistika għall-Malti se tkun realizzata fi żmien raġonevoli. Ahna nemmnu li anke f'pajjiż żgħir daqs Malta, il-ħidma għandha bżonn tinqasam bejn partijiet interessati differenti. Irridu naslu għal pjan direzzjonali fattibbli permezz ta' verżjoni lokalizzata tad-diviżjoni tripartitika tax-xogħol sostnut minn META: identifikazzjoni ta' komunità b'vizjoni maqsuma; estensjoni ta' infrastruttura biex jiġi ffaċilitat it-tqassim tar-riżorsi, u t-tiħiħ ta' konnessjonijiet bejn t-teknoloġija lingwistika u l-oqsma ġirien tar-riċerka u l-iżvilupp.

Appoġġ eċċellenti	Appoġġ tajjeb	Appoġġ medju	Appoġġ parzjali	Appoġġ baxx/xejn
	Ingliz	Ġermaniż Taljan Finlandiż Franciż Olandiż Portugiż Spanjol Ċek	Bask Bulgaru Daniż Estonjan Galizjan Grieg Irlandiż Katalan Norvegiż Pollakk Svediż Serb Slovakk Sloven Ungeriz	Islandiż Kroat Latvjan Litwan <b>Malti</b> Rumen

11: L-proċessar tad-Diskors: listat tal-appoġġ għal 30 lingwa Ewropeja

Appoġġ eċċellenti	Appoġġ tajjeb	Appoġġ medju	Appoġġ parzjali	Appoġġ baxx/xejn
	Ingliz	Franciż Spanjol	Ġermaniż Taljan Katalan Olandiż Pollakk Rumen Ungeriz	Bask Bulgaru Daniż Estonjan Finlandiż Galizjan Grieg Irlandiż Islandiż Kroat Latvjan Litwan <b>Malti</b> Norvegiż Portugiż Svediż Serb Slovakk Sloven Ċek

12: Traduzzjoni bil-magni: listat tal-appoġġ għal 30 lingwa Ewropeja



Appoġġ eċċellenti	Appoġġ tajjeb	Appoġġ medju	Appoġġ parzjali	Appoġġ baxx/xejn
	Ingliz	Ġermaniż Franciż Taljan Olandiż Spanjol	Bask Bulgaru Daniż Finlandiż Galizjan Grieg Katalan Norvegiż Pollakk Portugiż Rumen Svediż Slovakk Sloven Ċek Ungeriz	Estonjan Irlandiż Islandiż Kroat Latvjan Litwan <b>Malti</b> Serb

13: Analizi ta' Testi: I-stat tal-appoġġ għal 30 lingwa Ewropeja

Appoġġ eċċellenti	Appoġġ tajjeb	Appoġġ medju	Appoġġ parzjali	Appoġġ baxx/xejn
	Ingliz	Ġermaniż Franciż Taljan Olandiż Pollakk Svediż Spanjol Ċek Ungeriz	Bask Bulgaru Daniż Estonjan Finlandiż Galizjan Grieg Katalan Kroat Norvegiż Portugiż Rumen Serb Slovakk Sloven	Irlandiż Islandiż Latvjan Litwan <b>Malti</b>

14: Rizorsi tal-lingwi u tat-testi: I-stat tal-appoġġ għal 30 lingwa Ewropeja

## DWAR META-NET

**META-NET** huwa Network ta' Eċċellenza ffinanzjat mill-Kummissjoni Ewropea. In-network bhalissa jikkonsisti f'54 membru minn 33 pajjiż Ewropew [1]. **META-NET** irawwem Alleanza Ewropea ta' Teknoloġija Multilingwi (**META**), komunità dejjem tikber ta' professjonisti u organizzazzjonijiet tat-teknoloġija lingwistika fl-Ewropa. **META-NET** irawwem is-sisien teknoloġiċi għall-istabbiliment u ż-żamma ta' soċjetà tal-informazzjoni Ewropea tasew multilingwi:

- jagħmel il-komunikazzjoni u l-kooperazzjoni bejn il-lingwi possibbli;
- jipprovdi aċċess ugwali għall-informazzjoni u l-għarfien fi kwalunkwe lingwa;
- joffri teknoloġija tal-informatika f'network avanzat u affordabbli għaċ-ċittadini Ewropej.

In-network jappoġġja Ewropa li tingħaqad bhala suq diġitali u spazju ta' informazzjoni uniku. Jistimula u jippromwovi teknoloġiji multilingwi għal-lingwi Ewropej kollha. It-teknoloġiji jippermettu traduzzjoni awtomatika, produzzjoni tal-kontenut, ipproċessar ta' informazzjoni u ġestjoni ta' għarfien għal varjetà wiesgħa ta' applikazzjonijiet u dominji ta' suġġetti. Huma jippermettu wkoll tkompli l-iżvilupp ta' interfaces intuitivi bbażati fuq il-lingwa għal elettronika tad-dar, makkinarju, vetturi, kompjuters u robots. Imniedi fl-1 ta' Frar 2010, **META-NET** għadu wettaq diversi attivitajiet fit-tliet linji ta' azzjoni tan-network **META-VISION**, **META-SHARE** u **META-RESEARCH**.

**META-VISION** trawwem komunità dinamika u influwenti ta' partijiet interessati, li tingħaqad madwar

vizjoni komuni u aġenda ta' ricerka strateġika komuni (**SRA**). L-enfasi ewlenija ta' din l-attività hija li tinbena komunità tat-TL koerenti u koeżiva fl-Ewropa billi jingabru flimkien rappreżentanti minn gruppi frammentati hafna u diversi ta' partijiet interessati. Il-White Paper preżenti kien ippreparat flimkien ma' volumi għal 29 lingwa oħra. Il-vizjoni teknoloġija maqsuma giet żviluppata fi tliet Gruppi ta' Vizjoni settorjali. Il-Kunsill tat-Teknoloġija ta' **META** gie stabbilit biex jiddiskutu u jhejju l-**SRA** bbażata fuq il-vizjoni b'interazzjoni mal-komunità **LT** kollha.

**META-SHARE** tohloq faċilità miftuħa u mqasma għal skambju u tqassim ta' riżorsi. In-network 'peer to peer' ta' repożitorji se jinvolvi dejta lingwistika, għodod u servizzi tal-web li huma dokumentati bi kwalità għolja ta' metadejta u organizzati f'kategoriji standardizzati. Ir-riżorsi jridu jkunu faċilment aċċessibbli u mfittxija b'mod uniformi. Ir-riżorsi disponibbli jinkludu materjal miftuħ u hieles ta' sorsi kif ukoll oġġetti ristretti, kummerċjalment disponibbli, ibbażati fuq tariffi.

**META-RESEARCH** tibni pontijiet għal oqsma marbuta mat-teknoloġija. Din l-attività tfittex li tmexxi l-avvanzi f'oqsma oħra u tikkapitalizza fuq ricerka innovattiva li tista' tibbenefika mit-teknoloġija lingwistika. B'mod partikolari, din l-attività tiffoka fuq t-twettiq ta' ricerka fit-traduzzjoni awtomatika, il-għbir tad-data, it-thejjija ta' settijiet ta' data u l-organizzazzjoni ta' riżorsi lingwistiċi għal skopijiet ta' evalwazzjoni; il-kompilazzjoni ta' inventarji ta' għodod u metodi; u l-organizzazzjoni ta' workshops u avvenimenti ta' taħriġ għall-membri tal-komunità.

**office@meta-net.eu** – <http://www.meta-net.eu>

## EXECUTIVE SUMMARY

During the last 60 years, Europe has become a distinct political and economic structure. Culturally and linguistically it is rich and diverse. However, from Portuguese to Polish and Italian to Icelandic, everyday communication between Europe's citizens, within business and among politicians is inevitably confronted with language barriers. The EU's institutions spend about a billion euros a year on maintaining their policy of multilingualism, i. e., translating texts and interpreting spoken communication. Does this have to be such a burden? Language technology and linguistic research can make a significant contribution to removing the linguistic borders. Combined with intelligent devices and applications, language technology will help Europeans talk and do business together even if they do not speak a common language.

---

Language technology builds bridges.

---

Language barriers can bring business to a halt, especially for SMEs who do not have the financial means to reverse the situation. The only (unthinkable) alternative to this kind of a multilingual Europe would be to allow a single language to take a dominant position, to replace all other languages. One way to overcome the language barrier is to learn foreign languages. Yet without technological support, mastering the 23 official languages of the member states of the European Union and some 60 other European languages is an insurmountable obstacle for Europe's citizens, economy, political debate, and scientific progress.

The solution is to build key enabling technologies: language technologies will offer European stakeholders tremendous advantages, not only within the common European market, but also in trade relations with non-European countries, especially emerging economies. Language technology solutions will eventually serve as a unique bridge between Europe's languages. An indispensable prerequisite for their development is first to carry out a systematic analysis of the linguistic particularities of all European languages, and the current state of language technology support for them.

The automated translation and speech processing tools currently available on the market fall short of the envisaged goals. The dominant actors in the field are primarily privately-owned for-profit enterprises based in Northern America. As early as the late 1970s, the EU realised the profound relevance of language technology as a driver of European unity, and began funding its first research projects, such as EUROTRA. At the same time, national projects were set up that generated valuable results, but never led to a concerted European effort. In contrast to these highly selective funding efforts, other multilingual societies such as India (22 official languages) and South Africa (11 official languages) have set up long-term national programmes for language research and technology development.

The predominant actors in LT today rely on imprecise statistical approaches that do not make use of deeper linguistic methods and knowledge. For example, sentences are often automatically translated by comparing each new sentence against thousands of sentences pre-

viously translated by humans. The quality of the output largely depends on the size and quality of the available data. While the automatic translation of simple sentences in languages with sufficient amounts of available textual data can achieve useful results, shallow statistical methods are doomed to fail in the case of languages with a much smaller body of sample data or in the case of sentences with complex, non-repetitive structures. Analysing the deeper structural properties of languages is the only way forward if we want to build applications that perform well across the entire range of European languages.

---

### Language technology as a key for the future.

---

The European Union is thus funding projects such as EuroMatrix and EuroMatrixPlus (since 2006) and iTranslate4 (since 2010), which carry out basic and applied research, and generate resources for establishing high quality language technology solutions for all European languages. European research in the area of language technology has already achieved a number of successes. For example, the translation services of the European Union now use the Moses open-source machine translation software, which has been mainly developed in European research projects.

In Malta, the most advanced areas in language technology are currently speech synthesis and text corpora: In the area of Maltese speech synthesis, a government-supported project partly funded by EU regional development funds is under way to bring speech technology within the reach of disabled persons. The consortium, which consists of an SME (Crimson Wing Ltd), a foundation (FITA, Foundation for IT Access), and the University, has pledged that these resources will be made available for research purposes.

In the area of text corpora, the Maltese Language Resource Server (MLRS) has come to fruition and signifi-

cant efforts are under way at University, through the Institute of Linguistics (A. Gatt, C. Borg, R. Fabri) and the Department of Intelligent Computer Systems (M. Rosner), to maintain and develop it. Currently, the corpus comprises some 100M words, and further tools are planned, including a part-of-speech tagger and a spell-checker.

---

### Language Technology helps to unify Europe.

---

Drawing on the insights gained so far, it appears that today's 'hybrid' language technology mixing deep processing with statistical methods will be able to bridge the gap between all European languages and beyond. As this series of white papers shows, there is a dramatic difference between Europe's member states in terms of both the maturity of the research and in the state of readiness with respect to language solutions. This white paper for the Maltese language demonstrates that there is potential for a language technology industry and research environment in Malta. But although a number of technologies and resources for Maltese exist, there are far fewer than for "larger" European languages and certainly not enough to support the full range of language-sensitive applications that are available for those other languages.

According to the assessment detailed in this report, the achievement of a breakthrough in Maltese language technology requires a whole cycle of changes involving content providers, developers and users of language technology. Some changes in national language policy must be implemented before any breakthroughs for the Maltese language can be achieved.

META-NET's vision is high-quality language technology for all languages that supports political and economic unity through cultural diversity. This technology will help tear down existing barriers and build bridges between Europe's languages. This requires all stakehold-

ers – in politics, research, business, and society – to unite their efforts for the future.

This white paper series complements other strategic actions taken by META-NET (see the appendix for an

overview). Up-to-date information such as the current version of the META-NET vision paper [4] or the Strategic Research Agenda (SRA) can be found on the META-NET web site: <http://www.meta-net.eu>.

## LANGUAGES AT RISK: A CHALLENGE FOR LANGUAGE TECHNOLOGY

We are witnesses to a digital revolution that is dramatically impacting communication and society. Recent developments in information and communication technology are sometimes compared to Gutenberg's invention of the printing press. What can this analogy tell us about the future of the European information society and our languages in particular?

---

The digital revolution is comparable to Gutenberg's invention of the printing press.

---

After Gutenberg's invention, real breakthroughs in communication were accomplished by efforts such as Luther's translation of the Bible into vernacular language. In subsequent centuries, cultural techniques have been developed to better handle language processing and knowledge exchange:

- the orthographic and grammatical standardisation of major languages enabled the rapid dissemination of new scientific and intellectual ideas;
- the development of official languages made it possible for citizens to communicate within certain (often political) boundaries;
- the teaching and translation of languages enabled exchanges across languages;
- the creation of editorial and bibliographic guidelines assured the quality of printed material;
- the creation of different media like newspapers, radio, television, books, and other formats satisfied different communication needs.

In the past twenty years, information technology has helped to automate and facilitate many processes:

- desktop publishing software has replaced typewriting and typesetting;
- Microsoft PowerPoint has replaced overhead projector transparencies;
- e-mail allows documents to be sent and received more quickly than using a fax machine;
- Skype offers cheap Internet phone calls and hosts virtual meetings;
- audio and video encoding formats make it easy to exchange multimedia content;
- web search engines provide keyword-based access;
- online services like Google Translate produce quick, approximate translations;
- social media platforms such as Facebook, Twitter and Google+ facilitate communication, collaboration, and information sharing.

Although these tools and applications are helpful, they are not yet capable of supporting a fully-sustainable, multilingual European society in which information and goods can flow freely.

## 2.1 LANGUAGE BORDERS HOLD BACK THE EUROPEAN INFORMATION SOCIETY

We cannot predict exactly what the future information society will look like. However, there is a strong likelihood that the revolution in communication technology is bringing together people who speak different languages in new ways. This is putting pressure both on individuals to learn new languages and especially on developers to create new technology applications to ensure mutual understanding and access to shareable knowledge.

In the global economic and information space, there is increasing interaction between different languages, speakers and content thanks to new types of media.

The current popularity of social media (Wikipedia, Facebook, Twitter, YouTube, and, recently, Google+) is only the tip of the iceberg.

---

The global economy and information space confronts us with different languages, speakers and content.

---

Today, we can transmit gigabytes of text around the world in a few seconds before we recognise that it is in a language that we do not understand. According to a recent report from the European Commission, 57% of Internet users in Europe purchase goods and services in non-native languages; English is the most common foreign language followed by French, German and Spanish. 55% of users read content in a foreign language while 35% use another language to write e-mails or post comments on the web [5].

A few years ago, English might have been the lingua franca of the web – the vast majority of content on the web was in English – but the situation has now drastically changed. The amount of online content in other

European (as well as Asian and Middle Eastern) languages has exploded.

Surprisingly, this ubiquitous digital linguistic divide has not gained much public attention; yet, it raises a very pressing question: Which European languages will thrive in the networked information and knowledge society, and which are doomed to disappear?

## 2.2 OUR LANGUAGES AT RISK

While the printing press helped step up the exchange of information in Europe, it also led to the extinction of many European languages. Regional and minority languages were rarely printed and languages such as Cornish and Dalmatian were limited to oral forms of transmission, which in turn restricted their scope of use. Will the Internet have the same impact on our modern languages?

---

The variety of languages in Europe is one of its richest and most important cultural assets.

---

Europe's approximately 80 languages are one of our richest and most important cultural assets, and a vital part of this unique social model [6]. While languages such as English and Spanish are likely to survive in the emerging digital marketplace, many European languages could become irrelevant in a networked society. This would weaken Europe's global standing, and run counter to the strategic goal of ensuring equal participation for every European citizen regardless of language. According to a UNESCO report on multilingualism, languages are an essential medium for the enjoyment of fundamental rights, such as political expression, education and participation in society [7].



## 2.3 LANGUAGE TECHNOLOGY IS A KEY ENABLING TECHNOLOGY

In the past, investments in language preservation focussed primarily on language education and translation. According to one estimate, the European market for translation, interpretation, software localisation and website globalisation was €8.4 billion in 2008 and is expected to grow by 10% per annum [8]. Yet this figure covers just a small proportion of current and future needs in communicating between languages. The most compelling solution for ensuring the breadth and depth of language usage in Europe tomorrow is to use appropriate technology, just as we use technology to solve our transport and energy needs among others.

Language technology targeting all forms of written text and spoken discourse can help people to collaborate, conduct business, share knowledge and participate in social and political debate regardless of language barriers and computer skills. It often operates invisibly inside complex software systems to help us already today to:

- find information with a search engine;
- check spelling and grammar in a word processor;
- view product recommendations in an online shop;
- follow the spoken directions of a navigation system;
- translate webpages via an online service.

Language technology consists of a number of core applications that enable processes within a larger application framework. The purpose of the META-NET language white papers is to focus on how ready these core enabling technologies are for each European language.

---

Europe needs robust and affordable language technology for all European languages.

---

To maintain our position in the frontline of global innovation, Europe will need language technology, tailored to all European languages, that is robust and affordable and can be tightly integrated within key software environments. Without language technology, we will not be able to achieve a really effective interactive, multimedia and multilingual user experience in the near future.

## 2.4 OPPORTUNITIES FOR LANGUAGE TECHNOLOGY

In the world of print, the technology breakthrough was the rapid duplication of an image of a text using a suitably powered printing press. Human beings had to do the hard work of looking up, assessing, translating, and summarising knowledge. We had to wait until Edison to record spoken language – and again his technology simply made analogue copies.

Language technology can now simplify and automate the processes of translation, content production, and knowledge management for all European languages. It can also empower intuitive speech-based interfaces for household electronics, machinery, vehicles, computers and robots. Real-world commercial and industrial applications are still in the early stages of development, yet R&D achievements are creating a genuine window of opportunity. For example, machine translation is already reasonably accurate in specific domains, and experimental applications provide multilingual information and knowledge management, as well as content production, in many European languages.

As with most technologies, the first language applications such as voice-based user interfaces and dialogue systems were developed for specialised domains, and often exhibit limited performance. However, there are huge market opportunities in the education and entertainment industries for integrating language technologies into games, edutainment packages, libraries, simu-

lation environments and training programmes. Mobile information services, computer-assisted language learning software, eLearning environments, self-assessment tools and plagiarism detection software are just some of the application areas in which language technology can play an important role. The popularity of social media applications like Twitter and Facebook suggest a need for sophisticated language technologies that can monitor posts, summarise discussions, suggest opinion trends, detect emotional responses, identify copyright infringements or track misuse.

---

Language technology helps overcome the “disability” of linguistic diversity.

---

Language technology represents a tremendous opportunity for the European Union. It can help to address the complex issue of multilingualism in Europe – the fact that different languages coexist naturally in European businesses, organisations and schools. However, citizens need to communicate across the language borders of the European Common Market, and language technology can help overcome this final barrier, while supporting the free and open use of individual languages. Looking even further ahead, innovative European multilingual language technology will provide a benchmark for our global partners when they begin to support their own multilingual communities. Language technology can be seen as a form of “assistive” technology that helps overcome the “disability” of linguistic diversity and makes language communities more accessible to each other. Finally, one active field of research is the use of language technology for rescue operations in disaster areas, where performance can be a matter of life and death: Future intelligent robots with cross-lingual language capabilities have the potential to save lives.

## 2.5 CHALLENGES FACING LANGUAGE TECHNOLOGY

Although language technology has made considerable progress in the last few years, the current pace of technological progress and product innovation is too slow. Widely-used technologies such as the spelling and grammar correctors in word processors are typically monolingual, and are only available for a handful of languages. Online machine translation services, although useful for quickly generating a reasonable approximation of a document’s contents, are fraught with difficulties when highly accurate and complete translations are required. Due to the complexity of human language, modelling our tongues in software and testing them in the real world is a long, costly business that requires sustained funding commitments. Europe must therefore maintain its pioneering role in facing the technological challenges of a multiple-language community by inventing new methods to accelerate development right across the map. These could include both computational advances and techniques such as crowdsourcing.

---

Technological progress needs to be accelerated.

---

## 2.6 LANGUAGE ACQUISITION IN HUMANS AND MACHINES

To illustrate how computers handle language and why it is difficult to program them to process different tongues, let’s look briefly at the way humans acquire first and second languages, and then see how language technology systems work.

Humans acquire language skills in two different ways. Babies acquire a language by listening to the real interactions between their parents, siblings and other family members. From the age of about two, children produce

their first words and short phrases. This is only possible because humans have a genetic disposition to imitate and then rationalise what they hear.

Learning a second language at an older age requires more cognitive effort, largely because the child is not immersed in a language community of native speakers. At school, foreign languages are usually acquired by learning grammatical structure, vocabulary and spelling using drills that describe linguistic knowledge in terms of abstract rules, tables and examples.

---

Humans acquire language skills in two different ways: learning from examples and learning the underlying language rules.

---

Moving now to language technology, the two main types of systems ‘acquire’ language capabilities in a similar manner. Statistical (or ‘data-driven’) approaches obtain linguistic knowledge from vast collections of concrete example texts. While it is sufficient to use text in a single language for training, e. g., a spell checker, parallel texts in two (or more) languages have to be available for training a machine translation system. The machine learning algorithm then “learns” patterns of how words, short phrases and complete sentences are translated.

This statistical approach usually requires millions of sentences to boost performance quality. This is one reason why search engine providers are eager to collect as much written material as possible. Spelling correction in word processors, and services such as Google Search and Google Translate, all rely on statistical approaches. The great advantage of statistics is that the machine learns quickly in a continuous series of training cycles, even though quality can vary randomly.

The second approach to language technology, and to machine translation in particular, is to build rule-based

systems. Experts in the fields of linguistics, computational linguistics and computer science first have to encode grammatical analyses (translation rules) and compile vocabulary lists (lexicons). This is very time consuming and labour intensive. Some of the leading rule-based machine translation systems have been under constant development for more than 20 years. The great advantage of rule-based systems is that the experts have more detailed control over the language processing. This makes it possible to systematically correct mistakes in the software and give detailed feedback to the user, especially when rule-based systems are used for language learning. However, due to the high cost of this work, rule-based language technology has so far only been developed for a few major languages.

---

The two main types of language technology systems acquire language in a similar manner.

---

As the strengths and weaknesses of statistical and rule-based systems tend to be complementary, current research focusses on hybrid approaches that combine the two methodologies. However, these approaches have so far been less successful in industrial applications than in the research lab.

As we have seen in this chapter, many applications widely used in today’s information society rely heavily on language technology. Due to its multilingual community, this is particularly true of Europe’s economic and information space. Although language technology has made considerable progress in the last few years, there is still huge potential in improving the quality of language technology systems. In the following, we will describe the role of Maltese in European information society and assess the current state of language technology for the Maltese language.

# THE MALTESE LANGUAGE IN THE EUROPEAN INFORMATION SOCIETY

## 3.1 GENERAL FACTS

Maltese is the national language of the Maltese archipelago, which consists of the islands Malta, Gozo (*Għawdex*) and Comino (*Kemmuna*).

Together with English, Maltese is also the official language of Malta. According to the *Demographic Review 2009* by the National Statistics Office of Malta, the estimated Maltese population (excluding foreigners) in Malta for the end of the year 2009 was 396,278. It is estimated that today, due to emigration phases from Malta mostly in the 1950s and 1960s, roughly the same number of expatriate native speakers lives abroad (mostly in the United Kingdom, Australia, USA and Canada).

Although Maltese belongs to the South Arabic branch of the Semitic language family, it differs considerably from the other neo-Arabic languages. Its structure is the result of different language contact situations that emerged under different rulers of the islands in the course of a millennium. While the core of Maltese is Semitic, it also contains a Romance superstrate and English adstrate. Also, Maltese is the only Semitic language written in a (modified) Latin alphabet.

The Semitic core of the Maltese language stems from the Arab conquest in 870 AD and its subsequent repopulation with Arabic speaking settlers. The first direct contact with Romance languages was established in 1090 when Malta was conquered by the Normans, who brought Sicilian with them, while the population still used their Arabic vernacular in everyday life. Malta

was more and more cut off politically, culturally and linguistically from the Arabic world. In the following centuries, under the influence of the Romance languages of the rulers, more and more Romance loan words entered the Arabic dialect. When Malta was under British rule in 1800, the official language changed from Italian to English, which brought an increasing number of English loan words into Maltese. The following sentence taken from a newspaper article (*l-Orizzont* of September 7th, 1995; reproduced in [9, p. 135]) can illustrate the different influences of the languages in contact (Romance loan words are in boldface, English loans underlined):

- (5) *Il-hold-up sar minn żagħżuġħ*  
the-hold-up happened from young.man  
*li kien **liebes nuċċali skur tax-xemx.***  
that was wearing glasses dark of.the-sun  
‘The robbery was committed by a young man who was wearing dark sunglasses.’

One remarkable fact about Maltese is that despite its relatively small number of speakers and the small area in which it is spoken, there is a comparatively rich number of variants or dialects. In general, a main distinction can be made between the Standard variety spoken in the urban areas like Valletta and Sliema and non-standard varieties spoken in the rural areas. Outside of Malta, the Maltese spoken in Australia has developed into an ethnolect of its own called *Maltraljan* [10]. It differs from Standard Maltese mainly in terms of its lexicon (i. e., the

vocabulary) that is the result of extensively borrowing words from (Australian) English and subsequent change in meaning.

With English being the second official language in Malta, many Maltese are bilingual. Between the poles of monolingualism and full bilingualism, there is a continuum of language-mixing and codeswitching. Most Maltese speak only Maltese at home and among each other. English, on the other hand, is the language used in the written context of higher education and in communication with foreigners.

## 3.2 PARTICULARITIES OF THE MALTESE LANGUAGE

Maltese is the only Semitic language in the European Union and the only Semitic language written in a Latin alphabet. The Maltese alphabet makes use of some special graphemes that differ from other Latin alphabets (the sound values are given in the International Phonetic Alphabet): ċ [tʃ], ġ [dʒ], ġħ (mostly silent), ħ [h], ż [z] [3, 11].

Some particular characteristics of Maltese are:

- free word order
- Semitic morphology
- aspect-based temporal system
- lack of a morphological infinitive

---

Word order is relatively free in Maltese sentences.

---

Even though there are no case endings, Maltese has a very free word order. The sentence *Il-kelb gidem il-qattusa lbierah* ('The dog bit the cat yesterday.') has the word order S(ubject) V(erb) O(bject) but could also be expressed as:

(6) *Ilbierah il-kelb gidem il-qattusa.*  
yesterday the-dog (m) he.bit the-cat (f)  
(SVO)

'Yesterday, the dog bit the cat.'

(7) *Gidem il-qattusa l-kelb ilbierah.*  
he.bit the-cat (f) the-dog (m) yesterday  
(VOS)

'The dog bit the cat yesterday.'

(8) *Il-qattusa gidimba l-kelb ilbierah.*  
the-cat (f) he.bit.her the-dog (m) yesterday  
(OVS)

'The cat, it was bitten by the dog yesterday.'

As the English translations try to show, the different word orders have a different emphasis in meaning. In the first two examples, the word orders are unmarked, with the object following the verb. In the last example, the object *il-qattusa* ('the cat') precedes the verb. As pointed out in [12, p. 140], this word order is marked and emphasises the object for contrast. With the object in front, native speakers prefer to mark *il-qattusa* with the object enclitic *-ha* on the verb. Also, in spoken discourse, this contrast is expressed with different intonation. The word order in the second example (VOS) could be used for expressing a contrastive meaning as well, given the appropriate intonation, putting emphasis on *gidem il-qattusa* 'he bit the cat'. Without this contrastive meaning (and without the contrastive intonation) emphasis would be on the fact itself as in: "Haven't you heard what happened yesterday? The dog bit the cat yesterday!" (Fabri, personal conversation).

---

Maltese words can change internally during inflection and derivation.

---

As a Semitic language, Maltese shows a non-concatenative morphology, i. e., inflected and derived word forms change internally:



In languages like English, word forms are made up of stems and affixes, i. e., concatenatively. The verb *shoot* can be inflected for third person by attaching the affix *-s* to the stem as in *(he) shoot-s*. Also, from the verbal stem a noun can be derived by adding the affix *-er* as in *shoot-er*. Hence both inflection and derivation take place without internal changes to the structure, i. e., concatenatively.

In Maltese, there is a mixture of stem-based and root-and-pattern-based morphology. In the Semitic component, the basic “unit” within a word is often not a stem but a root made up of three (sometimes four) consonants in a fixed order that carry a general meaning. Word stems with their specific meaning are formed by arranging the consonants according to a certain pattern. For example, the root *k-t-b* carries the meaning of everything connected with “writing”. In the following, patterns are represented as numbers 1,2,3 for the root consonants and v for the vowels between them, for example 1v2v3. By applying the pattern 1v2v3 and filling the vowel positions between the root consonants 1,2 and 3 with the vowel sequence i-e, one gets the perfective verb *kiteb* ‘he wrote’. Inflection of this verb for plural takes place by affixation of the plural affix *-u*, giving the form *kitbu* ‘they wrote’. Applying the pattern 1v22v:3 (v: stands for a long vowel) to the root renders the agent noun *kittieb* ‘writer’. Inflection of the noun by adding the affix *-a* gives the plural *kittieba* ‘writers’. Note that the plural suffix *-a* looks similar to the feminine marker *-a* so that *kittieba* could also refer to a female writer. The other Semitic Maltese plural suffixes are *-in* as in *mħallef* ‘judge’, *mħallfin* ‘judges’; *-at/-iet* as in *kittieba* ‘(female) writer’, *kittiebat* ‘(female) writers’; *-ijiet* as in *żmien* ‘time’, *żminijiet* ‘times’.

Plural nouns in Maltese can also be formed non-concatenatively (the so-called broken plural forms), i. e., no affixation takes place, but the noun is changed internally, e. g., *knieb* ‘book’ vs. *kotba* ‘books’.

Loan verbs today are mostly imported using a special verb class that can accommodate undigested stems [13]. For example, the English stem *park-* became the basis of the Maltese verb forms *pparkjajt*, *pparkjat*, *pparkja* ‘I/ she/ he parked’. Today, this formerly marginal Semitic special verb class has increased in size due to the influx of English loan verbs. It is highly productive, often giving way to *ad-hoc* loans of English verbs which already have a Semitic counterpart in Maltese. For example ‘to download (a file)’ can be expressed using the Semitic verb *nizzel* (originally meaning ‘he caused to come down’). Taking the English stem *download* and importing it via the special verb class instead gives forms like *ddawnlowdjajt*, *ddawnlowdjat*, *ddawnlowdja* ‘I/ she/ he down-loaded’. This strategy is often criticised as corrupting the language [3].

---

### The Maltese temporal system is marked for aspect.

---

Verbs in Maltese are marked for aspect, i. e., as to whether an action is completed (perfective) or not completed (imperfective) – for a full account on tense and aspect in Maltese, see [14, 15]. In the absence of any other grammatical markers, verbs in the perfective are interpreted as “past tense” and verbs in the imperfective as “present tense”: *Andrew kiteb* ‘Andrew wrote’; *Andrew jikteb* ‘Andrew writes’. Combination of the imperfective verb with *kien*, the perfective form of the verb for ‘to be’, expresses habitual past: *Andrew kien jikteb* ‘Andrew used to write’. Adding word *qed* ‘progressive’ (like the English *-ing* form) gives *Andrew kien qed jikteb* ‘Andrew was writing’ etc.

Maltese verbs do not have morphological infinitives. Thus, in complex predicates like in the English sentence ‘Andrew wants to write’, both verbs are morphologically finite: *Andrew irid jikteb* (literally: ‘Andrew he wants he writes’) even though semantically, *jikteb* is not finite.

### 3.3 RECENT DEVELOPMENTS

With the rise of English to the status of an international language and language of technology after the Second World War, the number of English loan words in Maltese has grown to a great extent. Many of them have become “nativised”, i. e., they are adopted in regular use so much that even derived Semitic words cannot replace them. For example, instead of the commonly used word *ajruport* (from English *airport*), the Semitic word *mitjar* once was proposed (derived from *tar* ‘he flew’). However, it became never accepted by the language community. On the other hand, loan words enter the language very rapidly, being imported spontaneously, even though there are already proper Maltese words for them (for example *ddawnlowdja* vs *nizzel* ‘he downloaded’). This fuels fears among some that the language might become “corrupted” [3].

Another recent development for Maltese is its status as an official language of the European Union. This has both advantages and disadvantages [3]. On the one hand, Maltese has finally become an internationally recognised language, a status that it did not have for a long time, being marginalised as a “kitchen language” centuries before. On the other hand, Maltese EU translators are confronted with certain challenges: many technical and legal terms have yet to be “invented” for Maltese. This results eventually in lexical expansion of the language (definitely a positive aspect), which, however, has to be coordinated by a central body so that individual translators do not come up with different terms for the same concepts independently from each other (which is a serious problem). The central body to deal with this challenge is the National Council for the Maltese Language (*Il-Kunsill Nazżjonali tal-Ilsien Malti*).

Other developments in recent years concern the Maltese orthography. Maltese (together with English) became the official language of Malta on January 1, 1934 – in

the orthography released by the Union of Maltese Writers (*Għaqda tal-Kittieba tal-Malti*) in 1924. Since then, the orthography has undergone three revisions (1984, 1992 and 2008).

The last reform was released in 2008. Its aim was to reduce writers’ insecurities that resulting from a considerable number of spelling variants for certain words. As the *Kunsill’s* document *Deciżjonijiet 1* [16] points out, the great number of variants could be reduced by finding a consistent balance between grammatical and phonetic spelling. Thus the four variants *zobtu*, *zoptu*, *sobtu* and *soptu* (‘suddenly, unexpectedly’) could be reduced to the two variants *zoptu* for [‘zɔp.tɔ] and *soptu* for [‘sɔp.tɔ]. For a similar reason, the word *skond* [skɔnt] ‘according to’, was changed to *skont* since its other grammatical forms do not justify spelling with *d* (derived from Italian *secondo*), as, e. g., *skontok* [‘skɔn.tɔk] ‘according to you’.

For the third area (loan words), the principle remains to write loan words according to the Maltese orthography if they are regarded as “nativised” and if it does not result in conflicts with the pronunciation or with other Maltese writing rules. However, many Maltese prefer to write English loan words with their original spelling, since they have become used to them. In fact, during a public seminar on the treatment of English loan words in April 2008, there were emotional discussions among the audience when it came to words like *email* and their proposed new spellings as *imejl*. Factors like the habits of a language community make the standardisation of spellings even more difficult than finding the balance between grammatical and phonetic principles [17].

These examples only give a slight idea of the hard work that the *National Council for the Maltese Language* is undertaking as part of language cultivation in Malta. The next section will give an insight into the history of language cultivation in Malta.



### 3.4 LANGUAGE CULTIVATION IN MALTA

Compared to other languages of Europe, the status of Maltese as an official language (since 1934) itself is a recent development. Thus language cultivation, too, had a late start.

For centuries, Maltese was only the spoken medium of the Maltese population and marginalised in comparison to the respective official language of Malta's rulers. This started to change with the language movement of the mid-/ late 18th century when first systematic studies of the language were conducted by Agius de Soldanis (1750) and Mikiel Anton Vassalli (1797). Especially Vassalli promoted the Maltese language by promoting its use in every domain of everyday life. Fortunato Panzavecchia's bible translations of the mid 19th century contributed to further standardisation of the language [18]. And with the move towards a standardised orthography in the early 20th century, an important step was made by the foundation of the Union of Maltese Writers (*Għaqda tal-Kittieba tal-Malti*) in 1920. The orthographic system, which was developed by this organisation, became Malta's official orthography in 1934 and, with some changes and additions, has been in use since.

In 1964, after gaining independence from Great Britain, the status of Maltese as national language and as official language together with English was written into the constitution. When Malta joined the EU in 2004, Maltese became an official language of the EU. As noted in the section above, this results in certain challenges, which can only be solved by a body that coordinates standardisation and common practice in translation work.

The body in Malta to do this work is the National Council for the Maltese Language (*Il-Kunsill Nazzjonali tal-Ilsien Malti*). It was founded in 2005 as the first government organisation to officially deal with language mat-

ters and language planning for the Maltese language. The Council's tasks are, as formulated in the Maltese Language Act (ACT No. V of 2004): promoting the Maltese language, to "adopt a suitable linguistic policy backed by a strategic plan" and put it into practice. Another important task of the *Kunsill* is to update the Maltese orthography and decide on correct spellings (taking over the task from the Academy of Maltese and thus being mainly responsible for the Maltese orthography reform of 2008). On its website, the Council also offers training courses for proof-readers and Maltese language courses for foreigners [19].

Before the Council was founded, standardisation of orthography was the task of the Academy of Maltese (*Akkademja tal-Malti*). It emerged in 1964 from the Union of Maltese Writers (*Għaqda tal-Kittieba tal-Malti*), which had been the founding body for the first official orthography in 1924/1932. The Academy's main aim today is to promote academic studies in the Maltese language and literature, to promote the use of Maltese in every domain of everyday life and to build up contacts to people who are friends of the language and who use it outside of Malta [20]. The Academy works closely together with the National Council for the Maltese Language.

The motivation behind the Maltese Language Act was the idea that one national language which is shared by all individuals within that nation forms the basis for cultural and national identity. This of course calls for standardisation of the language. Indeed, from the language cultivation movement of the 19th century until today, Maltese has risen from a formerly marginalised vernacular to a national language of high prestige. This is also reflected in the ever-growing amount of literary works in Maltese during the same timespan and in the high number of influential organisations and bodies for the Maltese language and literature [3].

### 3.5 LANGUAGE IN EDUCATION

Particularly in a bilingual society like in Malta, several aspects play a role when it comes to language in education. One aspect is the language of instruction, i. e., the language that is used officially by the teachers during lessons in school or in seminars at university.

Another factor is the language used in certain school books. With English being the language of technology and natural sciences, most of the school books on these topics are in English. In fact, efforts to translate technical and scientific terms into Maltese have encountered several problems, one of them being the acceptance by the language community. Hence the school subjects, too, possibly determine the language of instruction for certain lessons, although it can also be that English school books (and the English terminology contained therein) are used while the language of instruction is Maltese.

Yet another aspect is the language used by individuals. Bilingual speakers not only use different languages in different social settings (“domains”), e. g., Maltese with the family at home, English with foreigners, Maltese or English during school lessons etc. They also tend to mix both languages, either by language mixing (e. g., English words are mixed into a conversation conducted in Maltese) or by code-switching (e. g., a conversation in Maltese switches to English and back again, with the English parts being larger than just single words, often consisting of several sentences). Thus even during school lessons that are taught in one language, conversations between teachers and students can switch between the languages [21].

Keeping these three factors in mind, it becomes clear that the actual exposure of students to the respective language in schools or at the university is something different from the chosen language of instruction.

Regarding the official language of instruction in education, both Maltese and English can be found in schools

and at the university, since Maltese and English share the status as official languages in Malta. In schools, both are taught as subjects from early on. Which language is used as language of instruction depends on the type of school. Private schools tend to use English more than Maltese (sometimes to a greater extent), while in state schools Maltese is slightly preferred to English. Church schools have their individual preferences in that some traditionally prefer one language over the other.

As was mentioned before, most science books that are used in school are in English. Thus, with the introduction of more and more scientific subjects later in school and even more so at the university, students are exposed to the two languages at the same time, using them for different situations: they might have their lessons taught in Maltese, but read their books and write their essays in English. Especially for students at university, conversations between them, friends and lecturers often take place in Maltese, sometimes code-switching/mixing between Maltese and English, or they are even in English only (the latter for example with inter-national students or lecturers).

At home with their family and friends, however, most Maltese speak Maltese, some mix languages and only a few families speak English only. As can be seen from the examples above, despite the fact that both Maltese and English are used as languages in education, there is a clear distribution when it comes to their use in society. Sciriha and Vassallo (2001, p. 29, cited in [3]) point out that “70% of the respondents claimed to use Maltese at work, while 90% said they communicate with their family members at home in Maltese. ... the percentages for spoken Maltese are extremely high but go down for other skills like reading and writing.”

This distribution of Maltese being used mainly as the spoken medium and English mainly as the written medium bears a certain risk, as it can have an impact on different skills of its native speakers when it comes to

speaking, reading or writing. In order to give reasons to this statement, one has to look at the basic characteristics of spoken and written language.

In general, written texts differ from spoken discourse in a number of ways. What they have in common is that both are ways of transferring information between two parties, i. e., speaker and hearer, and writer and reader, respectively. However, they differ in the way information is passed on between them. Putting it in a simple way, a written text, unlike spoken discourse, is set outside a concrete interactive communicative situation. Spoken discourse, on the one hand, depends on the interaction between speaker and hearer. The speaker has to structure the information in a certain way. This is important because of the limited human short-term memory: a hearer in a conversation can only take in a certain amount of information before he has to interrupt and ask the speaker to make sure that he understood.

A written text, on the other hand, is non-interactive in so far as the reader cannot ask for more specific information. He can however, browse back and forth in the text (something that a hearer cannot do in discourse). In that way, a written text itself serves as the long-term memory for the reader. Thus, a written text structures information differently than would be done in a spoken conversation. For example, a text has to provide more background information in order to provide a common ground with the reader before the actual information flow starts. This is not a problem, given that a text can serve as a long-term memory for the reader. In fact, it allows for a more elaborated structure than spoken discourse, i. e., it usually contains longer sentences and a higher number of subordinate clauses.

This register (i. e., “language style”) distinction is what in the literature, e. g., [22], has been dubbed *oral* versus *literate* text structures. Of course, a text can also be written in an oral register that resembles spoken conversations (e. g., in forum chats or informal emails). But it

is not the register normally used in, e. g., essays. Ideally, native speakers acquire the literate register already from an early age on, e. g., by their parents reading stories to them. Later in school, this knowledge is deepened by active exercise in writing essays, for example.

A literate register develops over time in a language with a literary tradition. Maltese, compared to its short history as an officially written language (since 1934) has a long and rich literary history. Even though the oldest literature discovered is very sparse (*Il Cantilena* by Pietro Caxaro, dating back to about 1450), a literary tradition started to form around the 1740s. In the 19th century, the amount of literature in Maltese was growing [3], and with it, Maltese was expanding. Today it is a language with a fully fledged literate register.

This register, however, needs to be exercised in order to keep up the status of the language as a both conversational and literary language. The trend in higher education to write more essays in English than in Maltese, at least theoretically, bears the risk of reducing Maltese to the oral register. A higher number of Maltese websites of all genres is desirable to cover both registers and their subtypes in order to ensure a stable status of the language in all its richness.

### 3.6 INTERNATIONAL ASPECTS

Bearing the previous sections in mind, it should be clear now that the international aspects of Maltese differ to a great extent from other languages. With under a million native speakers worldwide, Maltese is considered a “lesser-spoken” language. In its history, it was not the language of occupiers but rather one of the occupied. As a result of this, Maltese has never become what is traditionally considered an international language or lingua franca as was the case for, e. g., Latin, Spanish, Portuguese or English, all of which can be considered as the languages of conquerors. It did spread to other countries, where it is still spoken today (Australia, Canada,

USA and UK), but only as a community language. It took nearly 200 years from the first interest of Maltese grammarians in their own language until it eventually gained the status of an official language. And even then, the other official language, English, still served as the language for international relations.

A change for Maltese to become an internationally visible language came with Malta's joining of the EU in 2004. Since then, it has been an official language inside the European Union, with all the benefits and challenges which are connected to this status.

Academically, the interest in Maltese as subject of science goes back to as far as 1603 when Hieronymus Megiser published his *Thesaurus Polyglottus*, which included a list of Maltese words. The first scholar to systematically explore and promote the Maltese language was Mikiel Anton Vassalli. He published a grammar (1790), a dictionary (1797) and several alphabets (1788 and 1790) for Maltese and today is called "the father of the Maltese Language" [23]. In the 20th century, Sutcliffe's *Grammar of the Maltese Language* (1936) was published. From the 1960s, Maltese language Linguistics gained wider international academic awareness through the publications by Joseph Aquilina (e. g., *Papers in Maltese Linguistics* (1961) and *Maltese-English Dictionary*, two volumes (1987 and 1990)). Since then, more and more scholars outside Malta have taken an interest in Maltese. 2007 saw the foundation of the International Association of Maltese Linguistics (*Għaqda Internazzjonali tal-Lingwistika Maltija*) [24], an association of linguists who are interested in the Maltese language. The main aim of GHILM, as stated on its website, is to provide "a connection between interested scholars from all subdisciplines of Linguistics", thus facilitating research on Maltese. It also wants to bring together people from different backgrounds who work with the Maltese language (linguists, translators, students and others).

### 3.7 MALTESE ON THE INTERNET

A survey of the National Statistics Office of Malta in the second quarter of 2009 [25] shows that among a population of roughly 400,000 persons, 67 per cent had access to a computer and 64 per cent had access to the Internet. A recent Eurobarometer survey (published in May 2011) [26] among European Internet users' browsing habits showed that only 6.5 per cent of Maltese Internet users use exclusively Maltese on the Internet when reading, consuming content or communicating. Instead, 90.6 per cent choose to browse websites in English and 20.1 per cent Italian, respectively. These figures formed the basis of an article in the Maltese daily newspaper *The Times of Malta*, which provoked a lively discussion mostly among Maltese readers of the online edition [27]. The exact findings in the survey, however, point to the conclusion that this habit is not a deliberate choice: When asked which language Maltese considered their mother tongue, 89.5 per cent of the respondents claimed that Maltese was their mother tongue (opposed to only 7.6 per cent for English and 0.2 per cent for Italian).

Languages other than respondents' own used to read or watch content on the Internet were English (90.6 per cent) and Italian (20.1 per cent). Only 6.5 per cent responded that they only use their own language, which is not surprising, given that most Maltese are bilingual in Maltese and English and a considerable number speak Italian as well. When writing on the Internet, numbers in favour of Maltese are higher than when reading or watching content: 87 per cent claimed they used Maltese, 85 per cent English and 8 per cent Italian. The reason for the majority to use English as the language for consuming online content may be just the limited number of websites in Maltese rather than the favour for English *per se*. Remember that most respondents did not regard English as their own language and that the usage of Maltese increased when producing content on the

web, even though this use of Maltese in most cases takes place in chat forums and social platforms, hence in a colloquial style, i. e., in the oral register.

A peculiarity about the Maltese used by the younger generation in social platforms and chat forums is its phonetic spelling, without the silent characters like *gh* and *h*. Thus *ghax* ‘because’ is written as *ax*, *tieghi* ‘my’ as *tiei* etc. The reason for this may be the late introduction of Maltese special characters into the PC world. Although Maltese has been implemented in the Unicode framework since its inception, computers and operation systems followed much later. The *Maltese Standards Authority* only released a standardised Maltese keyboard layout in 2002, and Microsoft’s Windows operating system has been available in a Maltese language version since as late as 2006 (with Windows XP). In the case of mobile phones, the special Maltese letters are still not implemented. Hence it remains to be seen whether the *ad-hoc* orthography of the chat forums will give way to a spelling with special characters once they are available on mobile phones or whether this phonetic orthography will survive as a “sociolect” of the younger generation [28].

As for the amount of Maltese on the Internet in general, it is hard to come up with exact numbers, not least because the number of websites is changing constantly. But there are other factors which give an idea about the amount of Maltese online in comparison to other languages. A first look at the number of Wikipedia entries (on June 1st, 2011) showed that there were about 2,820 entries in Maltese in contrast to more than 3,640,000 entries in English and more than 1,238,000 entries in German. Comparing the number of top level domains (TLD), the TLD *.mt* occupies rank 213 (out of 358) with an unspecified number of registered *.mt* domains (a member of the Network Information Centre Malta gave an estimate of about 5,000), opposed to 21,336,063 registered domains for *.com* (commercial, rank 1) and

5,459,604 domains for *.de* (Germany, rank 2). Of course, the number of registered domains does not tell anything about the language in which the pages under a certain domain are written.

Some rough numbers of the amount of Maltese language on the Internet can be calculated using a procedure proposed by [29] (The authors are indebted to Dr. Albert Gatt (Institute of Linguistics, University of Malta) for drawing their attention to this paper.). The basic idea is that function words (e. g., *but*, *for*, *this* etc) are more frequent than content words (e. g., nouns, verbs, adjectives) and form a finite set in a language. Also, the percentage of the function words in a language is stable in a text sample as the size of the sample increases (Zipf’s Law). Thus, one can calculate the amount of words for any language on the Internet as follows:

In the first step, one calculates the amount of selected function words of Maltese in a corpus (i. e., a text collection) whose size is known. In the second step, one uses a search engine (e. g., Google) to find out the frequency for the same function words on the web. In the third step, the frequency from the corpus count is extrapolated to the Google search and then an average is calculated for the frequency of function words in the search results.

Some restrictions of this method should be mentioned: Firstly, the numbers gained by this method are only page hits. For example, 94,300 Google hits for the word *ghal* ‘for’ are not 94,300 instances of the word on the Internet, but 94,300 webpages which contain the word *ghal* at least once. Secondly, the search only finds webpages which have an individual URL [29]. Pages that are only accessible via a web interface are not retrieved in the web search. Thirdly, a search engine will only search for a string irrespective of its environment on a webpage. It does not make judgements about whether a certain string is actually a word of a language. Applied to Maltese function words, the method described above gener-



ates different estimates for Maltese. For websites in the domain .mt situated in Malta, the estimated size is 50 million words, while for .mt websites in all regions the size is 500 million words. The reason for this difference is that a lot of .mt domains are reserved for servers outside Malta.

The exact results of the Google searches (conducted on July 8, 2011) and their extrapolation can be retraced in Figure 1 and Figure 2 below. The column f/m (i. e., frequency per million) identifies how often in a million words the respective word occurs in the MLRS corpus. For example, in Figure 1, the word *ghal* ‘for’ appears nearly 3731 times among a million words. The Google search for *ghal* results in 94,300 pages with at least one instance of *ghal* on a webpage under the domain .mt in Malta. Multiplication by 1 million and division by 3730.96 makes an estimated 25,274,996 instances of any Maltese word on pages under the domain .mt inside Malta. If one does this calculation for the other words in the figure and averages the results, one arrives at a number slightly less than 50 million words. For all webpages worldwide listed under the domain .mt, the results are ten times as high.

Of course, for a serious study, this search and extrapolation would have to include more words to arrive at more reliable numbers for the amount of Maltese on the Internet. But comparing the results with Table 3 in [29], one can say that both numbers are very low: for webpages in Malta only, the amount is more than Latvian and less than Icelandic ten years ago (the numbers in Table 3 were calculated in March 2001). For webpages worldwide, the amount of Maltese is more than Hungarian and less than Czech ten years ago. Given that “the proportion of non-English text to English is growing” [29], Maltese might be even less represented online today than the languages just mentioned.

Apart from private home pages and blogs, there are a number of official websites in Maltese. First of all, there

is the home page of the Maltese government [30], which is available in both Maltese and English. Also, there are the Internet editions of the Maltese language daily and weekly newspapers: *In-Nazzjon*, *L-Orizzont* (daily), *Illum*, *Il-ĠENSillum*, *KullHadd*, *Leben is-Sewwa*, *It-Torċa* (weekly).

The websites of the Maltese TV and radio stations show a mixture of both English and Maltese to different degrees. For example, the website of the stations NET TV [31] and One TV [32] show a framework in English, with some articles in Maltese, even though their programme contains both Maltese and English titles. The church-owned radio station RTK [33] (Maltese and English) lets the user choose between the two languages. The website of the Public Broadcasting Services (PBS) [34] contains sections in English and sections in Maltese as does the website of Radio 101 [35]. This mixture between English and Maltese reflects the language use in everyday life. Within the programmes, however, the situation is a clearer, since the *Maltese Broadcasting Authority* has issued strict guidelines for the use of Maltese on TV and the radio. Following those, presenters should speak in either Maltese or English and not switch between the two languages [3]. Hence the programmes of the stations contain broadcasts in Maltese only and others in English only. Those are often available online as well, either as live stream or as podcasts.

Outside Malta, a big collection for Maltese texts is within the EUR-Lex [36] that hosts all official law and other documents of the European Union since 1951 in its 23 official languages. Many if not all of these openly available web documents are used in corpus projects, e. g., the *JRC-Acquis Multilingual Parallel Corpus* [37], which is a parallel corpus containing the complete text of the European Union Law in 22 languages. Another corpus that contains a growing number of visible web documents in Maltese is the corpus on the MLRS (Maltese Language Resource Server) [38].

Word	f/m	Google (.mt only, Region=mt)	Extrapolation
ghal	3730.96	94,300	25,274,996
qed	4770.79	118,000	24,733,849
minn	4833.58	173,000	35,791,276
kien	4073.83	93,800	23,025,015
biex	5276.78	179,000	33,922,202
dan	6412.28	434,000	67,682,634
kienet	1452.42	116,000	79,866,705
kienu	1465.56	135,000	92,114,959
kont	521.43	34,200	65,588,861
konna	301.39	19,400	64,368,426
jekk	2776.8	72,100	25,965,140
mhux	2101.32	79,500	37,833,362
Average			48,013,952

1: Google search, restricted to domain .mt and region Malta

Word	f/m	Google (.mt only)	Extrapolation
ghal	3730.96	1,340,000	359,156,892
qed	4770.79	966,00	202,482,188
minn	4833.58	1,240,000	256,538,632
kien	4073.83	3,100,000	760,954,679
biex	5276.78	6,530,000	1,237,497,110
dan	6412.28	3,980,000	620,684,062
kienet	1452.42	665,000	457,856,543
kienu	1465.56	436,000	297,497,202
kont	521.43	450,000	863,011,334
konna	301.39	81,600	270,745,546
jekk	2776.8	1,120,000	403,341,976
mhux	2101.32	1,040,000	494,926,998
Average			518,724,430

2: Google search, restricted to domain .mt only



# LANGUAGE TECHNOLOGY SUPPORT FOR MALTESE

Language technology is used to develop software systems designed to handle human language and are therefore often called “human language technology”. Human language comes in spoken and written forms. While speech is the oldest and in terms of human evolution the most natural form of language communication, complex information and most human knowledge is stored and transmitted through the written word. Speech and text technologies process or produce these different forms of language, using dictionaries, rules of grammar, and semantics. This means that language technology (LT) links language to various forms of knowledge, independently of the media (speech or text) in which it is expressed. Figure 3 illustrates the LT landscape.

When we communicate, we combine language with other modes of communication and information media – for example speaking can involve gestures and facial expressions. Digital texts link to pictures and sounds. Movies may contain language in spoken and written form. In other words, speech and text technologies overlap and interact with other multimodal communication and multimedia technologies.

In this section, we will discuss the main application areas of language technology, i. e., language checking, web search, speech interaction, and machine translation. These applications and basic technologies include

- spelling correction
- authoring support
- computer-assisted language learning

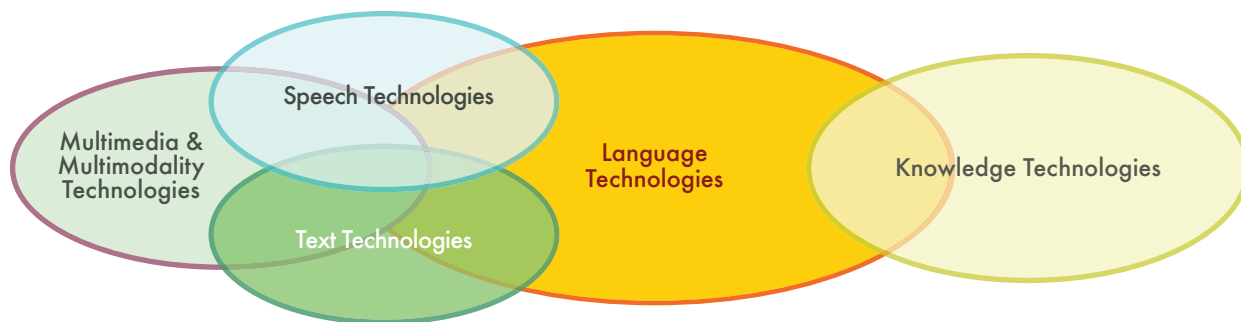
- information retrieval
- information extraction
- text summarisation
- question answering
- speech recognition
- speech synthesis

Language technology is an established area of research with an extensive set of introductory literature. The interested reader is referred to the following references: [39, 40, 41, 42, 43]. Before discussing the above application areas, we will briefly describe the architecture of a typical LT system.

## 4.1 APPLICATION ARCHITECTURES

Software applications for language processing typically consist of several components that mirror different aspects of language. While such applications tend to be very complex, figure 4 shows a highly simplified architecture of a typical text processing system. The first three modules handle the structure and meaning of the text input:

1. Pre-processing: cleans the data, analyses or removes formatting, detects the input languages, and so on.
2. Grammatical analysis: finds the verb, its objects, modifiers and other sentence elements; detects the sentence structure.



3: Language technologies

3. Semantic analysis: performs disambiguation (i. e., computes the appropriate meaning of words in a given context); resolves anaphora (i. e., which pronouns refer to which nouns); represents the meaning of the sentence in a machine-readable way.

After analysing the text, task-specific modules can perform other operations, such as automatic summarisation and database look-ups.

In the remainder of this section, we firstly introduce the core application areas for language technology, and follow this with a brief overview of the state of LT research and education today, and a description of past and present research programmes. Finally, we present

an expert estimate of core LT tools and resources for Maltese in terms of various dimensions such as availability, maturity and quality. The general situation of LT for the Maltese language is summarised in figure 9 (p. 76) at the end of this chapter. This table lists all tools and resources that are boldfaced in the text. LT support for Maltese is also compared to other languages that are part of this series.

## 4.2 CORE APPLICATION AREAS

In this section, we focus on the most important LT tools and resources, and provide an overview of LT activities in Malta.



4: A typical text processing architecture

### 4.2.1 Language Checking

Anyone who has used a word processor such as Microsoft Word knows that it has a spell checker that highlights spelling mistakes and proposes corrections. The first spelling correction programs compared a list of extracted words against a dictionary of correctly spelled words. Today these programs are far more sophisticated. Using language-dependent algorithms for **grammatical analysis**, they detect errors related to morphology (e. g., plural formation) as well as syntax-related errors, such as a missing verb or a conflict of verb-subject agreement (e. g., *she \*write a letter*). However, most spell checkers will not find any errors in the following text [44]:

I have a spelling checker,  
It came with my PC.  
It plane lee marks four my revue  
Miss steaks aye can knot sea.

For handling this type of error, analysis of the context is needed in many cases, e. g., for deciding in which position in a Maltese verb the silent *gh* has to be written, as in:

1. ... *in-negożjati li kien għamel il-Gvern ...*  
'... the negotiations that the government had made..'
2. *Pawlu, aghmel l-eżamijiet!*  
'Paul, do the exams!'
3. \*... *in-negożjati li kien aghmel il-Gvern ...*

Both *għamel* 'he made' and *aghmel* 'make!' are pronounced ['e:.mɛl].

This type of analysis either needs to draw on language-specific **grammars** laboriously coded into the software by experts, or on a statistical language model. In this case, a model calculates the probability of a particular word as it occurs in a specific position (e. g., between the words that precede and follow it). For example, *kien għamel* is much more probable word sequence than *kien*

*aghmel*. A statistical language model can be automatically derived using a large amount of (correct) language data, a **text corpus**. Up to now, these approaches have mostly been developed and evaluated using English language data. However, they do not necessarily transfer well to highly inflectional languages like Maltese, where a given word type, such as a verb, can yield a large number of orthographic forms.

As with other languages, a means to determine whether a given string is a valid word is not a sufficient condition for spelling-error detection, but it is a necessary condition. As yet, no such means exists for Maltese, though various attempts have been made.

One of the earliest was by [45] using a rudimentary form of rule-driven morphological analysis. Essentially a word was considered valid if it could be derived by rule from a citation form found in a dictionary. The problem with this approach is that it requires a complete list of all citation forms, and of course, the rules have to be very accurate. Results were somewhat limited by the list of citation forms, which was incomplete, and the imperfect nature of the rules.

A second approach looked to statistics for a solution. The intuitive idea is that for a given language, certain sequences of characters are highly unlikely. In English, for example, we never find the sequence *kk*, so if that occurs as a substring in a written word, we can guess, with a high degree of confidence, that the word is not valid. More generally, we can calculate the probability of any string as a function of the probabilities of all its substrings, adopting the principle that to count as a valid word, that probability must exceed a certain threshold. A statistical spell checker making use of such a principle was developed by [46]. It did not require a lexicon, being based instead on the distribution of character n-grams found in a newspaper corpus. It became clear that for this approach to succeed (i) a more accurate language model was needed requiring more language data than was then



5: Language checking (top: statistical; bottom: rule-based)

available, and (ii) that string probability alone was insufficient to accurately classify an orthographic word as an error. As suggested above, other information is necessary, such as part of speech information from the surrounding context.

Other attempts to develop a spell-checker for Maltese include an online checker that has been developed by Ramon Casha of the Linux User Group [47]. This is based on a wordlist of around 1 million word types originally collected from various corpora, and subsequently extended by various rules for handling inflections. Its accuracy has not been officially established. Microsoft has also been working on a spell checker for inclusion with their Maltese language interface pack though it is not clear when this will be released. The use of language checking is not limited to word processing tools. Language checking is also applied to automatically correct queries sent to search engines, e. g., Google's *Did you mean...* suggestions. Other application areas include various kinds of authoring support software.

As a result of the rapid increase in demand for technical products, many companies have begun to focus increasingly on the quality of technical documentation in the face of potential customer complaints about wrong linguistic usage and damage claims resulting from bad or badly understood instructions. Authoring support software can assist the writer of technical documentation to use vocabulary and sentence structures that are consis-

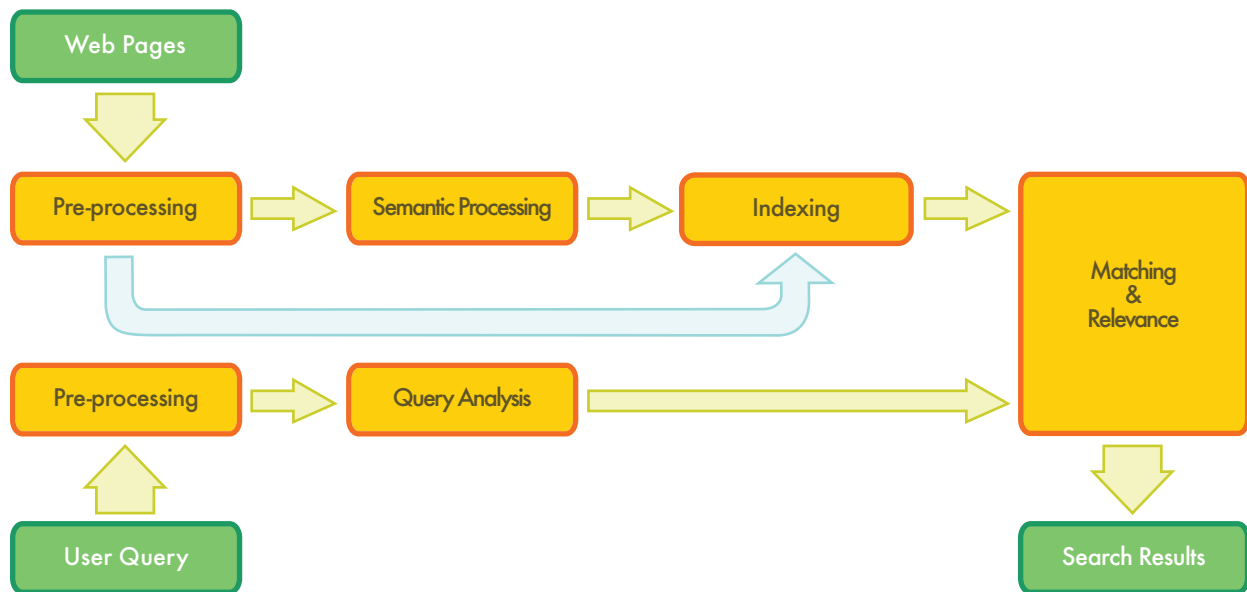
tent with certain formally expressed rules and (corporate) terminology restrictions.

Authoring support software for Maltese does not at the moment exist but there would be considerable scope for the use of such software at the production end of Maltese. One of the reasons for the comparative scarcity of written Maltese content, in business correspondence for example, is that the production of correct Maltese text is difficult. Many competent native speakers are inclined to make mistakes when it comes to the written language and so they prefer to write in English. The availability of the right kind of simple authoring support tools could alleviate this problem.

An evolving area of language technology is computer-assisted language learning but apart from an interactive CD picture dictionary [48], no such applications have been specifically developed for Maltese to date.

#### 4.2.2 Web Search

Search on the web, in intranets, or in digital libraries is probably the most widely used and yet underdeveloped language technology today. The search engine Google, which started in 1998, is nowadays used for about 80% of all search queries world-wide [49]. Since 2004, the verb *to google* even has an entry in the *Cambridge Advanced Learner's Dictionary*. Neither the search interface nor the presentation of the retrieved results has significantly changed since the first version. In the cur-



## 6: Web search

rent version, Google offers spelling correction for misspelled words and also, in 2009, incorporated basic semantic search capabilities into their algorithmic mix [50], which can improve search accuracy by analysing the meaning of the query terms in context. The success story of Google shows that with a lot of data at hand and efficient indexing techniques a mainly statistical approach can lead to satisfactory results.

However, for more sophisticated information requests, the integration of deeper linguistic knowledge is essential. In the research labs, experiments using **lexical resources** such as machine-readable thesauri and ontological language resources like WordNet have shown improvements by allowing a page to be found on the basis of synonyms of the search terms, e. g., Maltese *enerġija atomika*, *enerġija nukleari* (atomic energy, nuclear energy) or even more loosely related terms.

The next generation of search engines will have to include much more sophisticated language technology. If

a search query consists of a question or another type of sentence rather than a list of key-words, retrieving relevant answers to this query requires a syntactic and **semantic analysis** of the sentence as well as the availability of an index that allows for a fast retrieval of the relevant documents. For example, imagine a user inputs the query “Give me a list of all companies that were taken over by other companies in the last five years”. For a satisfactory answer, syntactic parsing needs to be applied to analyse the grammatical structure of the sentence and determine that the user is looking for companies that have been taken over and not companies that took over others. Also, the expression *last five years* needs to be processed in order to find out which years it refers to.

Finally, the processed query needs to be matched against a huge amount of unstructured data in order to find the piece or pieces of information the user is looking for. This is commonly referred to as information retrieval and involves the search for and ranking of relevant doc-

uments. In addition, generating a list of companies, we also need to extract the information that a particular string of words in a document actually refers to a company name. This kind of information is made available by so-called named-entity recognisers.

---

The next generation of search engines will have to include much more sophisticated language technology.

---

Even more demanding is the attempt to match a query to documents written in a different language. For cross-lingual information retrieval, we have to automatically translate the query to all possible source languages and transfer the retrieved information back to the target language. The increasing percentage of data available in non-textual formats drives the demand for services enabling multimedia information retrieval, i. e., information search on images, audio, and video data. For audio and video files, this involves a **speech recognition** module to convert speech content into text or a phonetic representation, to which user queries can be matched.

In Malta, there are a number of search websites that are specifically oriented towards Malta [51]. In addition there are a small number of Malta based SMEs that incorporate relatively sophisticated language processing techniques within search applications. Charonite [52], for example, is a local SME dealing with search engine optimisation. However, at the time of writing there are no commercially available search engines that are specifically oriented towards the Maltese language, apart from a prototype for cross lingual information retrieval developed within the scope of LT4EL [53], a European FP6 research project which used multilingual language technology tools and semantic encoding techniques for improving the retrieval of learning material.

### 4.2.3 Speech Interaction

Speech interaction is one of many application areas that depend on speech technology, i. e., technologies for processing spoken language. Speech interaction technology is used to create interfaces that enable users to interact in spoken language instead of using a graphical display, keyboard and mouse. Today, these voice user interfaces (VUI) are used for partially or fully automated telephone services provided by companies to customers, employees or partners. Business domains that rely heavily on VUIs include banking, supply chain, public transportation, and telecommunications. Other uses of speech interaction technology include interfaces to car navigation systems and the use of spoken language as an alternative to the graphical or touchscreen interfaces in smartphones.

---

Speech interaction is the basis for interfaces that allow a user to interact with spoken language.

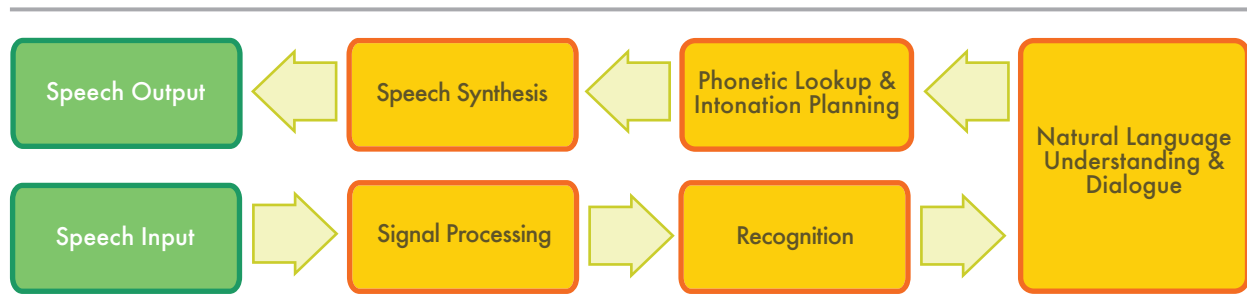
---

Speech interaction technology comprises four technologies:

1. Automatic **speech recognition** (ASR) determines which words are actually spoken in a given sequence of sounds uttered by a user.
2. Natural language understanding analyses the syntactic structure of a user's utterance and interprets it according to the system in question.
3. Dialogue management determines which action to take given the user input and system functionality.
4. **Speech synthesis** (text-to-speech or TTS) transforms the system's reply into sounds for the user.

One of the major challenges is to have an ASR system recognise the words uttered by a user as precisely as possible. This requires either a restriction of the range of possible user utterances to a limited set of keywords,





7: Speech-based dialogue system

or the manual creation of language models that cover a large range of natural language user utterances. Using machine learning techniques, language models can also be generated automatically from **speech corpora**, i. e., large collections of speech audio files and text transcriptions. Restricting utterances usually forces people to use the voice user interface (VUI) in a rigid way and can damage user acceptance; but the creation, tuning and maintenance of rich language models will significantly increase costs. VUIs that employ language models and initially allow a user to express their intent more flexibly – prompted by a *How may I help you?* greeting – are better accepted by users.

Companies tend to use pre-recorded utterances of professional speakers for generating the output of the voice user interface. For static utterances, where the wording does not depend on the particular contexts of use or the personal user data, this can deliver a rich user experience. But more dynamic content in an utterance may suffer from unnatural intonation because different parts of audio files have simply been strung together. Through optimisation, today's TTS systems are getting better at producing natural-sounding dynamic utterances.

Interfaces in speech interaction have been considerably standardised during the last decade in terms of their various technological components. There has also been strong market consolidation in speech recognition and

speech synthesis. The national markets in the G20 countries (economically resilient countries with high populations) have been dominated by just five global players, with Nuance (USA) and Loquendo (Italy) being the most prominent players in Europe. In 2011, Nuance announced the acquisition of Loquendo, which represents a further step in market consolidation.

Most speech interaction technology development in Malta has concentrated on text-to-speech (TTS). Some pioneering work was initially carried out by [54] and this was followed by a number of Master's dissertations [55]. Some work on a web-based TTS system was initiated by [56].

A significant recent development for Maltese speech synthesis was the winning of a government tender for the development of a speech synthesiser by the local company Crimson Wing Malta Ltd. This work is partly financed by the EU Regional Development fund and commissioned by the Maltese Foundation for Information Access (FITA). The prototype will be SAPI compliant and will include three voices (male, female, and child). According to a recent presentation [57] the work is advancing well and a prototype, expected in 2012, will be freely available for download.

Work on speech recognition is less advanced. A prototype for recognizing numerals was created by [58] in simple domains. With respect to speech, the fundamen-

tal problem remains a lack of suitably annotated data since this requires significant manual effort. Some attempts at automatic annotation have been made by [59]. The creation of a corpus and descriptive framework for the study of Maltese intonation was initiated by the Institute of Linguistics carried out by [60]. It is expected that the corpora being developed by Crimson Wing will be made available for research.

Looking beyond the state of today's technology, there will be significant changes due to the spread of smartphones as a new platform for managing customer relationships – in addition to the telephone, Internet, and email channels. This tendency will also affect the employment of technology for Speech Interaction. On the one hand, demand for telephony-based VUIs will decrease in the long run. On the other hand, the usage of spoken language as a user-friendly input modality for smartphones will gain significant importance. This tendency is supported by the observable improvement of speaker-independent speech recognition accuracy for speech dictation services that are already offered as centralised services to smartphone users. Given this 'outsourcing' of the recognition task to the infrastructure of applications, the application-specific employment of linguistic core technologies will supposedly gain importance compared to the present situation.

#### 4.2.4 Machine Translation

The idea of using digital computers for translation of natural languages came up in 1946 by A. D. Booth and was followed by substantial funding for research in this area in the 1950s and beginning again in the 1980s. Nevertheless, **Machine Translation** (MT) still fails to fulfill the high expectations it gave rise to in its early years. The most basic approach to machine translation is the automatic replacement of the words in a text written in one natural language with the equivalent words of another language.

---

At its basic level, Machine Translation simply substitutes words in one natural language with words in another language.

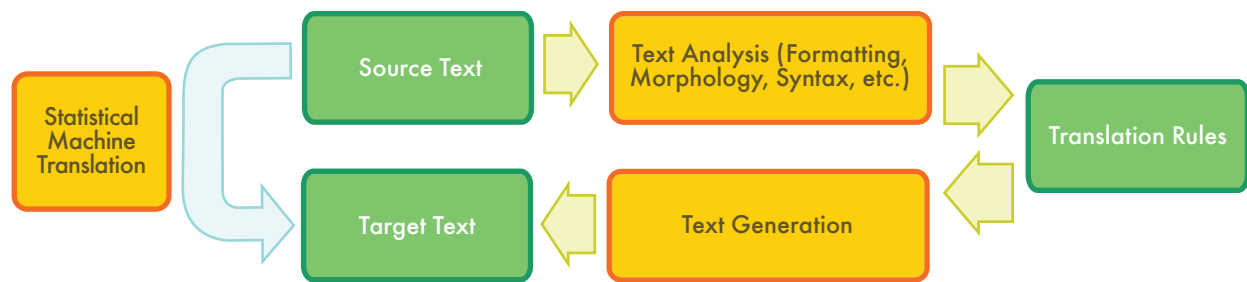
---

This can be useful in subject domains with a very restricted, formulaic language, e. g., weather reports. However, for a good translation of less standardised texts, larger text units (phrases, sentences, or even whole passages) need to be matched to their closest counterparts in the target language. The major difficulty here lies in the fact that human language is ambiguous, which yields challenges on multiple levels, e. g., word sense disambiguation at the lexical level ('Jaguar' can mean a car or an animal) or the attachment of prepositional phrases on the syntactic level as in:

1. *Il-Kuntistabbli osserva lir-ragel bit-teleskopju.*  
'The policeman observed the man with the telescope.'
2. *Il-Kuntistabbli osserva lir-ragel bir-rivolver.*  
'The policeman observed the man with the revolver.'

One way of approaching the task is based on linguistic rules. For translations between closely related languages, a direct translation may be feasible in cases like the example above. But often rule-based (or knowledge-driven) systems analyse the input text and create an intermediary, symbolic representation, from which the text in the target language is generated. The success of these methods is highly dependent on the availability of extensive lexicons with morphological, syntactic, and semantic information, and large sets of grammar rules carefully designed by a skilled linguist.

Beginning in the late 1980s, as computational power increased and became less expensive, more interest was shown in statistical models for MT. The parameters of these statistical models are derived from the analysis of bilingual text corpora, such as the Europarl **parallel corpus**, which contains the proceedings of the European



8: Machine translation (left: statistical; right: rule-based)

Parliament in 21 European languages. Given enough data, statistical MT works well enough to derive an approximate meaning of a foreign language text. However, unlike knowledge-driven systems, statistical (or data-driven) MT often generates ungrammatical output. On the other hand, besides the advantage that less human effort is required for grammar writing, data-driven MT can also cover particularities of the language that go missing in knowledge-driven systems, for example idiomatic expressions.

As the strengths and weaknesses of knowledge- and data-driven MT are complementary, researchers nowadays unanimously target hybrid approaches combining methodologies of both. This can be done in several ways. One is to use both knowledge- and data-driven systems and have a selection module decide on the best output for each sentence. However, for longer sentences, no result will be perfect. A better solution is to combine the best parts of each sentence from multiple outputs, which can be fairly complex, as corresponding parts of multiple alternatives are not always obvious and need to be aligned.

---

The quality of MT systems is still considered to have huge improvement potential.

---

In Malta work carried out in Machine Translation has been restricted to just a few Bachelors and Masters dissertations. A transfer system based on LFG was developed for English/Maltese by [61] and successfully translated weather forecasts. Later J. Bajada [62, 63] worked on statistical MT (SMT) with the emphasis on techniques for producing language and translation models. The earlier work concerned word-based models, whilst the latter developed techniques for gathering bilingual phrase data from a limited corpus.

Like in so many other areas, the underlying problem is a lack of sufficient quantities of suitably annotated bilingual data. For this reason, perhaps, the benchmark system against which to judge advances remains Google Translate.

The quality of MT systems is still considered to have huge improvement potential. Challenges include the adaptability of the language resources to a given subject domain or user area and the integration into existing workflows with term bases and translation memories. In addition, most of the current systems are English-centred and support only few languages from and into other languages, which leads to frictions in the total translation workflow, and, e. g., forces MT users to learn different lexicon coding tools for different systems.

Evaluation campaigns allow for comparing the quality of MT systems, the various approaches and the status of MT systems for the different languages. Figure 9 (p. 32), which was prepared during the EC Euromatrix+ project, shows the pairwise performances obtained for 22 of the 23 official EU languages (Irish was not compared). The results are ranked according to a BLEU score, which indicates higher scores for better translations [64]. A human translator would normally achieve a score of around 80 points.

The best results (shown in green and blue) were achieved by languages that benefit from considerable research efforts, within coordinated programs, and from the existence of many parallel corpora (e. g., English, French, Dutch, Spanish, German), the worst (in red) by languages that did not benefit from similar efforts, or that are very different from other languages (e. g., Hungarian, Maltese, Finnish).

### 4.3 OTHER APPLICATION AREAS

Building language technology applications involves a range of subtasks that do not always surface at the level of interaction with the user, but provide significant service functionalities “behind the scenes”. They form important research issues that have now evolved into individual subdisciplines of computational linguistics. Question answering, for example, is an active area of research for which annotated corpora have been built and scientific competitions have been initiated. The concept of question answering goes beyond keyword-based search (in which the search engine responds by delivering a collection of potentially relevant documents) and enables users to ask a concrete question to which the system provides a single answer, e. g.,

*Question: How old was Neil Armstrong when he stepped on the moon?*

*Answer: 38.*

While question answering is obviously related to the core area of web search, it is nowadays an umbrella term for such research issues as which different types of questions exist, and how they should be handled; how a set of documents that potentially contain the answer can be analysed and compared (do they provide conflicting answers?); and how specific information (the answer) can be reliably extracted from a document without ignoring the context.

---

Language technology applications often provide significant service functionalities “behind the scenes” of larger software systems.

---

This is in turn related to the information extraction (IE) task, an area that was extremely popular and influential at the time of the “statistical turn” in Computational Linguistics, in the early 1990s. IE aims at identifying specific pieces of information in specific classes of documents; this could be, e. g., the detection of the key players in company takeovers as reported in newspaper stories. Another scenario that has been worked on is reports on terrorist incidents, where the problem is to map the text to a template specifying the perpetrator, the target, time and location of the incident, and the results of the incident. Domain-specific template-filling is the central characteristic of IE, which for this reason is another example of a “behind the scenes” technology that constitutes a well-demarcated research area but for practical purposes then needs to be embedded into a suitable application environment.

Two “borderline” areas, which sometimes play the role of standalone application and sometimes that of supportive, “under the hood” component are text summarisation and **text generation**. Summarisation, obviously, refers to the task of making a long text short, and is offered for instance as a functionality within MS Word. It works largely on a statistical basis, by first identifying

‘important’ words in a text (that is, for example, words that are highly frequent in this text but markedly less frequent in general language use) and then determining those sentences that contain many important words. These sentences are then marked in the document, or extracted from it, and are taken to constitute the summary. In this scenario, which is by far the most popular one, summarisation equals sentence extraction: the text is reduced to a subset of its sentences. All commercial summarisers make use of this idea. An alternative approach, to which some research is devoted, is to actually synthesise *new* sentences, i. e., to build a summary of sentences that need not show up in that form in the source text. This requires a certain amount of deeper understanding of the text and therefore is much less robust. All in all, a text generator is in most cases not a stand-alone application but embedded into a larger software environment, such as into the clinical information system where patient data is collected, stored and processed, and report generation is just one of many functionalities.

## 4.4 EDUCATIONAL PROGRAMMES

Language technology is a highly interdisciplinary field, involving the expertise of linguists, computer scientists, mathematicians, philosophers, psycholinguists, and neuroscientists, among others.

In Malta the vast majority of research and education in LT has taken place at the University of Malta. However, it was established rather late. One reason for this was the late appearance of Computer Science as a curriculum subject at the University. The turbulent political leadership of the country during the 1970s and 1980s had not foreseen the information revolution to come and it was not until the early 1990s that an undergraduate option in Computing with Mathematics was offered through the Faculty of Science.

The roots of change came in 1994, when a national strategic initiative was undertaken to recognise and strengthen the role of IT in commercial, political, and above all, educational sectors. One immediate consequence of this was the introduction of a substantial four-year Bachelors programme – the BSc. IT (Hons) – at University as well as the founding of a new Department of Computer Science and Artificial Intelligence (CSAI, renamed “Department of Intelligent Computer Systems (ICS)” in 2009). A course in NLP was included as an advanced option, and this led, four years later, to a series of undergraduate final year projects tackling language processing issues including computational approaches to Maltese [66, 45, 67, 61, 46, 62, 68, 69, 70, 71]. The Department of Computer Communications Engineering also participated in the programme, and this led to another set of undergraduate projects in speech technology.

Another important influence on research is the University’s Institute of Linguistics (IOL), founded in 1988 with the aim of teaching as well as promoting and coordinating research in both General and Applied Linguistics, furthering research involving the description of particular languages, not least Maltese, fostering the study of the various sub-fields of linguistics, and promoting interdisciplinary research involving academics in practical cooperation that cuts across departmental and faculty boundaries abroad. The Institute of Linguistics runs two undergraduate programmes: a B.A. in General Linguistics and a new B.Sc. in Human Language Technology which will be on offer in October 2011. It is also possible to do a Masters Degree and a Ph.D. in Linguistics with the Institute.

In 1997, an interdisciplinary group of computer scientists and linguists embarked on Maltilex (M. Rosner, R. Fabri, J. Caruana, M. Montebello and others), a project to create a computational lexicon, which was sustained by a small grant from the University sup-

ported by the Mid-Med Bank. A simple web-based interface was developed to enable the creation and maintenance of entries, as reported in [72] at the first ACL Workshop on Computational Approaches to Semitic Languages [83]. Several thousand such entries were created by hand, but the project ran into legal problems, the compilation of entries having been largely inspired by Joseph Aquilina's existing paper dictionary [73, 74].

Effort then shifted from paper dictionaries to extraction of lexical entries from other sources. Two [75, 68] used techniques based on alignment derived from bioinformatics to cluster lexical entries and this was used as a means of structuring the lexicon automatically.

Despite lack of funding, the Maltilex effort continued in a somewhat piecemeal fashion, supported by staff at the IOL and CSAI Department. It was not until 2005 that Malta's Council for Science and Technology (MCST) launched the country's first Research and Technology Development Initiative and a joint proposal for a Maltese Language Resource Server (MLRS) was accepted, providing sufficient financial support to employ a researcher full time between 2006 and 2008. The project had the twin goals of creating both a lexicon and a corpus [76], and it laid the foundations for the present MLRS server.

The research mentioned above mainly deals with the written language. Two branches of speech-related work are also ongoing.

The first, initiated from the signal-processing tradition within the Engineering Faculty, yielded a prototype speech synthesiser [54]. His work has influenced several other projects aimed at improving speech synthesis from a low-resource perspective including [58, 55, 77, 57].

The second tackles the issue of intonation [78] from a linguistic perspective. Some pioneering work to create a corpus and descriptive framework for the study of Maltese intonation was carried out by [60].

Outside Malta, two research groups that are in active collaboration with local LT-oriented efforts deserve a special mention.

At the University of Arizona, a group led by linguist Adam Ussishkin is particularly interested in the psycholinguistic issues pertaining to Semitic languages including Maltese. To study these issues an online corpus has been made available [79].

At the University of Bremen, Prof. Thomas Stolz has been actively involved with the academic study of Maltese but is particularly known for having hosted the first conference on Maltese Linguistics in Bremen [80], founded a periodical [81] and the International Association of Maltese Linguistics, also based in Bremen, that exists alongside the Malta-based Council for the Maltese Language.

As mentioned, the LT-sensitive communities existing at the University of Malta mainly inhabit the Faculty of ICT, the Institute of Linguistics. There is also a potential interest in Faculty of Arts (Department of Maltese) and other Humanities subjects though up until now computational linguistics tends to be regarded as an exotic topic located in the more scientific computer science faculties or in the humanities and, therefore, the research topics dealt with only overlap only partially.

Curiously, Malta does not lack for LT-related international events. LREC 2010 was held in Valletta, drawing 1200 participants. The annual EAMT conference was also held in Malta in 1994, and there have also been a number of smaller workshops held during the last 10 years.

## 4.5 NATIONAL PROJECTS AND INITIATIVES

Malta joined the EU in 2004 and this event immediately conferred to Maltese the status of being an official EU language. With this status came new obligations – in



particular to translate large quantities of official documents, and in addition, a recognition, at European level, that as a national language, it should have “first-class” status from a technological as well as a social perspective, and be accorded all the rights and privileges enjoyed by “larger” European languages (i. e., having larger numbers of native speakers).

The government’s National IT Strategy 2008-10 included a number of objectives related to Maltese Language including (i) the development of online government in Maltese, (ii) creation of Maltese language tools, in collaboration with the University, and (iii) support for Maltese online communities. At the time of writing in 2011, not all the objectives have been realised. However the longer term effects of this strategy are beginning to take shape.

Currently the language technology scene in Malta is under the influence of four main initiatives:

1. First of all, a government-supported project partly funded by EU regional development funds is under way to bring speech technology within the reach of disabled persons. The project is currently focused on Maltese speech synthesis, and at this point the relevant language models are in the process of being developed. The consortium, which consists of an SME (Crimson Wing Ltd), a foundation (FITA, Foundation for IT Access), and the University, has pledged that these resources will be made available for research purposes. It remains to be seen whether components of the speech synthesiser will be made available to resource sharing networks inspired by CLARIN and META.
2. Second, as is evident from the current report, Malta participates in METANET4U and is thus in receipt of significant EU funding aimed at the enhancement and distribution of resources and tools that are specifically for Maltese. The University of Malta is a member of META-NET and intends to fulfill its

obligations towards the aims of META, particularly regarding the identification of stakeholders, actually and potential.

3. Third, the Maltese Language Resource Server (MLRS) [82, 38] has come to fruition and significant efforts are under way at University, through the Institute of Linguistics (A. Gatt, C. Borg, R. Fabri) and the Department of Intelligent Computer Systems (M. Rosner), to maintain and develop it. Currently MLRS is online at <http://mlrs.research.um.edu.mt>. The corpus comprises some 100M words, and the system includes basic services that include KWIC search and display, pattern-directed search, various kinds of statistical analysis etc. Further tools are planned including a part-of-speech tagger and a spell-checker.
4. Finally, a new undergraduate programme in Human Language Technology is destined to be launched by the Institute of Linguistics in October 2011. This will cover a full range of topics and will inevitably have a positive long-term effect on the study of Maltese from a computational perspective.

Besides these, a project to develop an electronic version of the Aquilina dictionary [73, 74] is currently in preparation. This is a collaborative effort between the University of Malta who are supplying the linguistic expertise, the University of Arizona, who have already digitised the dictionary into machine readable form, and the publishers Midsea Books of Valletta. The dual aims of the project are to update the content, and to confer upon researchers the flexibility to swiftly access the text. An effort is in progress locally, to organise the right level of lexicographic expertise necessary to update the content. We should also mention Malta’s relationship to CLARIN, a proposed EU research infrastructure addressing the provision of language resources for the Humanities and Social Sciences. During specification phase, the University was able to participate thanks to a

small support grant from the local Council for Science and Technology. However, it has turned out to be more challenging to secure the longer term funding required for the construction phase of CLARIN. Identification of a suitable government entity to take responsibility for the programme has so far been without success. Consequently, Malta's future participation in the construction phase currently hangs in the balance.

## 4.6 AVAILABILITY OF TOOLS AND RESOURCES

Figure 9 provides a rating for language technology support for the Maltese language. This rating of existing tools and resources was generated by leading experts in the field who provided estimates based on a scale from 0 (very low) to 6 (very high) using seven criteria.

For Maltese, the most evident characteristics revealed by the figure are that

- many entries are blank, and
- the highest grade scored is 3.2.

The fact that most entries are blank reflects the immature state of LT-related research and development in Malta. Although there are signs that the situation is improving, investment in language technology remains at a low level, and as a result, despite modest local achievements, the effort is fragmentary, both in terms of coverage of different areas, and in terms of sustainability of research: there have been too many projects involving just one area, just one researcher, and just one or two years. The collective efforts don't add up as they should. So what has been achieved? We can see by looking at the non-blank entries, whose average score yields the following ordering:

- Tools:
  1. Tokenisation, Speech Synthesis

### 2. Speech Recognition

- Resources:
  1. Reference Corpora
  2. Parallel Corpora
  3. Lexicons, Terminology (this should be understood to include wordlists)
  4. Language Models

With respect to tools, low level text extraction and processing tools are available, including a tokeniser. A POS-tagger is under development, but its performance is not state-of-the-art, pending further training with better annotated data. Higher level tools (syntactic or semantic analysis, classification tools, information extraction etc.) are entirely lacking. A consequence is that, for example, there are no treebanks available for Maltese.

Prototype speech recognition tools have been developed at University but are not readily available at the time of writing. However, the government-funded speech engine mentioned earlier should yield a working speech synthesiser by 2013. Whilst this is a very positive development, it is highly focused on the synthesis side of speech. Hardly any work on **speech recognition** is planned at this stage.

With respect to resources, the situation is a little more structured, in so far as there already exists MLRS, an extensible computational infrastructure in the form of a server providing the basic functionality to enable access over the web to available corpora, some services, and a rudimentary system to facilitate the submission of contributions. MLRS currently provides some very basic services for the extraction, representation, search and analysis of text.

The existing MLRS corpus is currently around 100 million tokens in length. It is predominantly textual and monolingual. It is also somewhat non-representative: there is an abundance of legalistic material, but a shortage of academic text and works of fiction.

	Quantity	Availability	Quality	Coverage	Maturity	Sustainability	Adaptability
<b>Language Technology: Tools, Technologies and Applications</b>							
Speech Recognition	0.8	0.8	0.8	0.8	0.8	0.8	0.8
Speech Synthesis	2.4	0.8	3.2	3.2	2.4	2.4	2.4
Grammatical analysis	0.8	0.8	0.8	0.8	0.8	0.8	0.8
Semantic analysis	0	0	0	0	0	0	0
Text generation	0	0	0	0	0	0	0
Machine translation	1.6	1.6	1.6	1.6	1.6	1.6	1.6
<b>Language Resources: Resources, Data and Knowledge Bases</b>							
Text corpora	3.2	3.2	2.4	2.4	2.4	3.2	3.2
Speech corpora	2.4	0.8	2.4	1.6	2.4	2.4	2.4
Parallel corpora	3.2	3.2	2.4	1.6	1.6	1.6	1.6
Lexical resources	2.4	2.4	1.6	2.4	2.4	2.4	2.4
Grammars	0	0	0	0	0	0	0

9: State of language technology support for Maltese

As things stand, these materials can only be searched and analysed through the server and cannot be accessed directly. The reasons are legalistic. With access restricted in this way, the complications of IPR and copyright have been neatly sidestepped. The price is that these complications will eventually have to be confronted in the future, and in fact META is in the process of formulating a set of licence agreements to suit the distribution of resources, like MLRS.

## 4.7 CROSS-LANGUAGE COMPARISON

The current state of LT support varies considerably from one language community to another. In order to com-

pare the situation between languages, this section will present an evaluation based on two sample application areas (machine translation and speech processing) and one underlying technology (text analysis), as well as basic resources needed for building LT applications. The languages were categorised using the following five-point scale:

1. Excellent support
2. Good support
3. Moderate support
4. Fragmentary support
5. Weak or no support

LT support was measured according to the following criteria:

**Speech Processing:** Quality of existing speech recognition technologies, quality of existing speech synthesis technologies, coverage of domains, number and size of existing speech corpora, amount and variety of available speech-based applications.

**Machine Translation:** Quality of existing MT technologies, number of language pairs covered, coverage of linguistic phenomena and domains, quality and size of existing parallel corpora, amount and variety of available MT applications.

**Text Analysis:** Quality and coverage of existing text analysis technologies (morphology, syntax, semantics), coverage of linguistic phenomena and domains, amount and variety of applications, quality and size of existing (annotated) text corpora, quality and coverage of lexical resources (e. g., WordNet) and grammars.

**Resources:** Quality and size of existing text corpora, speech corpora and parallel corpora, quality and coverage of existing lexical resources and grammars.

Figures 10 to 13 show that the Maltese language has only low to medium LT support and thus compares well with other less spoken languages of Europe. LT resources and tools for Maltese clearly do not yet reach the quality and coverage of comparable resources and tools for “major” languages like German, and certainly not that of those for the English language, which is in the lead in almost all LT areas. And there are still plenty of gaps in English language resources with regard to high quality applications.

## 4.8 CONCLUSIONS

*In this series of white papers, we have made an important initial effort to assess language technology support for 30 European languages, and provide a high-level comparison across these languages. By identifying the gaps, needs and deficits, the European language technology community and related stakeholders are now in*

*a position to design a large scale research and development programme aimed at building a truly multilingual, technology-enabled Europe.*

The results of this white paper series show that there is a dramatic difference in language technology support between the various European languages. While there are good quality software and resources available for some languages and application areas, others, usually smaller languages, have substantial gaps. Many languages lack basic technologies for text analysis and the essential resources. Others have basic tools and resources but the implementation of for example semantic methods is still far away. Therefore a large-scale effort is needed to attain the ambitious goal of providing high-quality language technology support for all European languages, for example through high quality machine translation.

In this report, we have tried to convey the paradoxical state of Maltese Language Technology. The paradox arises because there are significant efforts made by a small number of well-qualified people across a spectrum of LT-related activities to improve the state of the art, whether this be in terms of tools, or resources, or both. It is also clear that within the wider context of educational, commercial and cultural activities in the country, there is a place for LT to make an important contribution. The problem is that efforts that have been made are uncoordinated, short term, and fragmentary, so progress is slower than it has to be.

Sustained and directed coordination of effort is, in our opinion, the only way in which the benefits of LT for Maltese will be realised in a reasonable time. We believe that even in a country as small as Malta, the work needs to be shared out amongst different stake-holders. We must arrive at a workable roadmap via a localised version of the tripartite division of labour advocated by META: identification of a community with a shared vision; extension of an infrastructure to facilitate the sharing of resources, and reinforcement of connections between LT and neighbouring fields of research and development.

Excellent support	Good support	Moderate support	Fragmentary support	Weak/no support
	English	Czech Dutch Finnish French German Italian Portuguese Spanish	Basque Bulgarian Catalan Danish Estonian Galician Greek Hungarian Irish Norwegian Polish Serbian Slovak Slovene Swedish	Croatian Icelandic Latvian Lithuanian <b>Maltese</b> Romanian

10: Speech processing: state of language technology support for 30 European languages

Excellent support	Good support	Moderate support	Fragmentary support	Weak/no support
	English	French Spanish	Catalan Dutch German Hungarian Italian Polish Romanian	Basque Bulgarian Croatian Czech Danish Estonian Finnish Galician Greek Icelandic Irish Latvian Lithuanian <b>Maltese</b> Norwegian Portuguese Serbian Slovak Slovene Swedish

11: Machine translation: state of language technology support for 30 European languages

Excellent support	Good support	Moderate support	Fragmentary support	Weak/no support
	English	Dutch French German Italian Spanish	Basque Bulgarian Catalan Czech Danish Finnish Galician Greek Hungarian Norwegian Polish Portuguese Romanian Slovak Slovene Swedish	Croatian Estonian Icelandic Irish Latvian Lithuanian <b>Maltese</b> Serbian

12: Text analysis: state of language technology support for 30 European languages

Excellent support	Good support	Moderate support	Fragmentary support	Weak/no support
	English	Czech Dutch French German Hungarian Italian Polish Spanish Swedish	Basque Bulgarian Catalan Croatian Danish Estonian Finnish Galician Greek Norwegian Portuguese Romanian Serbian Slovak Slovene	Icelandic Irish Latvian Lithuanian <b>Maltese</b>

13: Speech and text resources: State of support for 30 European languages



## ABOUT META-NET

**META-NET** is a Network of Excellence partially funded by the European Commission. The network currently consists of 54 research centres in 33 European countries [1]. **META-NET** forges **META**, the Multilingual Europe Technology Alliance, a growing community of language technology professionals and organisations in Europe. **META-NET** fosters the technological foundations for a truly multilingual European information society that:

- makes communication and cooperation possible across languages;
- grants all Europeans equal access to information and knowledge regardless of their language;
- builds upon and advances functionalities of networked information technology.

The network supports a Europe that unites as a single digital market and information space. It stimulates and promotes multilingual technologies for all European languages. These technologies support automatic translation, content production, information processing and knowledge management for a wide variety of subject domains and applications. They also enable intuitive language-based interfaces to technology ranging from household electronics, machinery and vehicles to computers and robots.

Launched on 1 February 2010, **META-NET** has already conducted various activities in its three lines of action **META-VISION**, **META-SHARE** and **META-RESEARCH**.

**META-VISION** fosters a dynamic and influential

stakeholder community that unites around a shared vision and a common strategic research agenda (SRA). The main focus of this activity is to build a coherent and cohesive LT community in Europe by bringing together representatives from highly fragmented and diverse groups of stakeholders. The present White Paper was prepared together with volumes for 29 other languages. The shared technology vision was developed in three sectorial Vision Groups. The **META** Technology Council was established in order to discuss and to prepare the SRA based on the vision in close interaction with the entire LT community.

**META-SHARE** creates an open, distributed facility for exchanging and sharing resources. The peer-to-peer network of repositories will contain language data, tools and web services that are documented with high-quality metadata and organised in standardised categories. The resources can be readily accessed and uniformly searched. The available resources include free, open source materials as well as restricted, commercially available, fee-based items.

**META-RESEARCH** builds bridges to related technology fields. This activity seeks to leverage advances in other fields and to capitalise on innovative research that can benefit language technology. In particular, the action line focuses on conducting leading-edge research in machine translation, collecting data, preparing data sets and organising language resources for evaluation purposes; compiling inventories of tools and methods; and organising workshops and training events for members of the community.

office@meta-net.eu – <http://www.meta-net.eu>

---

## REFERENZI REFERENCES

- [1] Georg Rehm and Hans Uszkoreit. Multilingual Europe: A challenge for language tech. *MultiLingual*, 22(3):51–52, April/May 2011.
- [2] Aljoscha Burchardt, Markus Egg, Kathrin Eichler, Brigitte Krenn, Jörn Kreutel, Annette Leßmöllmann, Georg Rehm, Manfred Stede, Hans Uszkoreit, and Martin Volk. *Die Deutsche Sprache im Digitalen Zeitalter – The German Language in the Digital Age*. META-NET White Paper Series. Georg Rehm and Hans Uszkoreit (Series Editors). Springer, 2012.
- [3] Ray Fabri. Maltese. In C. Delcourt and P. van Sterkenburg, editors, *The Languages of the 25. Revue belge de Philologie et d’Histoire: RBPH*, number 85 (2007) 3, pages 17–28. John Benjamins, Amsterdam-Philadelphia, 2011.
- [4] Aljoscha Burchardt, Georg Rehm, and Felix Sasaki. The Future European Multilingual Information Society – Vision Paper for a Strategic Research Agenda, 2011.  
<http://www.meta-net.eu/vision/reports/meta-net-vision-paper.pdf>.
- [5] Directorate-General Information Society & Media of the European Commission. User Language Preferences Online, 2011. [http://ec.europa.eu/public\\_opinion/flash/fl\\_313\\_en.pdf](http://ec.europa.eu/public_opinion/flash/fl_313_en.pdf).
- [6] European Commission. Multilingualism: an Asset for Europe and a Shared Commitment, 2008.  
[http://ec.europa.eu/languages/pdf/comm2008\\_en.pdf](http://ec.europa.eu/languages/pdf/comm2008_en.pdf).
- [7] Directorate-General of the UNESCO. Intersectoral Mid-term Strategy on Languages and Multilingualism, 2007. <http://unesdoc.unesco.org/images/0015/001503/150335e.pdf>.
- [8] Directorate-General for Translation of the European Commission. Size of the Language Industry in the EU, 2009. <http://ec.europa.eu/dgs/translation/publications/studies>.
- [9] Arne Ambros. *Bongornu, kif int? Einführung in die maltesische Sprache (Bongornu, kif int? Introduction to the Maltese Language)*. Reichert, Wiesbaden, 1998.
- [10] Roderick Bovingdon. *The Maltese language of Australia: Maltraljan: a lexical compilation with linguistic notations and a social, political and historical background*. LINCOM Europa, Munich, 2001.
- [11] Albert J. Borg and Marie Azzopardi-Alexander. *Maltese*. Routledge, London and New York, 1997.

- [12] Ray Fabri. *Kongruenz und die Grammatik des Maltesischen (Agreement and the Grammar of Maltese)*. Niemeyer, Tübingen, 1993.
- [13] Manwel Mifsud. *Loan verbs in Maltese: a descriptive and comparative study*. Brill, Leiden etc, 1995.
- [14] Ray Fabri. The Tense and Aspect System of Maltese. In Rolf Thieroff, editor, *Tempussysteme in europäischen Sprachen II*, pages 327–343. Niemeyer, Tübingen, 1995.
- [15] Kare Ebertn. Aspect in Maltese. In Östen Dahl, editor, *Tense and Aspect in the Languages of Europe*, pages 753–788. Mouton de Gruyter, Berlin, 2000.
- [16] Il-Kunsill Nazzjonali tal-Ilsien Malti, editor. *Deċiżjonijiet 1 tal-Kunsill Nazzjonali tal-Ilsien Malti dwar il-Varjanti Ortografici (Decisions 1 of the National Council for the Maltese Language about the Orthographic Variants)*. Il-Kunsill Nazzjonali tal-Ilsien Malti, Il-Furjana, Malta, 2008.
- [17] Il-Kunsill Nazzjonali tal-Ilsien Malti, editor. *Innaqqsu l-Incertezzi (Let's talk about uncertainties)*. Il-Kunsill Nazzjonali tal-Ilsien Malti, 2008. [http://www.kunsilltalmalti.gov.mt/filebank/documents/il-ktieb\\_finali\\_sa%20Nov%2008.pdf](http://www.kunsilltalmalti.gov.mt/filebank/documents/il-ktieb_finali_sa%20Nov%2008.pdf).
- [18] Reinhold Kontzi. *Sprachkontakt im Mittelmeer: Gesammelte Aufsätze zum Maltesischen (Language Contact in the Mediterranean: Collected Essays on the Maltese Language)*. Narr, Tübingen, 2005.
- [19] Il-Kunsill Nazzjonali tal-Ilsien Malti. <http://www.kunsilltalmalti.gov.mt/>.
- [20] Akkademja tal-Malti. <http://www.akkademjatalmalti.com/>.
- [21] Antoinette Camilleri. *Bilingualism in Education: The Maltese Experience*. Brill, New York, 1995.
- [22] Douglas Biber. *Variation across speech and writing*. Cambridge University Press, Cambridge, 1991.
- [23] Joseph M. Brincat. *Maltese and other Languages: a Linguistic History of Malta*. Midsea Books, Malta, 2011.
- [24] Għaqda Internazzjonali tal-Lingwistika Maltija. <http://www.fb10.uni-bremen.de/ghilm/default.aspx>.
- [25] National Statistics Office of Malta. [http://www.nso.gov.mt/statdoc/document\\_file.aspx?id=2651](http://www.nso.gov.mt/statdoc/document_file.aspx?id=2651).
- [26] EUROPA Press Releases. Digital Agenda: more than half EU Internet surfers use foreign language when online. <http://europa.eu/rapid/pressReleasesAction.do?reference=IP/11/556&format=HTML&aged=0&language=EN&guiLanguage=en>.
- [27] The Times of Malta. Maltese language hardly used on the internet, 2011. 16/05/2011.
- [28] Ray Fabri. The Language of Young People and Language Change in Maltese. In Sandro Caruana, Ray Fabri, and Thomas Stolz, editors, *Variation and Change: The dynamics of Maltese in space, time, and society*. Akademie Verlag, to appear.

- [29] Adam Kilgarriff and Gregory Grefenstette. Introduction to the special issue on the web as corpus. *Computational Linguistics*, (29):333–347, 2003.
- [30] Government of Malta. <http://www.gov.mt/>.
- [31] Net Television. <http://www.nettv.com.mt/>.
- [32] One Productions Limited. <http://www.one.com.mt/>.
- [33] RTK Limited. <http://www.rtk.org.mt/>.
- [34] Public Broadcasting Services Limited. <http://www.pbs.com.mt/>.
- [35] Radio 101. <http://www.radio101.com.mt/>.
- [36] EUR-Lex: Access to European Union. <http://eur-lex.europa.eu>.
- [37] The JRC-Acquis Multilingual Parallel Corpus. <http://langtech.jrc.it/JRC-Acquis.html>.
- [38] Maltese Language Resource Server . <http://mlrs.research.um.edu.mt/>.
- [39] Kai-Uwe Carstensen, Christian Ebert, Cornelia Ebert, Susanne Jekat, Hagen Langer, and Ralf Klabunde, editors. *Computerlinguistik und Sprachtechnologie: Eine Einführung (Computational Linguistics and Language Technology: An Introduction)*. Spektrum Akademischer Verlag, 2009.
- [40] Daniel Jurafsky and James H. Martin. *Speech and Language Processing*. Prentice Hall, 2nd edition, 2009.
- [41] Christopher D. Manning and Hinrich Schütze. *Foundations of Statistical Natural Language Processing*. MIT Press, 1999.
- [42] Language Technology World (LT World). <http://www.lt-world.org>.
- [43] Ronald Cole, Joseph Mariani, Hans Uszkoreit, Giovanni Battista Varile, Annie Zaenen, and Antonio Zampolli, editors. *Survey of the State of the Art in Human Language Technology*. Cambridge University Press, 1998.
- [44] Jerrold H. Zar. Candidate for a Pullet Surprise. *Journal of Irreproducible Results*, page 13, 1994.
- [45] Gordon Mangion. Spelling Correction for Maltese. Technical report, Dept. Computer Science and Artificial Intelligence, University of Malta, Msida MSD2080, 1999.
- [46] Ruth Mizzi. The Development of a Statistical Spell Checker for Maltese. Technical report, Dept. Computer Science and Artificial Intelligence, University of Malta, 2000.
- [47] Malta Linux User Group Spellchecker. <http://linux.org.mt/spellcheck>.
- [48] Lydia Sciriha. *The Maltese Interactive Picture Dictionary*. Protea Textware, Melbourne, 1997. ISBN: 978-09-587-3300-7.

- [49] Spiegel Online. Google zieht weiter davon (Google is still leaving everybody behind), 2009. <http://www.spiegel.de/netzwelt/web/0,1518,619398,00.html>.
- [50] Juan Carlos Perez. Google Rolls out Semantic Search Capabilities, 2009. [http://www.pcworld.com/businesscenter/article/161869/google\\_rolls\\_out\\_semantic\\_search\\_capabilities.html](http://www.pcworld.com/businesscenter/article/161869/google_rolls_out_semantic_search_capabilities.html).
- [51] 16 Malta Search engines. <http://www.philb.com/csc/malta.htm>.
- [52] Charonite. <http://www.charonite.com>.
- [53] Language Technology for eLearning. <http://www.let.uu.nl/lr4el/>.
- [54] Paul Micallef. *A Text to Speech Synthesis System for Maltese*. PhD thesis, University of Surrey, 1997.
- [55] Paulseph-John Farrugia. Text to Speech Technologies for Mobile Telephony Services. Technical report, Dept. CSAI, University of Malta, 2005.
- [56] Ian Buhagiar and Paul Micallef. Web Based Maltese Language Text to Speech Synthesiser. In *Proceedings of WICT08*. University of Malta, 2008.
- [57] Mark Borg, Keith Bugeja, Colin Vella, Gordon Mangion, and Carmel Gafà. Preparation of a free-running text corpus for Maltese concatenative speech synthesis. GHILM 3rd Conference on Maltese Linguistics, Valletta, 2011.
- [58] Sinclair Calleja. Speech Synthesis. Master's thesis, Dept. CSAI, University of Malta, 2002.
- [59] Anthony Psaila. Speech Annotation using Hidden Markov Models. Master's thesis, Dept. Computer Communications Engineering, University of Malta, 2008.
- [60] Alexandra Vella and Paulseph-John Farrugia. MalToBI – Building an Annotated Corpus of Spoken Maltese. In *Online Proceedings of the 3rd International Conference on Speech Prosody*, Dresden, 2006. ISCA Special Interest Group on Speech Prosody (SProSIG). [http://aune.lpl.univ-aix.fr/~sprosig/sp2006/contents/papers/PS5-31\\_0136.pdf](http://aune.lpl.univ-aix.fr/~sprosig/sp2006/contents/papers/PS5-31_0136.pdf).
- [61] Robert Farrugia. SAMILS – A Semi-Automatic Machine Indexing for Legal Systems. Technical report, Dept. CSAI, University of Malta, 2000.
- [62] Jo-Ann Bajada. Investigation of Translations Equivalences from Parallel Texts. Technical report, Dept. CSAI, University of Malta, 2004.
- [63] Jo-Ann Bajada. Phrase Extraction for Machine Translation. Master's thesis, Dept. CSAI, University of Malta, 2009.
- [64] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. BLEU: A Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting of ACL*, Philadelphia, PA, 2002.

- [65] Philipp Koehn, Alexandra Birch, and Ralf Steinberger. 462 Machine Translation Systems for Europe. In *Proceedings of MT Summit XII*, 2009.
- [66] David Galea. A System for the Analysis of Maltese Verbs. Technical report, Dept. CSAI, University of Malta, 1999.
- [67] Paulseph-John Farrugia. An Automatic Translation System for Maltese/English. Technical report, Dept. CSAI, University of Malta, 1999.
- [68] Duncan Attard. A Lexicon Server Toolkit for Maltese. Technical report, Dept. CSAI, University of Malta, 2005.
- [69] Alex Farrugia. MULTIMORPH: A Computational Analysis of the Maltese Broken Plural. Technical report, Dept. CSAI, University of Malta, 2008.
- [70] Christopher Farrugia. A Portal for Acquisition, Representation and Presentation of Current Affairs in Malta. Technical report, Dept. CSAI, University of Malta, 2009.
- [71] Gilbert Vella. Automatic Summarization of Legal Documents. Technical report, Dept. CSAI, University of Malta, 2010.
- [72] Michael Rosner, Joe Caruana, and Ray Fabri. Maltilex: A computational lexicon for Maltese. In Michael Rosner, editor, *Computational Approaches to Semitic Languages: Proceedings of the Workshop held at COLING-ACL98, Université de Montréal, Canada*, page 97–105, 1998.
- [73] Joseph Aquilina. *Maltese-English Dictionary Vol. I, A–L*. Midsea Books, Malta:Valletta, 1987.
- [74] Joseph Aquilina. *Maltese-English Dictionary Vol. II, M–Z*. Midsea Books, Malta:Valletta, 1990.
- [75] Angelo Dalli. Computational Lexicon for Maltese. Master’s thesis, Dept. CSAI, University of Malta, 2001.
- [76] Michael Rosner. Electronic language resources for Maltese. In Bernard Comrie, Ray Fabri, Manwel Mifsud, Thomas Stolz, and Martine Vanhove, editors, *Introducing Maltese Linguistics. Proceedings of the 1st International conference on Maltese Linguistics (Bremen/Germany, October, 2007)*, Studies in Language Companion Series, pages 251–276, Amsterdam; Philadelphia, 2009. John Benjamins.
- [77] Roberta Camilleri. Speech Annotation System. Master’s thesis, Dept. Computer Communications Engineering, University of Malta, 2010.
- [78] Alexandra Vella. On Maltese Prosody. In Bernard Comrie, Ray Fabri, Manwel Mifsud, Thomas Stolz, and Martine Vanhove, editors, *Introducing Maltese Linguistics. Proceedings of the 1st International conference on Maltese Linguistics (Bremen/Germany, October, 2007)*, Studies in Language Companion Series, pages 47–68, Amsterdam; Philadelphia, 2009. John Benjamins.



- [79] Adam Ussishkin, Jerid Francom, and Dainon Woudstra. Creating a Web-based Lexical Corpus and Information-extraction Tools for the Semitic Language Maltese. In *Proceedings of the SALT MIL Workshop, Donostia*, September 2009. ISBN: 978-84-692-4940-6.
- [80] Bernard Comrie, Ray Fabri, Manwel Mifsud, Thomas Stolz, and Martine Vanhove, editors. *Introducing Maltese Linguistics. Proceedings of the 1st International conference on Maltese Linguistics (Bremen/Germany, October, 2007)*, Studies in Language Companion Series, Amsterdam ; Philadelphia, 2009. John Benjamins.
- [81] International Association of Maltese Linguistics (GHILM), editor. *ILSIENNA/Our Language*, Bochum, 2009. Universitätsverlag Brockmeyer.
- [82] Michael Rosner, Ray Fabri, Duncan Attard, and Albert Gatt. Maltese language resource server. In *Proceedings of CSAW06*, page 90–98, University of Malta, November 2006.
- [83] Michael Rosner, editor. *Computational Approaches to Semitic Languages: Proceedings of the Workshop held at COLING-ACL98, Université de Montréal, Canada, Quebec*, 1998. Université de Montréal.



## MEMBRI TA' META-NET META-NET MEMBERS

Awstrija	Austria	Zentrum für Translationswissenschaft, Universität Wien: Gerhard Budin
Belġju	Belgium	Computational Linguistics and Psycholinguistics Research Centre, University of Antwerp: Walter Daelemans
		Centre for Processing Speech and Images, University of Leuven: Dirk van Compernelle
Bulgarija	Bulgaria	Institute for Bulgarian Language, Bulgarian Academy of Sciences: Svetla Koeva
Ĉekja	Czech Republic	Institute of Formal and Applied Linguistics, Charles University in Prague: Jan Hajič
Ċipru	Cyprus	Language Centre, School of Humanities: Jack Burston
Danimarka	Denmark	Centre for Language Technology, University of Copenhagen:
		Bolette Sandford Pedersen, Bente Maegaard
Estonja	Estonia	Institute of Computer Science, University of Tartu: Tiit Roosmaa, Kadri Vider
Finlandja	Finland	Computational Cognitive Systems Research Group, Aalto University: Timo Honkela
		Department of Modern Languages, University of Helsinki: Kimmo Koskenniemi, Krister Lindén
Franza	France	Centre National de la Recherche Scientifique, Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur and Institute for Multilingual and Multimedia Information: Joseph Mariani
		Evaluations and Language Resources Distribution Agency: Khalid Choukri
Ġermanja	Germany	Language Technology Lab, DFKI: Hans Uszkoreit, Georg Rehm
		Human Language Technology and Pattern Recognition, RWTH Aachen University: Hermann Ney
		Department of Computational Linguistics, Saarland University: Manfred Pinkal
Gran Brittanja	UK	School of Computer Science, University of Manchester: Sophia Ananiadou
		Institute for Language, Cognition and Computation, Center for Speech Technology Research, University of Edinburgh: Steve Renals
		Research Institute of Informatics and Language Processing, University of Wolverhampton: Ruslan Mitkov
Greċja	Greece	R.C. "Athena", Institute for Language and Speech Processing: Stelios Piperidis
Irlanda	Ireland	School of Computing, Dublin City University: Josef van Genabith
Islanda	Iceland	School of Humanities, University of Iceland: Eiríkur Rögnvaldsson
Isvezja	Sweden	Department of Swedish, University of Gothenburg: Lars Borin

Isvizzera	Switzerland	Idiap Research Institute: Hervé Bourlard
Italja	Italy	Consiglio Nazionale delle Ricerche, Istituto di Linguistica Computazionale “Antonio Zampolli”: Nicoletta Calzolari  Human Language Technology Research Unit, Fondazione Bruno Kessler: Bernardo Magnini
Kroazja	Croatia	Institute of Linguistics, Faculty of Humanities and Social Science, University of Zagreb: Marko Tadić
Latvja	Latvia	Tilde: Andrejs Vasiljevs  Institute of Mathematics and Computer Science, University of Latvia: Inguna Skadiņa
Litvanja	Lithuania	Institute of the Lithuanian Language: Jolanta Zabarskaitė
Lussemburgu	Luxembourg	Arax Ltd.: Vartkes Goetcherian
Malta	Malta	Department Intelligent Computer Systems, University of Malta: Mike Rosner
Norvegja	Norway	Department of Linguistic, Literary and Aesthetic Studies, University of Bergen: Koenraad De Smedt  Department of Informatics, Language Technology Group, University of Oslo: Stephan Oepen
Olanda	Netherlands	Utrecht Institute of Linguistics, Utrecht University: Jan Odijk  Computational Linguistics, University of Groningen: Gertjan van Noord
Polonja	Poland	Institute of Computer Science, Polish Academy of Sciences: Adam Przepiórkowski, Maciej Ogrodniczuk  University of Łódź: Barbara Lewandowska-Tomaszczyk, Piotr Pęzik  Department of Computer Linguistics and Artificial Intelligence, Adam Mickiewicz University: Zygmunt Vetulani
Portugall	Portugal	University of Lisbon: António Branco, Amália Mendes  Spoken Language Systems Laboratory, Institute for Systems Engineering and Computers: Isabel Trancoso
Rumanija	Romania	Research Institute for Artificial Intelligence, Romanian Academy of Sciences: Dan Tufiş  Faculty of Computer Science, University Alexandru Ioan Cuza of Iaşi: Dan Cristea
Serbja	Serbia	University of Belgrade, Faculty of Mathematics: Duško Vitas, Cvetana Krstev, Ivan Obradović  Pupin Institute: Sanja Vranes
Slovakkja	Slovakia	Ľudovít Štúr Institute of Linguistics, Slovak Academy of Sciences: Radovan Garabík
Slovenja	Slovenia	Jožef Stefan Institute: Marko Grobelnik
Spanja	Spain	Barcelona Media: Toni Badia, Maite Melero

Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra: Núria Bel

Aholab Signal Processing Laboratory, University of the Basque Country:  
Inma Hernaez Rioja

Center for Language and Speech Technologies and Applications, Universitat Politècnica de Catalunya: Asunción Moreno

Department of Signal Processing and Communications, University of Vigo:  
Carmen García Mateo

Ungerija

Hungary

Research Institute for Linguistics, Hungarian Academy of Sciences: Tamás Váradi

Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics: Géza Németh, Gábor Olasz



Kważi 100 esperti tat-teknoloġija lingwistika – rappreżentanti tal-pajjiżi u tal-lingwi li huma rappreżentati f’META-NET – ddiskutew u ffinalizzaw ir-riżultati ewlenin u l-messaġġi ta’ Serje ta’ White Papers tal-META-NET f’laqgħa f’Berlin, fil-Ġermanja, dwar Ottubru 21/22, 2011. – About 100 language technology experts – representatives of the countries and languages represented in META-NET – discussed and finalised the key results and messages of the White Paper Series at a META-NET meeting in Berlin, Germany, on October 21/22, 2011.





---

# IS-SERJE TA' WHITE PAPERS TA' META-NET

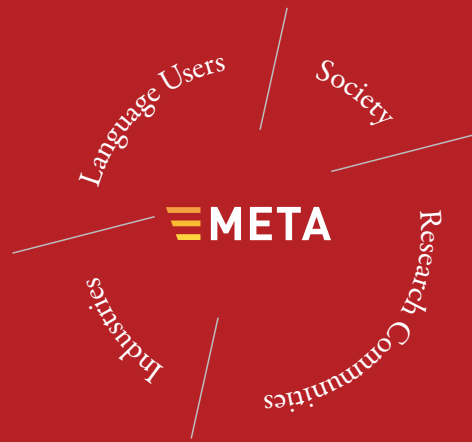
# THE META-NET WHITE PAPER SERIES

---

Bask	Basque	euskara
Bulgaru	Bulgarian	български
Ĉek	Czech	čeština
Daniž	Danish	dansk
Estonjan	Estonian	eesti
Finlandiž	Finnish	suomi
Franciž	French	français
Ġermaniž	German	Deutsch
Galizjan	Galician	galego
Grieg	Greek	ελληνικά
Ingliž	English	English
Irlandiž	Irish	Gaeilge
Islandiž	Icelandic	íslenska
Katalan	Catalan	català
Kroat	Croatian	hrvatski
Latvjan	Latvian	latviešu valoda
Litwan	Lithuanian	lietuvių kalba
Malti	Maltese	Malti
Norvegiž Bokmål	Norwegian Bokmål	bokmål
Norvegiž Nynorsk	Norwegian Nynorsk	nynorsk
Olandiž	Dutch	Nederlands
Pollakk	Polish	polski
Portugiž	Portuguese	português
Rumen	Romanian	română
Serb	Serbian	српски
Slovakk	Slovak	slovenčina
Sloven	Slovene	slovenščina
Spanjol	Spanish	español
Svediž	Swedish	svenska
Taljan	Italian	italiano
Ungeriž	Hungarian	magyar

---





“The Maltese language traces its roots to the arrival of the Arabs in Malta over 1000 years ago. Originally Siculo-Arabic, Maltese later followed its own path to linguistics development, evolving and adapting itself in response to political events in our islands that left their mark not only on the language but also on the Maltese People. In our language, we can perceive a reflection of our history and a confirmation of the tenacity of our People to hold on to their own identity in spite of having been ruled by different foreign powers. Today, Maltese is more alive than ever before and will again adapt itself, this time in response to the technological developments of the present. In this way, it will guarantee its own survival in the future.”

– His Excellency Dr. George Abela (President of Malta)

“Tkun verament hasra jekk ma nisfruttawx l-iżviluppi fit-teknoloġija biex napplikawhom għall-qasam lingwistiku. Is-sinerġija bejn dawn iż-żewġ xjenzi għandha tiapoġġa għas-servizz tal-bniedem biex jiffaċilitalna ħajjitna u jgħinha nkissru l-barrieri f’ dinja globalizzata. Għalhekk l-appoġġ tat-teknoloġija lingwistika għall-Malti għandu jservi biex ilsienna jibqa’ jkun kultivat, użat u mpoġġi fuq l-istess livell ta’ lingwi oħra.”

– L-Onor. Dolores Cristina (Ministru tal-Edukazzjoni u x-Xogħol)

“It would be really unfortunate if we did not exploit the developments in technology to apply them to the linguistic field. The synergy between these two sciences has to be brought to the service of the people so that it makes our life easier and helps to break the barriers in a globalised world. Thus the technology support for the Maltese language should serve our language to be continuously cultivated, used and placed on the same level as other languages.”

– The Hon. Dolores Cristina (Minister for Education and Employment)

“Lejn mappa tal-Malti fl-era diġitali”

– Prof. Ray Fabri (Il-Kunsill Nazzjonali tal-Ilsien Malti)

“Mapping out the territory of Maltese in a digital age”

– Prof. Ray Fabri (Council for the Maltese Language)