An Update and Extension of the META-NET Study META-NET "Europe's Languages in the Digital Age"

Georg Rehm & Hans Uszkoreit (META-NET, DFKI GmbH, Berlin, Germany), Ido Dagan (META-NET, Bar-Ilan University, Tel Aviv, Israel), Vartkes Goetcherian (META-NET, Arax Ltd., Luxembourg), Mehmet Ugur Dogan & Coskun Mermer (META-NET, Tübitak Bilgem, Gebze, Turkey), Tamás Varadi (EFNIL, META-NET, Hungarian Academy of Sciences, Budapest, Hungary), Sabine Kirchmeier-Andersen (EFNIL, META-NET, Danish Language Council, Copenhagen, Denmark), Gerhard Stickel (EFNIL, Institut für Deutsche Sprache, Mannheim, Germany), Meirion Prys Jones (NPLD, Network to Promote Linguistic Diversity, Cardiff, Wales), Stefan Oeter (Council of Europe, Committee of Experts, University of Hamburg, Hamburg, Germany), Sigve Gramstad (Council of Europe, Committee of Experts, Bergen, Norway)

This poster extends and updates the cross-language comparison of LT support for 30 European languages as published in the META-NET Language White Paper Series. The updated confirms the original results and paints an alarming picture: it demonstrates that there are even more dramatic differences in LT support between the European languages.

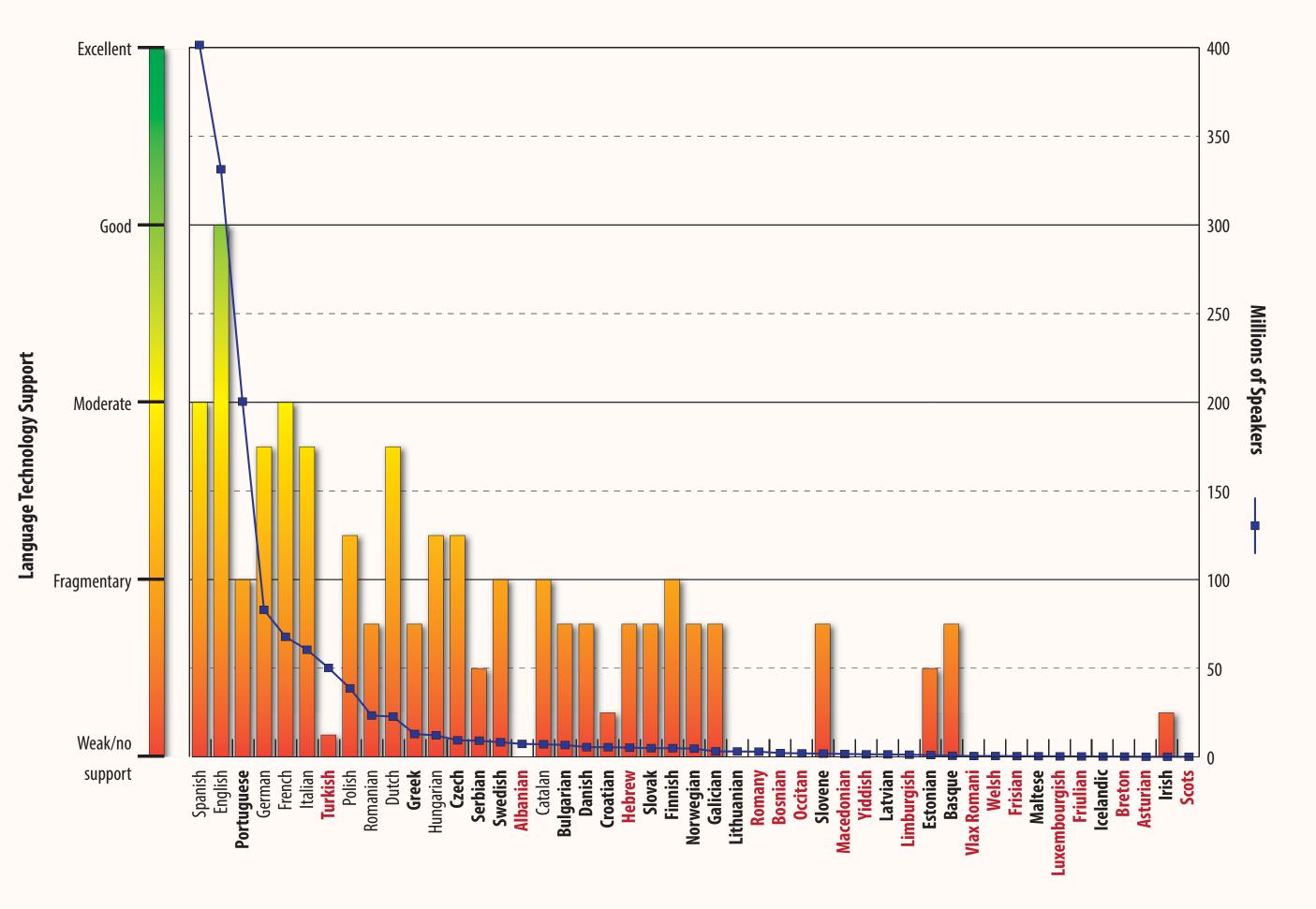
Introduction: META-NET

- 60 leading research centres in 34 European countries dedicated to the technological foundations of a multilingual European information society.
- Multilingual Europe Technology Alliance (META), more than 760 organisations and experts representing multiple stakeholders as of May 2014.
- The goal is monolingual, crosslingual and multilingual technology support for all European languages.

The Language White Paper Series "Europe's Languages in the Digital Age"

- Language White Paper Series covers 31 languages in 31 volumes, available online at http://www.meta-net.eu/whitepapers.
- Reports on the state of our languages in the digital age and the level of support through language technology.
- Press release "At least 21 European Languages in Danger of Digital Extinction," circulated on the European Day of Languages 2012 (Sept. 26).

The Set of Languages



- Original set comprised 30 languages; a few languages represented by META-NET members could not be addressed due to lack of funding.
- Multiple regional and minority languages could not be represented due to focus on official EU and national languages of the four META-NET projects' partners.
- This extended set now finally contains all languages represented by META-NET, EFNIL and NPLD as well as many of the languages monitored by the Council of Europe's Committe of Experts on Regional and Minority Languages.
- We excluded languages with less than 100,000 speakers (according to Ethnologue) and also all languages which did not originate in Europe.

Conclusions

- This updated comparison confirms the original results and paints an alarming picture: there are even more dramatic differences in LT support between the European languages, i.e., the technological gap keeps widening.
- We should focus even more on fostering technology development for smaller and/or less-resourced languages and also on digital language preservation.
- Research and technology transfer between the languages along with increased collaboration across languages must receive more attention. Regional, national and international organisations as well as funding agencies should team up!
- META-NET suggests setting up a shared programme to develop resources and technologies for all European languages (cf. META-NET SRA).

Overall language technology support levels and number of speakers for 47 European languages. Languages new to the crosslanguage comparison are in red. Languages in bold face have no weak or no support for machine translation.

Support for Speech Processing

	Support for Flatime franslation								Supportion Specent rocessing								
Excellent	Good	Moderate	Fragmentary	Weak/none			Excellent	Good	Good Moderate			Weak/none					
Englis	ish	French Spanish	Catalan Dutch German Hungarian Italian Polish Romanian	Albanian Asturian Basque Bosnian Breton Bulgarian Croatian Croatian Czech Danish Estonian Finnish Frisian Friulian Galician Greek	Icelandic Irish Latvian Limburgish Lithuanian Luxembourgish Macedonian Maltese Norwegian Occitan Portuguese Romany Scots Serbian Slovak	Swedish Turkish Vlax Romani Welsh Yiddish			English	Czech Dutch Finnish French German Italian Portuguese Spanish	Basque Bulgarian Catalan Danish Estonian Galician Greek Hungarian Irish Norwegian Polish Serbian Slovak Slovene Swedish	Albanian Asturian Bosnian Breton Croatian Frisian Friulian Friulian Hebrew Icelandic Latvian Limburgish Lithuanian Luxembourgish Macedonian Maltese	Romanian Romany Scots Vlax Romani Welsh Yiddish				

Support for Machine Translation

		Hebrew Slovene			Iurkish	Uccitan

Support for Text Analytics

Support for Speech and Text Resources

Excellent	Good	Moderate	Fragmentary	Weak/none		Weak/none		Excellent	Good	Moderate	Fragmentary		Weak/none
Engl		Dutch French German Italian Spanish Hebrew	Basque Bulgarian Catalan Czech Danish Finnish Galician Greek Hungarian Norwegian Polish Portuguese Romanian Slovak Slovene Swedish	Albanian Asturian Bosnian Breton Croatian Estonian Estonian Frisian Friulian Icelandic Irish Latvian Limburgish Lithuanian Luxembourgish Macedonian Maltese Occitan	Romany Scots Serbian Turkish Vlax Romani Welsh Yiddish		English	Czech Dutch French German Hungarian Italian Polish Spanish Swedish	Basque Bulgarian Catalan Croatian Danish Estonian Finnish Galician Greek Hebrew Norwegian Portuguese Romanian Serbian Slovak Slovene	Albanian Asturian Bosnian Breton Frisian Friulian Icelandic Irish Latvian Limburgish Lithuanian Luxembourgish Macedonian Maltese Occitan Romany Scots	Turkish Vlax Romani Welsh Jiddish		

Set of Languages

						_	_		
No. of Speakers	White Paper?			No. of Speakers	White Paper?			No. of Speakers	Wh Pap
n 7,436,990		1	Friulian	300,000			Occitan	2,048,310	
n 110,000			Galician	3,185,000	Yes		Polish	39,042,570	Ye
e 657,872	Yes		German	83,812,810	Yes		Portuguese	202,468,100	Ye
n 2,216,000		1	Greek	13,068,650	Yes		Romanian	23,623,890	Ye
n 225,000		1	Hebrew	5,302,770			Romany	3,017,920	
n 6,795,150	Yes		Hungarian	12,319,330	Yes		Scots	100,000	
ז 7,220,420	Yes		lcelandic	243,840	Yes		Serbian	9,262,890	Ye
n 5,533,890	Yes		Irish	106,210	Yes		Slovak	5,007,650	Ye
n 9,469,340	Yes		Italian	61,068,677	Yes		Slovene	1,906,630	Ye
n 5,592,490	Yes		Latvian	1,472,650	Yes		Spanish	405,638,110	Ye
n 22,984,690	Yes		Limburgish	1,300,000			Swedish	8,381,829	Ye
n 334,800,758	Yes		Lithuanian	3,130,970	Yes		Turkish	50,733,420	
n 1,078,400	Yes		Luxembourgish	320,710			Vlax Romani	540,780	
n 4,994,490	Yes		Macedonian	1,710,670			Welsh	536,890	Ye
n 68,458,600	Yes		Maltese	429,000	Yes		Yiddish	1,510,430	
467,000			Norwegian	4,741,780	Yes		·		
r r r r r r r	Speakers n 7,436,990 n 110,000 e 657,872 n 2,216,000 n 225,000 n 7,220,420 n 7,220,420 n 5,533,890 h 9,469,340 h 5,592,490 h 22,984,690 h 334,800,758 n 1,078,400 h 4,994,490 h 68,458,600	Speakers Paper? n 7,436,990 n 110,000 e 657,872 p 2,216,000 n 2,216,000 n 2,216,000 n 6,795,150 n 6,795,150 n 7,220,420 Yes 1 n 5,533,890 Yes Yes h 9,469,340 Yes Yes h 5,592,490 Yes Yes h 334,800,758 Yes Yes h 4,994,490 Yes Yes h 68,458,600	Speakers Paper? n 7,436,990 n 110,000 e 657,872 Yes 1 n 2,216,000 n 2,216,000 n 2,25,000 n 6,795,150 Yes 1 n 7,220,420 Yes 1 n 5,533,890 Yes 1 h 9,469,340 Yes 1 h 5,592,490 Yes 1 h 334,800,758 Yes 1 n 1,078,400 Yes 1 h 4,994,490 Yes 1 h 68,458,600	SpeakersPaper?n7,436,990Friuliann110,000Galiciane657,872Yesf2,216,000Germann2,25,000Hebrewn6,795,150Yesn6,795,150Yesn7,220,420Yesn5,533,890Yesh9,469,340Yesh22,984,690Yesh334,800,758Yesh1,078,400Yesh4,994,490Yesh68,458,600Yesh68,458,600Yesh68,458,600Yes	SpeakersPaper?Speakersn7,436,990Friulian300,000n110,000Galician3,185,000e657,872YesGerman83,812,810n2,216,000Greek13,068,650n225,000Hebrew5,302,770n6,795,150YesHungarian12,319,330n7,220,420YesIcelandic243,840n5,533,890YesItalian61,068,677h9,469,340YesLatvian1,472,650h22,984,690YesLithuanian3,130,970h1,078,400YesLithuanian3,20,710h4,994,490YesMaltese429,000	SpeakersPaper?SpeakersPaper?n7,436,990Friulian300,000n110,000Galician3,185,000Yese657,872YesGerman83,812,810Yesn2,216,000Greek13,068,650Yesn225,000Hebrew5,302,770n6,795,150YesHungarian12,319,330Yesn7,220,420YesIcelandic243,840Yesn5,533,890YesItalian61,068,677Yesh9,469,340YesLatvian1,472,650Yesh22,984,690YesLimburgish1,300,000h1,078,400YesMacedonian1,710,670Yesh68,458,600YesMaltese429,000Yes	SpeakersPaper?SpeakersPaper?n7,436,990Image: SpeakersSpeakersPaper?n110,000Image: Speakers300,000Image: Speakersn110,000Image: SpeakersSpeakersSpeakerse657,872YesGalician3,185,000Yesi2,216,000Image: SpeakersImage: SpeakersYesi2,216,000Image: SpeakersImage: SpeakersYesi2,216,000Image: SpeakersImage: SpeakersYesi2,216,000Image: SpeakersImage: SpeakersYesi2,216,000Image: SpeakersImage: SpeakersYesi2,216,000Image: SpeakersImage: SpeakersYesi6,795,150YesImage: SpeakersYesi6,795,150YesIcelandic243,840Yesi5,533,890YesItalian61,068,677Yesi5,592,490YesItalian61,068,677Yesi334,800,758YesItimburgish1,300,000Image: Speakersi334,800,758YesItimburgish320,710Image: Speakersi1,078,400YesImage: SpeakersImage: SpeakersImage: Speakersi1,078,400YesImage: SpeakersImage: SpeakersImage: Speakersi1,078,400YesImage: SpeakersImage: SpeakersImage: Speakersi1,078,400Yes<	SpeakersPaper?SpeakersPaper?n7,436,990IFriulian300,000IOccitann110,000IGalician3,185,000YesPolishe657,872YesGerman83,812,810YesPortuguesen2,216,000IIGerek13,068,650YesRomaniann225,000IHebrew5,302,770RomanyRomanyn6,795,150YesIcelandic243,840YesSerbiann5,533,890YesItalian61,068,677YesSlovakh9,469,340YesLatvian1,472,650YesSlovensh22,984,690YesLimburgish1,300,000ISwedishh1,078,400YesLithuanian3,130,970YesTurkishh4,994,490YesMacedonian1,710,670WelshWelsh	SpeakersPaper?SpeakersPaper?SpeakersPaper?n7,436,990IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII<

		No. of Speakers	White Paper?		No. of Speakers	White Paper?		No. of Speakers	Wh Pap
Alba	nian	7,436,990		Friulian	300,000		Occitan	2,048,310	
Astı	ırian	110,000		Galician	3,185,000	Yes	Polish	39,042,570	Ye
Ba	isque	657,872	Yes	German	83,812,810	Yes	Portuguese	202,468,100	Ye
Bos	nian	2,216,000		Greek	13,068,650	Yes	Romanian	23,623,890	Ye
Br	eton	225,000		Hebrew	5,302,770		Romany	3,017,920	
Bulg	arian	6,795,150	Yes	Hungarian	12,319,330	Yes	Scots	100,000	
Ca	talan	7,220,420	Yes	lcelandic	243,840	Yes	Serbian	9,262,890	Ye
Cro	atian	5,533,890	Yes	Irish	106,210	Yes	Slovak	5,007,650	Ye
(Czech	9,469,340	Yes	Italian	61,068,677	Yes	Slovene	1,906,630	Ye
D	anish	5,592,490	Yes	Latvian	1,472,650	Yes	Spanish	405,638,110	Ye
[Dutch	22,984,690	Yes	Limburgish	1,300,000		Swedish	8,381,829	Ye
Er	nglish	334,800,758	Yes	Lithuanian	3,130,970	Yes	Turkish	50,733,420	
Este	onian	1,078,400	Yes	Luxembourgish	320,710		Vlax Romani	540,780	
Fi	nnish	4,994,490	Yes	Macedonian	1,710,670		Welsh	536,890	Ye
F	rench	68,458,600	Yes	Maltese	429,000	Yes	Yiddish	1,510,430	
Fr	isian	467,000		Norwegian	4,741,780	Yes			

	No. of Speakers	White Paper?		No. of Speakers	White Paper?		No. of Speakers	White Paper?
Albanian	7,436,990		Friulian	300,000		Occitan	2,048,310	
Asturian	110,000		Galician	3,185,000	Yes	Polish	39,042,570	Yes
Basque	657,872	Yes	German	83,812,810	Yes	Portuguese	202,468,100	Yes
Bosnian	2,216,000		Greek	13,068,650	Yes	Romanian	23,623,890	Yes
Breton	225,000		Hebrew	5,302,770		Romany	3,017,920	
Bulgarian	6,795,150	Yes	Hungarian	12,319,330	Yes	Scots	100,000	
Catalan	7,220,420	Yes	lcelandic	243,840	Yes	Serbian	9,262,890	Yes
Croatian	5,533,890	Yes	lrish	106,210	Yes	Slovak	5,007,650	Yes
Czech	9,469,340	Yes	Italian	61,068,677	Yes	Slovene	1,906,630	Yes
Danish	5,592,490	Yes	Latvian	1,472,650	Yes	Spanish	405,638,110	Yes
Dutch	22,984,690	Yes	Limburgish	1,300,000		Swedish	8,381,829	Yes
English	334,800,758	Yes	Lithuanian	3,130,970	Yes	Turkish	50,733,420	
Estonian	1,078,400	Yes	Luxembourgish	320,710		Vlax Romani	540,780	
Finnish	4,994,490	Yes	Macedonian	1,710,670		Welsh	536,890	Yes
French	68,458,600	Yes	Maltese	429,000	Yes	Yiddish	1,510,430	
Frisian	467,000		Norwegian	4,741,780	Yes			



Strategic Partners

Formed in 2003, the European Federation of National Institutions for Language has institutional members from 30 countries whose role includes monitoring the official language(s) of their country, advising on language use or developing language policy. It provides a forum for these institutions to exchange information about their work and to gather and publish information about language use and policy within the EU. EFNIL encourages the study of the official EU languages and a coordinated approach towards mother-tongue and foreignlanguage learning, as a means of promoting linguistic and cultural diversity within the EU.



The Network to Promote Linguistic Diversity is a pan-European network which works with constitutional, regional and smaller state languages. It has 35 members, 10 of these being either member state or regional governments and the others major NGOs who have a role or are interested in language planning and management. NPLD was established in 2007 and has already asserted itself as the main voice of those linguistic communities that are not the official languages of the EU. NPLD's formation is a reflection of the growing interest in lesser used languages in Europe. Many governments from across the continent have established departments charged with the specific task of revitalizing and promoting the use of these languages. Many of these governments are represented within NPLD.

Languages included in the updated cross-language comparison (new languages in **bold**, number of world-wide speakers according to Ethnologue)

CONSEIL COUNCIL OF EUROPE DE L'EUROPE **Committee of Experts**

Council of Europe Committee of Experts on the Language Charter. The European Charter for Regional or Minority languages is a treaty of the Council of Europe with the purpose to protect and promote the regional and minority languages used in Europe. The two main political goals are the preservation of Europe's cultural heritage and diversity, and the promotion of democracy. The historic cultural and linguistic diversity in Europe is an integral part of European identity, and policies that acknowledge and promote this diversity also facilitate intercultural exchange and the participation in democratic processes. 33 European states have signed the treaty, and 25 states of those have ratified. The Languages Charter is applied to more than 190 regional or minority languages (or language situations), with around 40 million users. Most of these languages are small, less than 50,000 users. Only a handful are spoken by more than a million.

META ⁼ NET	META [⊟] NET	META ^{TE} NET	META	META	META ^{III} NET	META	META ^E NET	META	META	META	META	META	META ⁼ NET	META ^{III} NET	META	META	METAENET
THE GALICIAN O IDIOMA LANGUAGE GALEGO NA IN THE ERA DIXITAL DIGITAL AGE	THE GREEK Η ΕΛΛΗΝΙΚΗ LANGUAGE IN ΓΛΩΣΣΑ ΣΤΗΝ THE DIGITAL ΨΗΦΙΑΚΗ AGE EFIOXH	THE GERMAN DIE DEUTSCHE LANGUAGE IN SPRACHE IM THE DIGITAL DIGITALEN AGE ZEITALTER	THE ITALIAN LA LINGUA LANGUAGE IN ITALIANA THE DIGITAL NELL'ERA AGE DIGITALE	THE IRISH AN LANGUAGE IN GHAEILGE THE DIGITAL SA RÉ AGE DHIGITEACH	THE ICELANDIC ÍSLENSK LANGUAGE IN TUNGA Á THE DIGITAL STAFÆNNI AGE ÖLD	THE A MAGYAR HUNGARIAN NYELV A LANGUAGE IN DIGITÁLIS THE DIGITAL KORBAN AGE	THE SWEDISH SVENSKA LANGUAGE IN SPRÅKET I DEN THE DIGITAL DIGITALA AGE TIDSÅLDERN	THE LATVIAN LATVIEŠU LANGUAGE IN VALODA THE DIGITAL DIGITĀLAJĀ	THE SPANISH LA LENGUA LANGUAGE IN ESPAÑOLA THE DIGITAL EN LA ERA AGE DIGITAL	THE SLOVENE SLOVENSKI LANGUAGE IN JEZIK V THE DIGITAL DIGITALNI AGE DOBI	THE SLOVAK SLOVENSKÝ LANGUAGE IN JAZYK THE DIGITAL V DIGITÁLNOM AGE VEKU	WIX Your Your Market THE SERBIAN СРПСКИ LANGUAGE IN JEЗИК У	WARKE WE CAN AVERAGE THE WELSH Y GYMRAEG LANGUAGE IN YN YR OES	THE NORSK NORWEGIAN I DEN LANGUAGE IN DIGITALE THE DIGITAL TIDSALDEREN AGE	THE NORSK NORWEGIAN IDEN LANGUAGE IN DIGITALE THE DIGITAL TIDSALDEREN AGE	THE MALTESE IL-LINGWA LANGUAGE IN MALTIJA FLERA THE DIGITAL DIGITALI AGE	STRATEGIC RESEARCH AGENDA FOR MULTILINGUAL EUROPE 2020
Genera (See Sa Mada Makawar A Sa Madayan	Nata Canada sana Anatan Sana Kata Saka Awadi Saka Awadi	Agacha Bundhan Kada Sha Kagina kana Kagina kana Kagina kana Kana Kada Kata Sha Kata Sha Kata Sha	Nataré da Cadadi Banda San Kada Saya Manak Spanan	An Alte Angel-Consult faces 19 Stability faces of Stability faces of Stability Angel faces of St	Gene Stangensternen Stagen Hogenster Stage Antegenster	Some forme Indeed Profession Marcel Gan Van Gan	iankan Bandar Manakan Manakan Manakan	ingen Saladi André Marine André Marine Rigner Garan André Marine Antre Saladi Antre Saladi	Mana bahar Manda Manan	Service	Mang Samada Katala Katala Katala Katala Katala Katala Katala Katala Katala Katala Katala Katala Katala Katala Katala	Dada Wa Hanan Para Kata Kataba Kata Kataba Kata Kataba Kata Kataba Kata Kataba	juang fan	Bonnaldsstratigner Kannagarije Santa Ange Valler sama Gyr S Samagarat	NONCORRUNAL Connected in Security Connected in Security Connected in Security Connected in Security Connected in Security Connected in Security Connected in Security Connected in Security Connected in Security Connected Internected Inte	Mak hara	Mith Schwarg Court
A Springer	A Springer		🖉 Springer	<u>∲</u> Springer			④ Springer	<u>€</u> Springer	<u> </u>	🖉 Springer				<u>€</u> Springer	<u>€</u> Springer	오 Springer	