



Language contact and language choice in parliamentary speech in Suriname

Robert Borges & Margot van den Berg

Introduction

- outline
- Suriname & DNA
- relevance
- research questions



Outline

- Introduction
- Methodology
- Corpus Summary
- Results
- Conclusions

Where are we?

Introduction

- outline
SR & DNA
relevance
research Qs

Methodology

Corpus Summary

Results

Conclusions

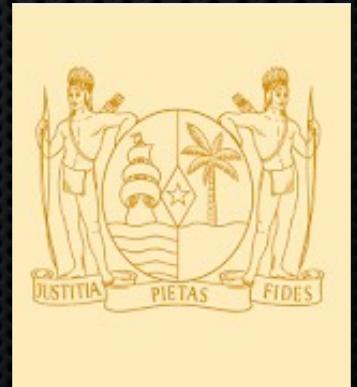
Suriname & Multilingualism

- Population development (Borges 2017)
 - 1651: colony of England – Proto Sranan, Maroon languages
 - 1667: colony of Netherlands – Dutch
 - 19thC: Asian / Indian Languages
 - 20thC: contd in migration, region & beyond
- Intense Multilingual, Multiethnic society
 - “leaky” diglossia
 - prestige highly context dependent

Were are we?
Introduction
outline
● SR & DNA
relevance
research Qs
Methodology
Corpus Summary
Results
Conclusions

De Nationale Assemblée

- Main legislative body of Suriname
- 51 members + guests
 - population relatively well represented geographically and re “ethno-linguistic” diversity
- Strict rules of communicative practice
 - Dutch as default language
 - not explicitly stated or enforced



Where are we?
Introduction
outline
● SR & DNA
relevance
research Qs
Methodology
Corpus Summary
Results
Conclusions

Parliamentary Corpora

- Corpora constructed from parliamentary debates e.g. within the EU (Fiser and Lenardic 2018)
- utilized in e.g. in discourse analysis and sociolinguistics (Hirst et al. 2014; Rheault et al. 2016; Bayley 2004).
- limited domains/registers

Where are we?
Introduction
outline
SR & DNA
● relevance
research Qs
Methodology
Corpus Summary
Results
Conclusions

DNA & SPCL

- Dutch – Sranan language contact
- future of creole studies (?)
 - language contact with adstrate language
 - multilingual language practice

Were are we?
Introduction
outline
SR & DNA
● relevance
research Qs
Methodology
Corpus Summary
Results
Conclusions

Research Questions

- feature variation
 - the Dutch pronominal *er*
 - Word order
 - imperfective marking
- language choice among Surinamese parliamentarians
- method: efficiency, representative sample for hypothesis testing and conclusions

Were are we?

Introduction

outline

SR & DNA

relevance

● research Qs

Methodology

Corpus Summary

Results

Conclusions

Methodology

- data source
- “auto”corpus building
- strengths & weaknesses



Plenty of Data

- DNA meetings are a matter of public record
 - A/V recorded and posted to their youtube channel
 - 1000+ videos: >1min – 6+ hours
- Data are downloaded in orig. quality

Where are we?
Introduction
Methodology
● data source
auto-corpus
SWOT
Corpus Summary
Results
Conclusions

A Semiautomated Corpus

- Corpus linguistics and associated tools typically restricted to “born-written” data



Where are we?

Introduction

Methodology

data source

- auto-corpus
- SWOT

Corpus Summary

Results

Conclusions

A Semiautomated Corpus

- Corpus linguistics and associated tools typically restricted to “born-written” data
 - labor bottleneck
 - ±10hrs work – 1hour data
- better way?



Where are we?
Introduction
Methodology
data source
● auto-corpus
SWOT
Corpus Summary
Results
Conclusions

A Semiautomated Corpus

- Corpus linguistics and associated tools typically restricted to “born-written” data
 - labor bottleneck
 - ±10hrs work – 1hour data
- better way?
 - speech to text



Where are we?
Introduction
Methodology
data source
● auto-corpus
SWOT
Corpus Summary
Results
Conclusions

A Semiautomated Corpus

- Process:
 - start w/ video & .wav files
 - create ELAN (Sloetjes & Wittenberg 2008) project
 - recognizers to delimit speech annotations
 - “chunk” audio file by annotation time stamps w/ elan2split (Cavar 2016)
 - Custom .py sends chunks to google SR api, time aligns result in .eaf
 - diarization recognizer to identify speakers of utterances

Where are we?
Introduction
Methodology
 data source
● auto-corpus
 SWOT
Corpus Summary
Results
Conclusions

An “auto”-corpus

- Result:
 - transcribed spoken-language data
 - utterances ascribed to their speaker



Where are we?
Introduction
Methodology
 data source
● auto-corpus
 SWOT
Corpus Summary
Results
Conclusions

SWOT Analysis

- Strengths

- fast, automated
- “big(er)” data

- Opportunities

- hypothesis testing
- wide application

- Weakness

- accuracy & unverified result
- additional labor required

- Threat

- unverified claim based on “quick ‘n dirty” testing

Where are we?
Introduction
Methodology

data source
auto-corpus

● SWOT
Corpus Summary
Results
Conclusions

Corpus Summary

- pilot data
- output stats
- labor hours



Data for the Pilot

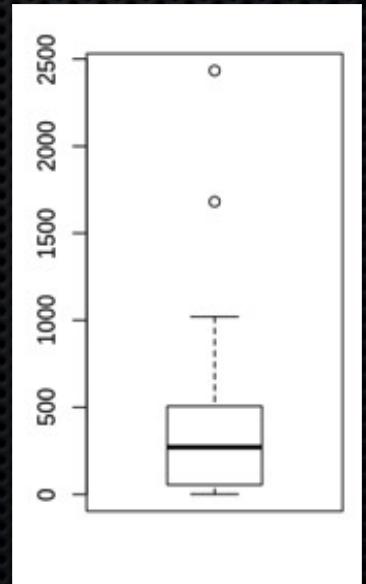
- videos selected pseudorandomly

20181113	2	20190411_01	60
20181123_01	85	20190411_02	90
20181123_02	111	20190411_03	24
20181123_03	18	20190411_04	48
20181101	19	20190405_04	101
20181121_01	21	20190405_05	100
20181121_02	87	20190409_01	102
20190405_06	73	20190409_02	30
20190405_07	33	20190405_01	57
20190405_08	1	20190405_02	57
20190101_02	359	20190405_03	43
20190402	384	20190404_02	311

Where are we?
Introduction
Methodology
[Corpus Summary](#)
• pilot data
output stats
labor hours
Results
Conclusions

Output Summary Statistics

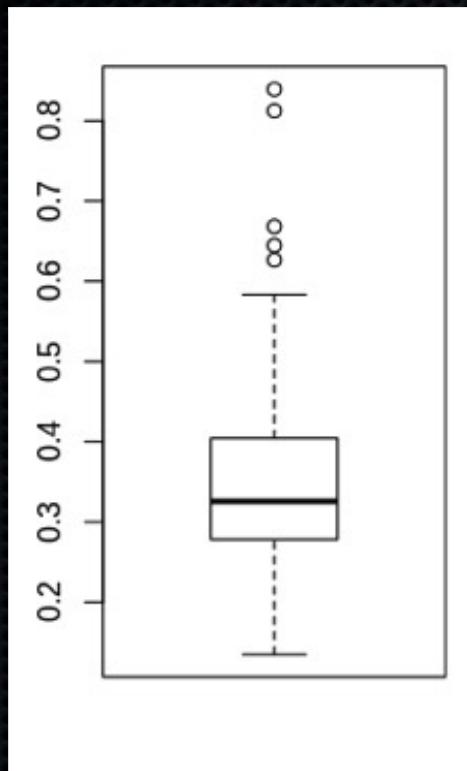
- Transcribed: 36h47m
- Annotations: 21,560
- Words: 204,168 / 14,275
- Speakers: 54



Where are we?
Introduction
Methodology
Corpus Summary
pilot data
● **output stats**
labor hours
Results
Conclusions

Output Summary Statistics

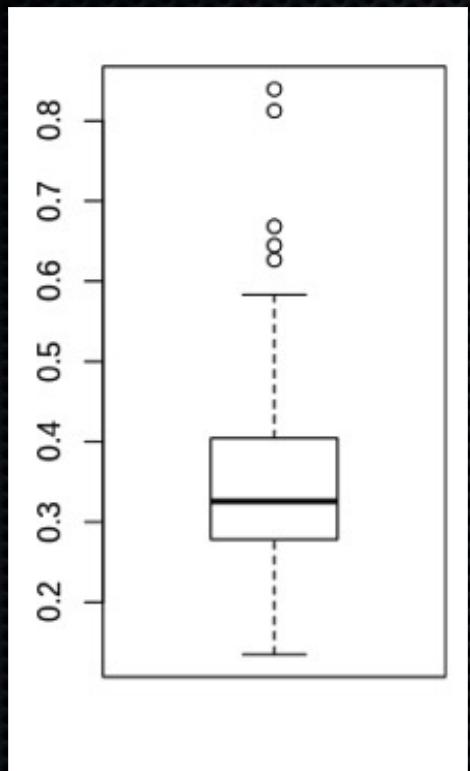
diversity



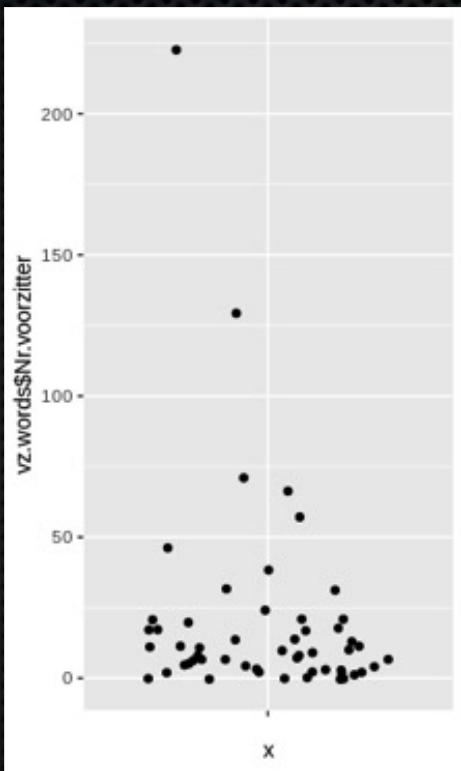
Where are we?
Introduction
Methodology
Corpus Summary
pilot data
• output stats
labor hours
Results
Conclusions

Output Summary Statistics

diversity



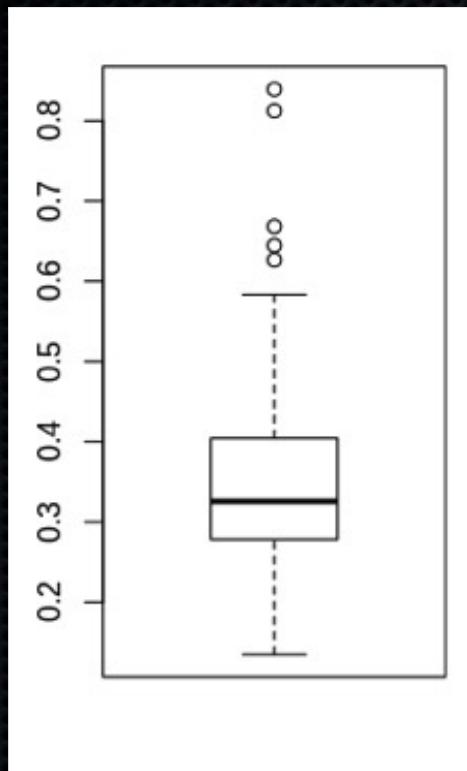
n “voorzitter”



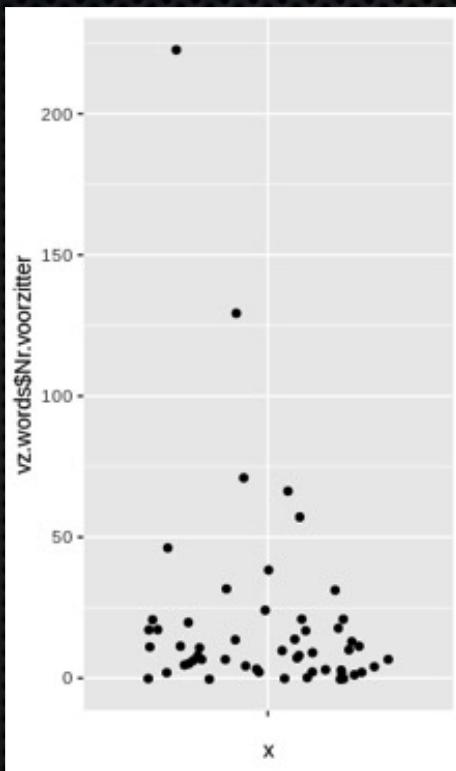
Where are we?
Introduction
Methodology
Corpus Summary
pilot data
● output stats
labor hours
Results
Conclusions

Output Summary Statistics

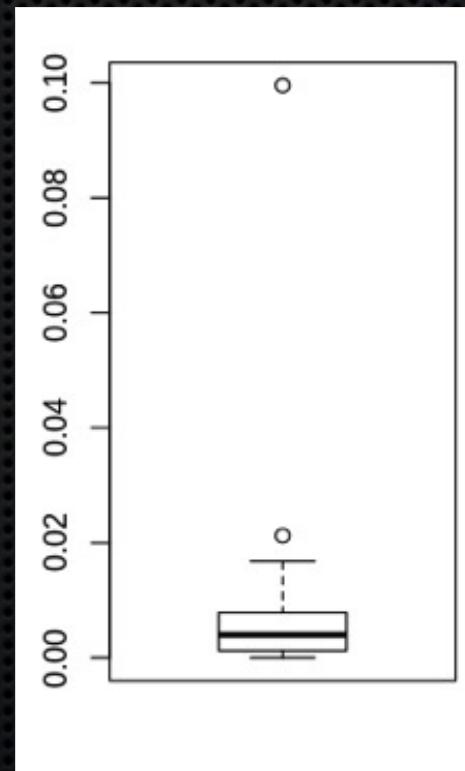
diversity



n “voorzitter”



% voorzitter



Where are we?
Introduction
Methodology
Corpus Summary
pilot data
● output stats
labor hours
Results
Conclusions

Labor Hours

- Preparing files and transcribing: ca. 12h
- Diarization check (ID speakers): ca. 8h

Where are we?
Introduction
Methodology
Corpus Summary
pilot data
output stats
● **labor hours**
Results
Conclusions

Results

- *er*
- word order
- IPFV
- pragmatics



er

“het zo beweeglijke woordje ‘er’ biedt veel moeilijkheden, zodat het of wegbleft of in de zin verdwaalt” (Schoonhoven 1939)

SU Dutch: *Mag ik even door?*

NL Dutch: *Mag ik er even door?*

Dummy subject, dummy adverbial, expletive,
discourse particle (accessibility marker)

- Where are we?
- Introduction
- Methodology
- Corpus Summary
- Results
 - *er*
 - word order
 - IPFV
 - Pragmatics
 - Conclusions

Referential *er* in Netherlandic Dutch

Examples from van Dijk & Coopmans (2000)

Type 1 - Locative

Jan went to the store

Hij kocht er een boek

he bought there a book

Type 2 - Prepositional

Did you hear about the accident?

Ik heb er met hem over gesproken

I have there with him about spoken

Where are we?
Introduction
Methodology
Corpus Summary
Results
• *er*
word order
IPFV
Pragmatics
Conclusions

Type 4 - Quantitative

Did you find those red puppets?

Ik heb er gisteren twee gekocht

I have there yesterday two bought

Non-referential *er* in NL-ic Dutch

Type 3 - Presentative

Er is een hond in de tuin

There is a dog in the garden

In de tuin is er een hond

In the garden is there a dog

Ik geloof dat er een hond in de tuin is

I believe that there a dog in the garden is

Wat gaat er gebeuren met de hond in de tuin?

Wat goes there happen with the dog in the garden

Where are we?

Introduction

Methodology

Corpus Summary

Results

- *er*

word order

IPFV

Pragmatics

Conclusions

er in our DNA data

	frequency	percentage
locative	104	8,07
prepositional	21	1,63
presentative	970	75,25
quantitative	20	1,55
question word	174	13,50
Total	1289	100,00

- Where are we?
 - Introduction
 - Methodology
 - Corpus Summary
 - Results
 - *er*
 - word order
 - IPFV
 - Pragmatics
 - Conclusions

Variation in European Dutch *er*

- Algemene Nederlandse Spraakkunst (1984: 820):
“Voor de aan- of afwezigheid van er zijn geen strikte regels te geven”
 - the postverbal distribution of *er* is governed by region, register, clause type and verbal specificity as shown by Grondelaers (2002), Grondelaers, Speelmans & Carbonez (2001), Grondelaers c.s. (2002)
 - Belgian Dutch x Netherlandic Dutch
- Where are we?
Introduction
Methodology
Corpus Summary
Results
● *er*
word order
IPFV
Pragmatics
Conclusions

Surinamese Dutch *er*?

Influence from Sranantongo: Absence of *er*

“Wat gebeurt met water etcetera”

What happens with water etcetera

“Wat gebeurt er dan, is er dan bescherming?”

What happens then, is there then protection

- Where are we?
- Introduction
- Methodology
- Corpus Summary
- Results
 - *er*
 - word order
 - IPFV
 - Pragmatics
 - Conclusions

Surinamese Dutch *er*?

Influence from Sranantongo: Word order

“minister heeft tijdens haar spreekbeurt bij Essent aangegeven er komt wijziging met Nova en alles”

“ben ik nog steeds van mening dat er was geen reden om dus dat niet aanmerking te laten komen voor eventuele verlenging”

Where are we?
Introduction
Methodology
Corpus Summary
Results

er
• word order
IPFV
Pragmatics
Conclusions

Imperfective Marking

- Muysken: SU Dutch speakers utilize *bezig zijn* and *aan het* imperfective constructions on the Sranan model *e* (Borges et al. 2017).

Where are we?
Introduction
Methodology
Corpus Summary
Results
er
word order
● IPFV
Pragmatics
Conclusions

Imperfective Marking

- Muysken: SU Dutch speakers utilize *bezig zijn* and *aan het* imperfective constructions on the Sranan model *e* (Borges et al. 2017).
- Testing in DNA data:
 - Underwhelming! 102 *bezig*, 179 *aan het* hits (not all IPFV)
 - nothing peculiar cf EU Dutch
- explanation: data types?

Where are we?
Introduction
Methodology
Corpus Summary
Results

er
word order
● IPFV
Pragmatics
Conclusions

Pragmatics: code switching

- Frequent use of code switching (relatively speaking).
 - completely acceptable, but...
 - not modeled, e.g. by voorzitter
- function (???)
 - not translation (competence not at issue)
 - parenthetical support of arguments
 - camaraderie, appealing to “the local” / sense of belonging, persuasion

Where are we?
Introduction
Methodology
Corpus Summary
Results

er
word order
IPFV
● Pragmatics
Conclusions

R. Doekie

I. Bink

R. Dodson

Conclusions

- future applicability
- insights re SPCL
- insights in general



The Future of this Project

- Continued corpus building
 - hypothesis testing
 - ¿empirically-based descr. of characteristics?
- Structured testing of labor/accuracy tradeoff
- “Deeper” analysis – NLPTK / spaCy, existing modules for Dutch POS and dependency tagging

Where are we?
Introduction
Methodology
Corpus Summary
Results
Conclusions
● future applicability
insights re SPCL
insights in general

Take-home for P&C Studies

- usage based, empirically driven accounts
 - contact-induced language change
 - language practice

Where are we?
Introduction
Methodology
Corpus Summary
Results
Conclusions
future applicability
● insights re SPCL
insights in general

Take-home, in general

- Use technology to make your work easier

Where are we?
Introduction
Methodology
Corpus Summary
Results
Conclusions
future applicability
insights re SPCL
• insights in general

End matter

- Acknowledgements
 - Uppsala University, Hugo Valentin Centre
 - Utrecht University, Department of Linguistics
 - Hans Sloetjes (MPI), Damir Cavar (Indiana U.)
 - WoGEL group @ Uppsala U. dept of LingFil

End matter

- Bayley, Paul, ed. (2004). Cross-Cultural Perspectives on Parliamentary Discourse. Discourse Approaches to Politics, Society, and Culture v. 10. Amsterdam ; Philadelphia: Benjamins.
- Bies, de R. (2008) Woordenboek van de Surinaamse Bijdrage aan het Nederlands, Paramaribo: Universiteit van Suriname
- Borges, Robert, Pieter Muysken, Sophie Villerius, Kofi Yakpo (2017). The tense-mood-aspect systems of the languages of Suriname. In Kofi Yakpo, Pieter C. Muysken (Eds.), *Boundaries and Bridges: Language Contact in Multilingual Ecologies* (pp. 311–362). Berlin, Boston: De Gruyter
- Cavar, Damir. 2017. Elan2Split. <https://bitbucket.org/dcavar/elan2split>
- Dijk, Chantal van & Coopmans, Peter (2013). On the acquisition of ‘daar’ and ‘er’. *Linguistics in the Netherlands*, 30(1), 73-88.
- DNA, De Nationale Assemblée (1990). Reglement Van Orde Voor De Nationale Assemblée.
- Donselaar, van, Jan (1989) J. Woordenboek van het Surinaams-Nederlands
- Fiser, Darja and Jakob Lenardic (2018). ‘Parliamentary Corpora in the CLARIN Infrastructure’. In: Selected Papers from the CLARIN Annual Conference 2017, Budapest, 18-20 September 2017. 147. Linköping University Electronic Press, Linköping Universitet / Department of Translation, Faculty of Arts, University of Ljubljana, Slovenia, pp. 75-85.
- Grondelaers S, M. Brysbaert, D. Speelman & D. Geeraerts (2002) “Er” als accessibility marker: on- en offline evidentie voor een procedurele interpretatie van presentatieve zinnen” Gramma/TTT 9:1-22
- Grondelaers, S. (2000) De distributie van niet-anaforisch er buiten de eerste zinsplaats. Sociolexicologische, functionele en psycholinguistische aspecten van er’s status als presentatief signaal. PhD Thesis, KU Leuven
- Hirst, Graeme, Vanessa Wei Feng, Christopher Cochrane, and Nona Naderi. (2014). ‘Argumentation, Ideology, and Issue Framing in Parliamentary Discourse’. In: Proceedings of the Workshop on Frontiers and Connections between Argumentation Theory and Natural Language Processing in Forlì-Cesena, Italy, July 21-25, 2014, edited by Elena Cabrio, Serena Villata, and Adam Wyner, pp. 50-56.
- Kleine De, C. (2002) Surinamese Dutch. In Carlin, E. Arends, J. (eds.) *Atlas of the Languages of Suriname*. Leiden: KITLV Press
- Rheault, Ludovic et al. (2016). ‘Measuring Emotion in Parliamentary Debates with Automated Textual Analysisâ’. In: PLOS ONE 11.12.
- Schoonhoven, C.W. 1939 ‘Het Nederlands in Suriname’ Nieuwe Taalgids 32, p. 156
- Sloetjes, Han and Peter Wittenburg (2008). ‘Annotation by Category: ELAN and ISO DCR’. In: LREC 2008.
- Vervoorn, A. J. (ca 1973) Het Nederlands en de Surinaamse Letterkunde. Schakels. Den Haag: Kabinet voor Surinaamse en Nederlands-Antiliaanse zaken
- <http://sulinguista.blogspot.com/2016/03/surinaams-nederlands-wat-is-dat.html>

Thanks for your attention!

Email us with questions or comments:

Bob robert.d.borges@gmail.com

Margot m.c.vandenbergh1@uu.nl



Questions?

Google Speech Rec. API

Afrikaans (South Africa)	Spanish (Costa Rica)	Lithuanian (Lithuania)	Hebrew (Israel)
Amharic (Ethiopia)	Spanish (Ecuador)	Hungarian (Hungary)	Arabic (Israel)
Armenian (Armenia)	Spanish (El Salvador)	Malayalam (India)	Arabic (Jordan)
Azerbaijani (Azerbaijan)	Spanish (Spain)	Marathi (India)	Arabic (United Arab Emirates)
Indonesian (Indonesia)	Spanish (United States)	Dutch (Netherlands)	Arabic (Bahrain)
Malay (Malaysia)	Spanish (Guatemala)	Nepali (Nepal)	Arabic (Algeria)
Bengali (Bangladesh)	Spanish (Honduras)	Norwegian Bokmål (Norway)	Arabic (Saudi Arabia)
Bengali (India)	Spanish (Mexico)	Polish (Poland)	Arabic (Iraq)
Catalan (Spain)	Spanish (Nicaragua)	Portuguese (Brazil)	Arabic (Kuwait)
Czech (Czech Republic)	Spanish (Panama)	Portuguese (Portugal)	Arabic (Morocco)
Danish (Denmark)	Spanish (Paraguay)	Romanian (Romania)	Arabic (Tunisia)
German (Germany)	Spanish (Peru)	Sinhala (Sri Lanka)	Arabic (Oman)
English (Australia)	Spanish (Puerto Rico)	Slovak (Slovakia)	Arabic (State of Palestine)
English (Canada)	Spanish (Dominican Republic)	Slovenian (Slovenia)	Arabic (Qatar)
English (Ghana)	Spanish (Uruguay)	Sundanese (Indonesia)	Arabic (Lebanon)
English (United Kingdom)	Spanish (Venezuela)	Swahili (Tanzania)	Arabic (Egypt)
English (India)	Basque (Spain)	Swahili (Kenya)	Persian (Iran)
English (Ireland)	Filipino (Philippines)	Finnish (Finland)	Hindi (India)
English (Kenya)	French (Canada)	Swedish (Sweden)	Thai (Thailand)
English (New Zealand)	French (France)	Tamil (India)	Korean (South Korea)
English (Nigeria)	Galician (Spain)	Tamil (Singapore)	Chinese, Mandarin (Traditional, Taiwan)
English (Philippines)	Georgian (Georgia)	Tamil (Sri Lanka)	Chinese, Cantonese (Traditional, Hong Kong)
English (Singapore)	Gujarati (India)	Tamil (Malaysia)	Japanese (Japan)
English (South Africa)	Croatian (Croatia)	Telugu (India)	Chinese, Mandarin (Simplified, Hong Kong)
English (Tanzania)	Zulu (South Africa)	Vietnamese (Vietnam)	Chinese, Mandarin (Simplified, China)
English (United States)	Icelandic (Iceland)	Turkish (Turkey)	
Spanish (Argentina)	Italian (Italy)	Urdu (Pakistan)	
Spanish (Bolivia)	Javanese (Indonesia)	Urdu (India)	
Spanish (Chile)	Kannada (India)	Greek (Greece)	
Spanish (Colombia)	Khmer (Cambodia)	Bulgarian (Bulgaria)	
	Lao (Laos)	Russian (Russia)	
	Latvian (Latvia)	Serbian (Serbia)	
		Ukrainian (Ukraine)	