

UniDive: WG3

TASK 2

Learning about **ELG** and
data behind it

Number of resources per year

2020 – 92

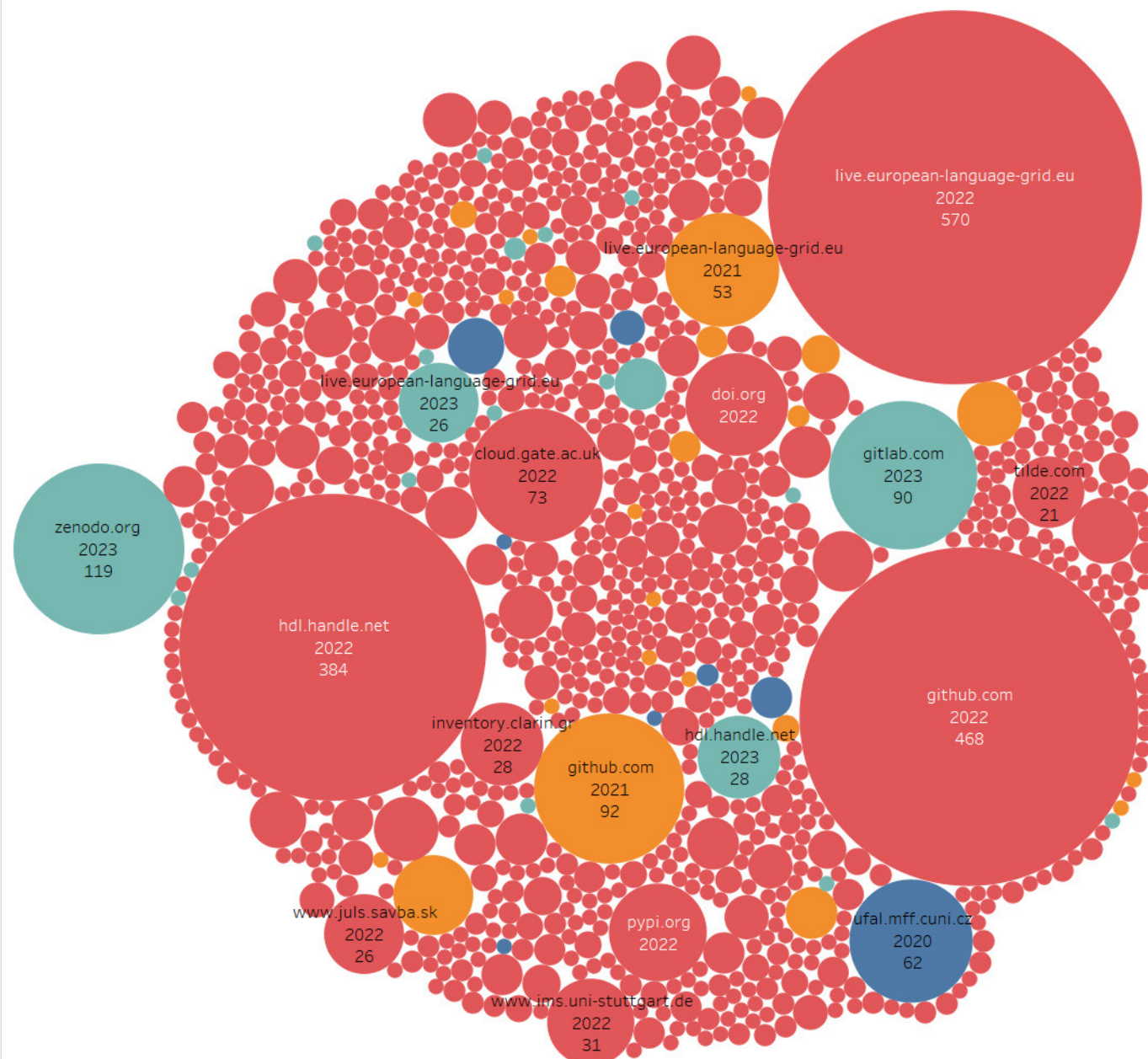
2021 – 237

2022 – 3 256

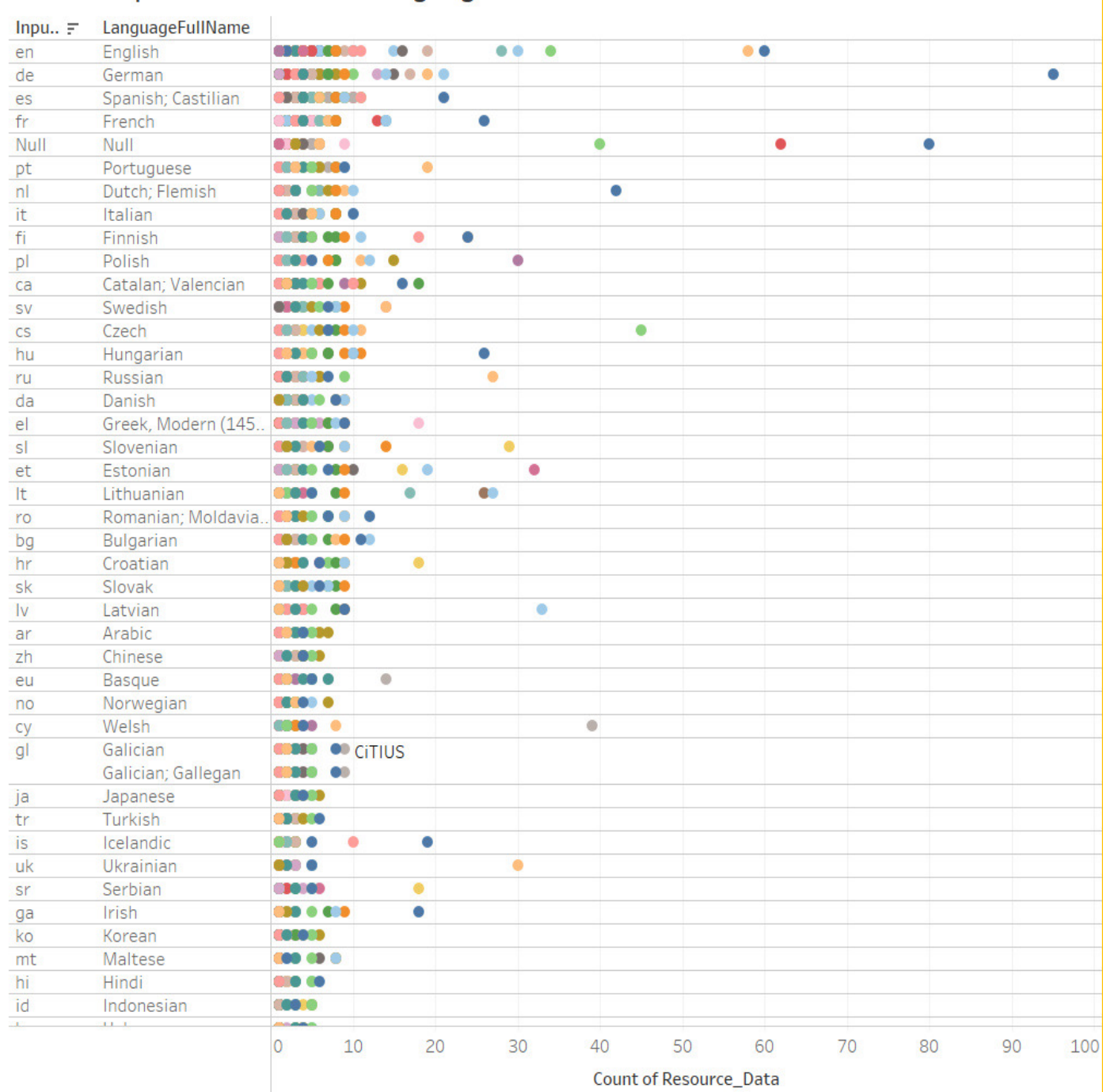
2023 – 290...

TOTAL – 3 875

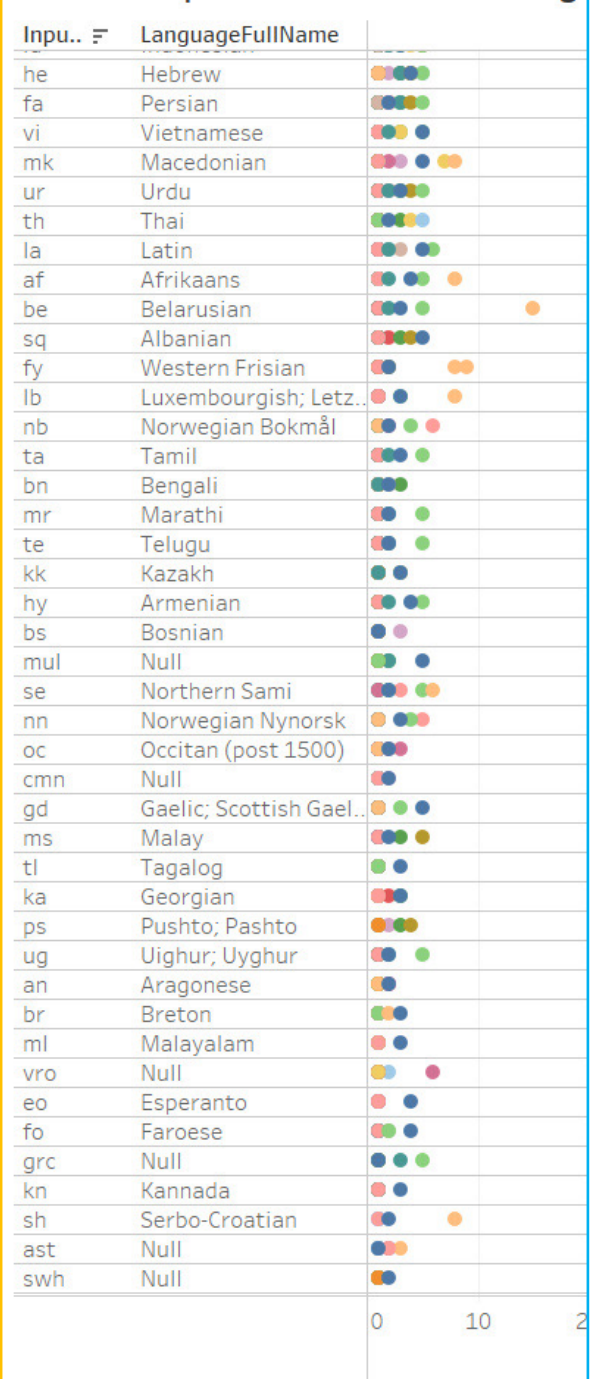
Number of resources added per year
and its Landing Page



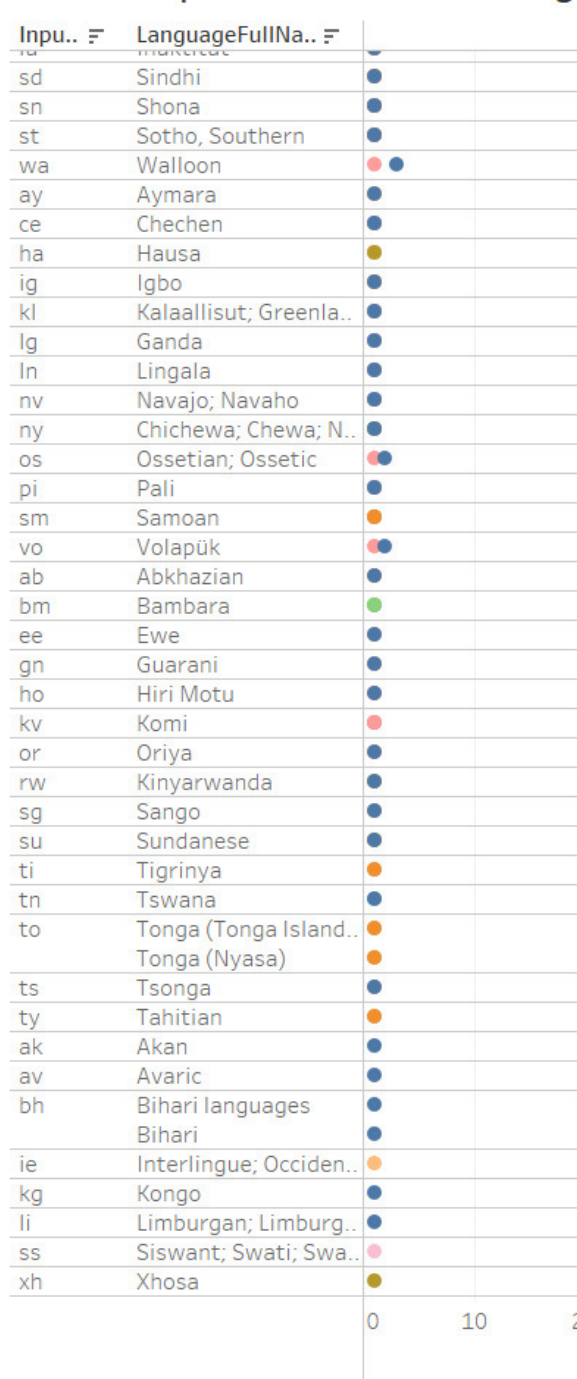
Resource providers for each language



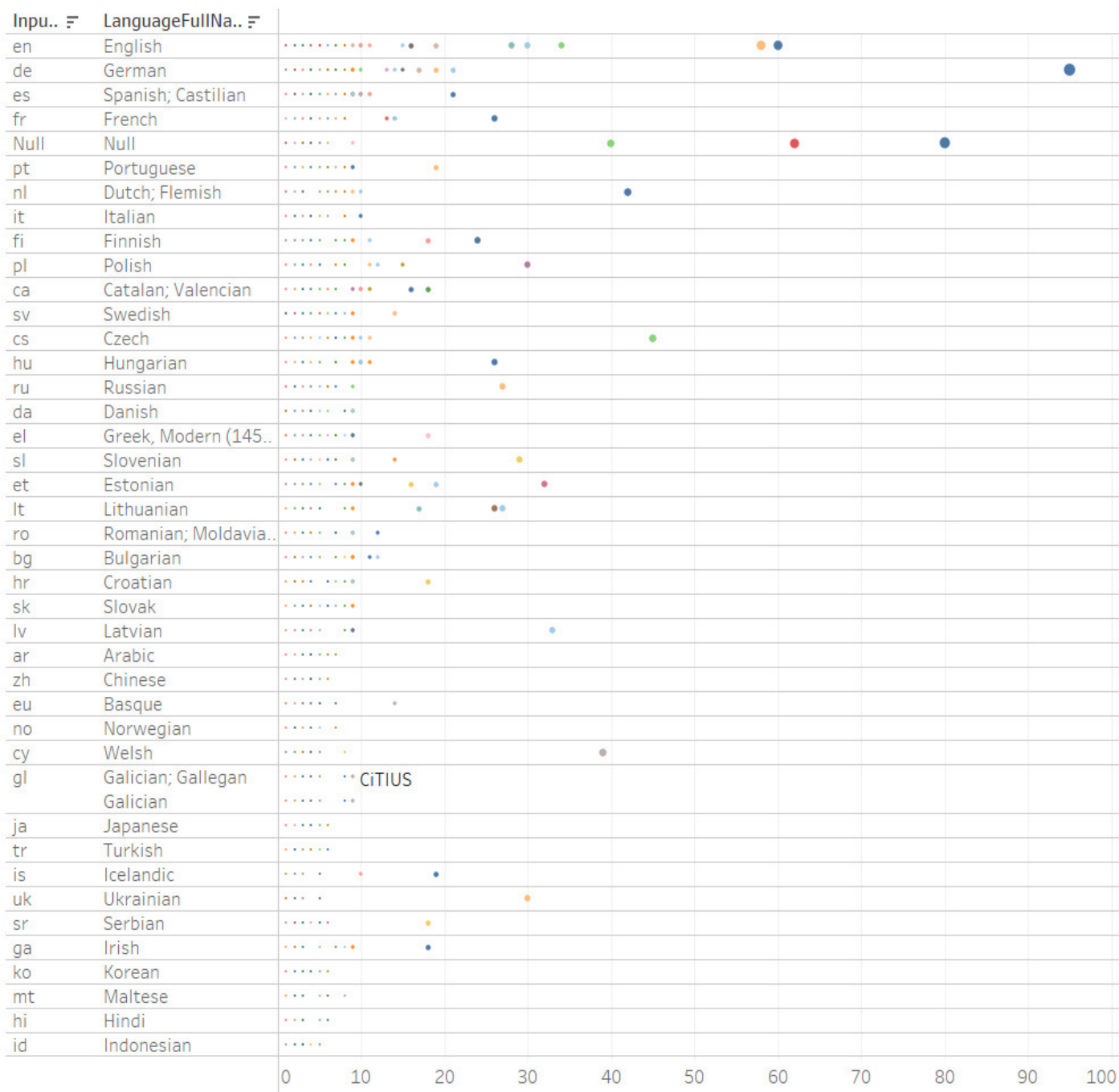
Resource providers for each language



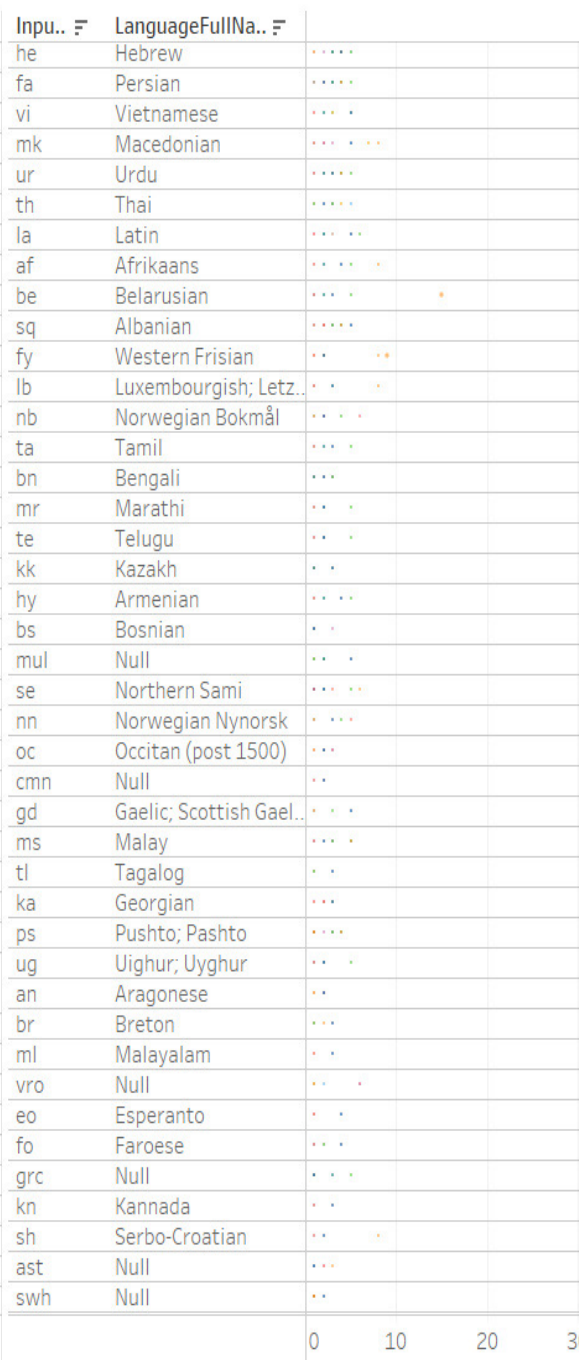
Resource providers for each language



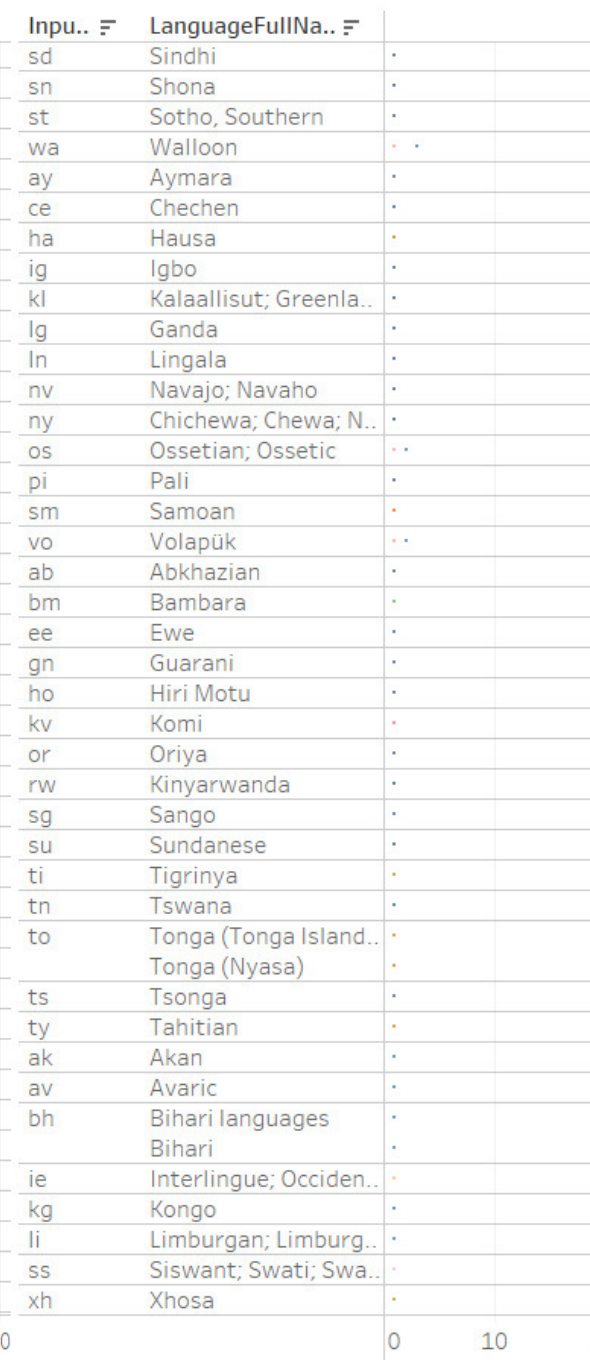
Resource providers for each language



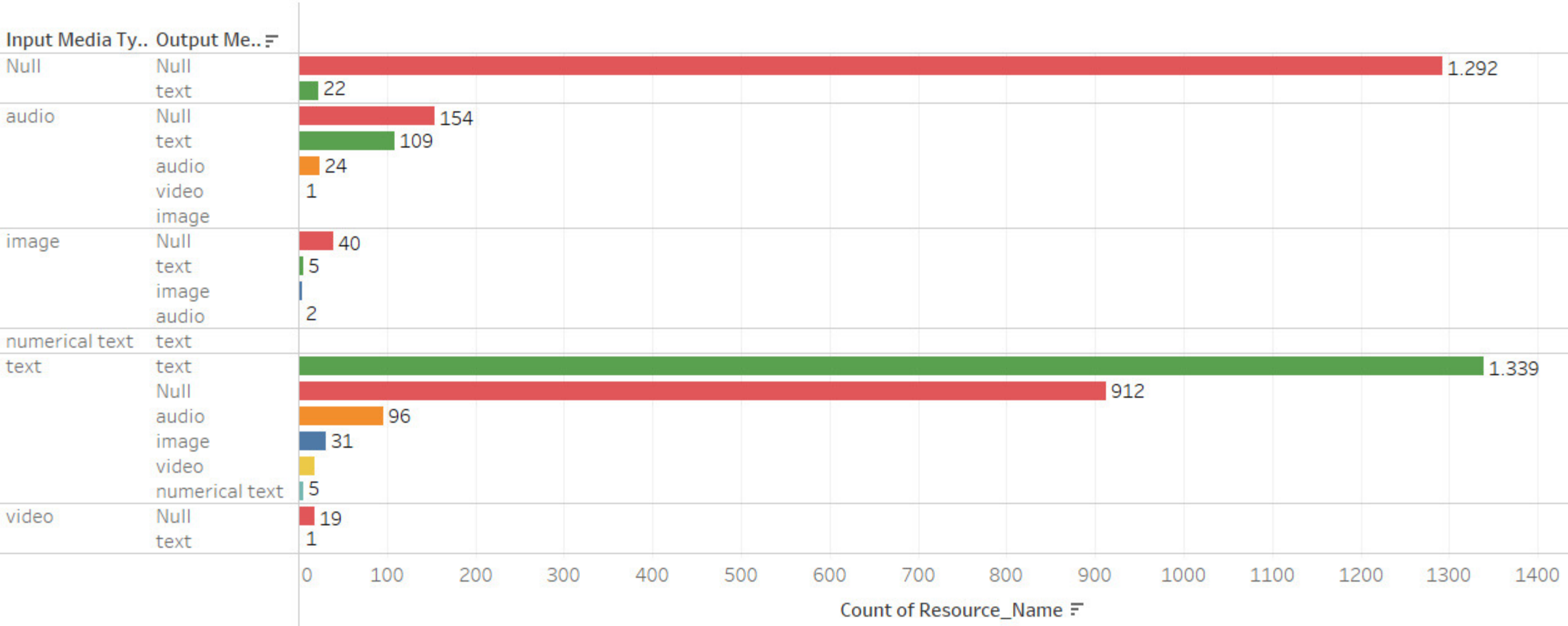
Resource providers for each language



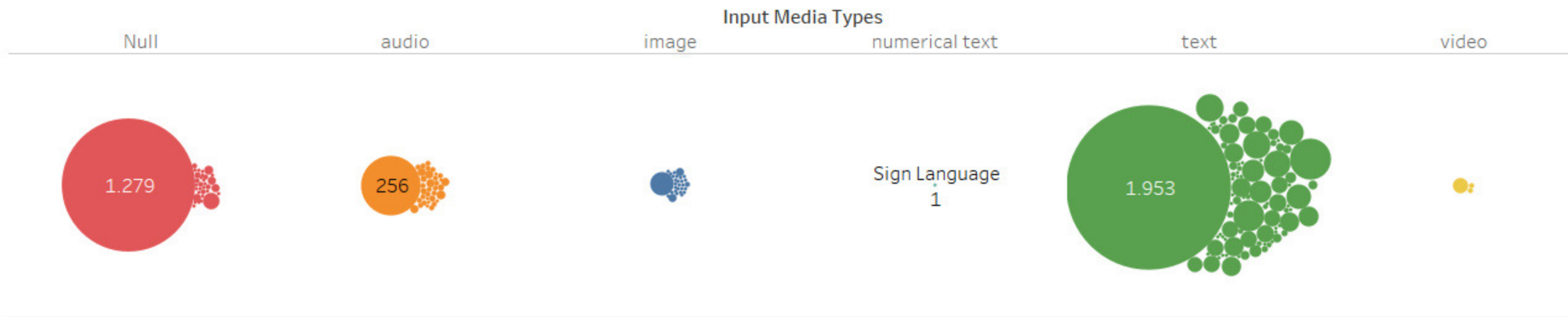
Resource providers for each language



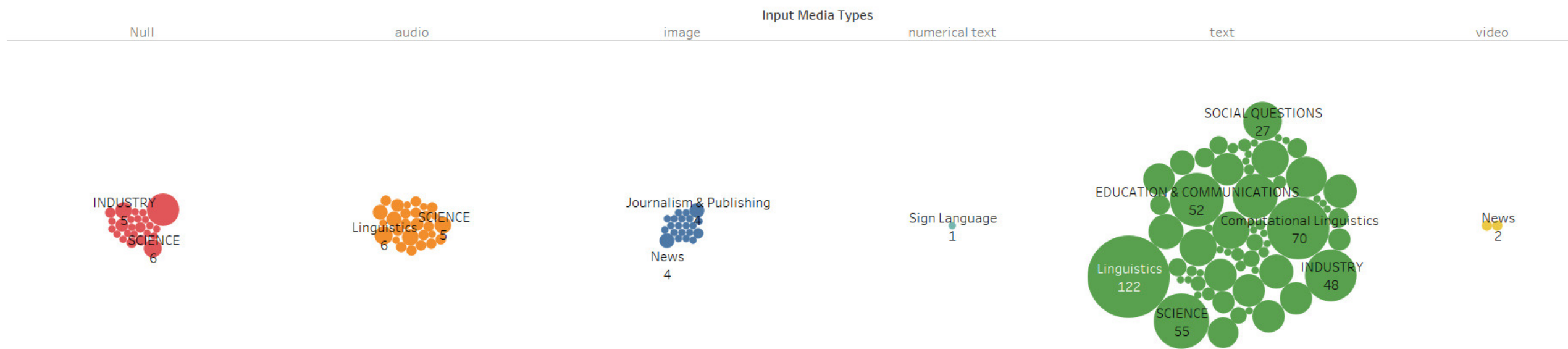
Output media type per Input media type



Different domains per Input media type



Different domains per Input media type

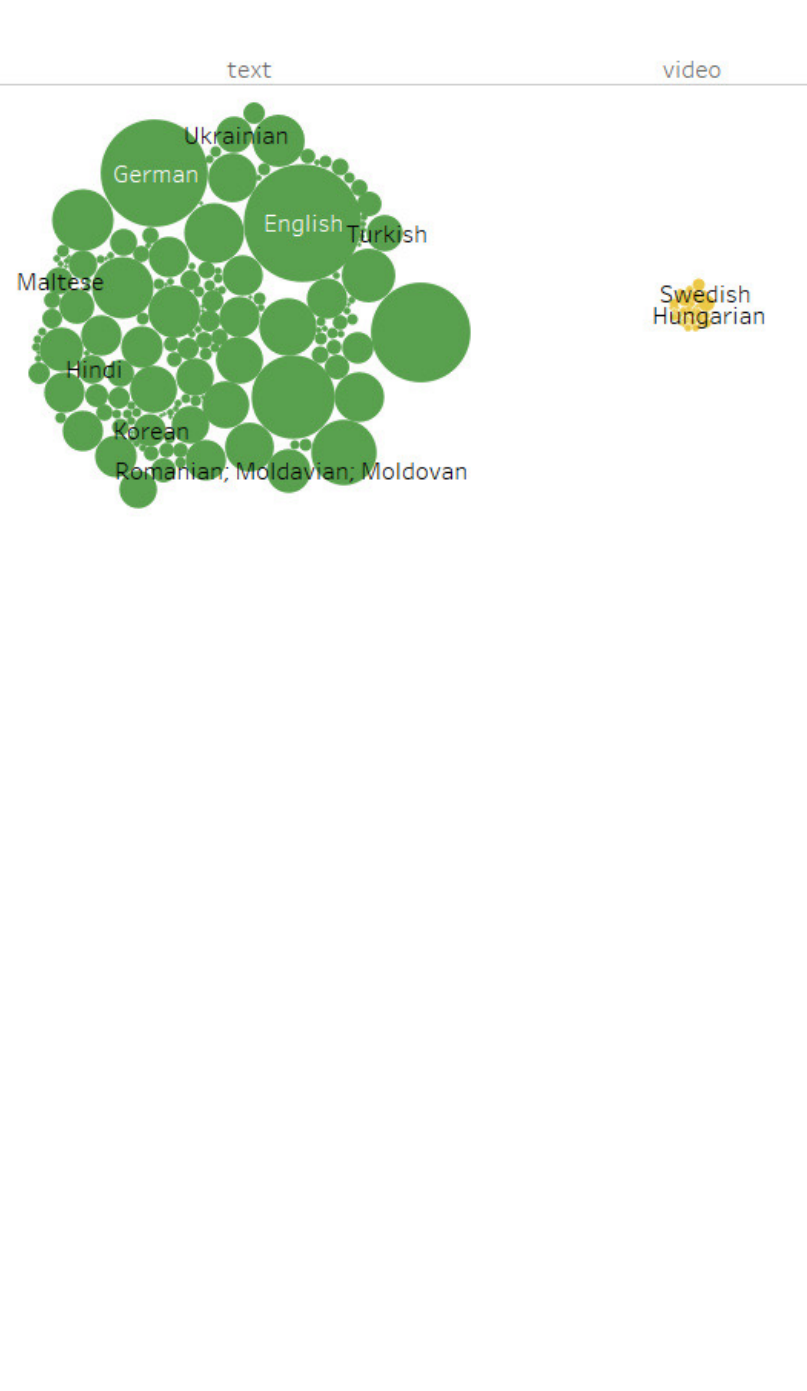


Different languages per Input media type

LanguageFullName	Input Media Types				
	Null	audio	image	text	video
English	●	●●	●	●●●●	●
German	●	●●	●	●●●●	●
Spanish; Castilian	●	●●	●	●●●●	●
French	●	●●	●	●●●●	●
Portuguese	●	●	●	●●●●	●
Catalan; Valencian	●	●	●	●●●●	●
Italian	●	●	●	●●●●	●
Dutch; Flemish	●	●	●	●●●●	●
Finnish	●	●	●	●●●●	●
Russian	●	●	●	●●●●	●
Swedish	●	●	●	●●●●	●
Polish	●	●	●	●●●●	●
Greek, Modern (1453-)	●	●	●	●●●●	●
Hungarian	●	●	●	●●●●	●
Estonian	●	●	●	●●●●	●
Czech	●	●	●	●●●●	●
Danish	●	●	●	●●●●	●
Romanian; Moldavian; Moldovan	●	●	●	●●●●	●

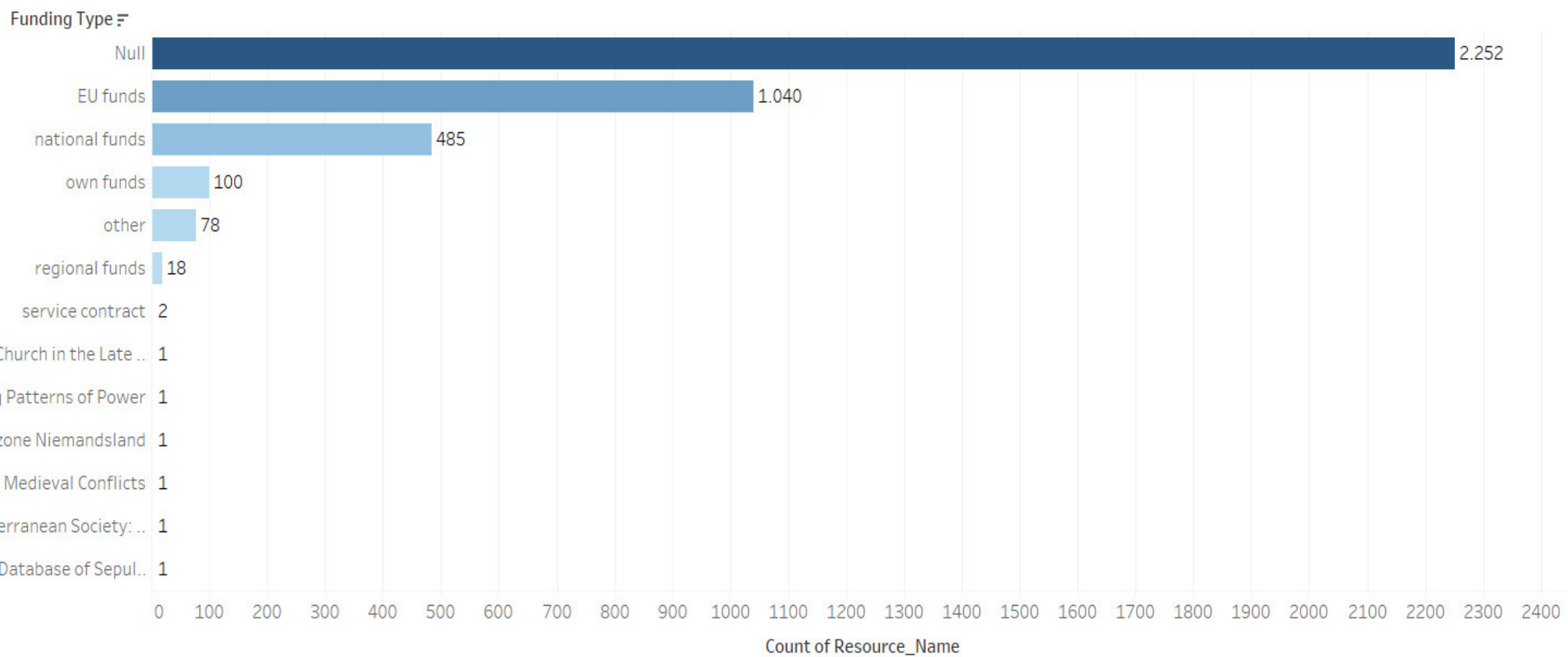
Different languages per Input media type

LanguageFullName	Input Media Types				
	Null	audio	image	text	video
Siswant; Swati; Swazi	●				
Sango	●				
Pali	●				
Oriya	●				
Limburgan; Limburger; Limburgi..	●				
Kongo	●				
Komi	●	●			
Interlingue; Occidental	●				
Hiri Motu	●				
Guarani	●				
Ewe	●				
Chechen	●				
Bambara	●				
Akan	●				
Abkhazian	●				
Bihari languages	●				
Bihari	●				
Avaric	●				



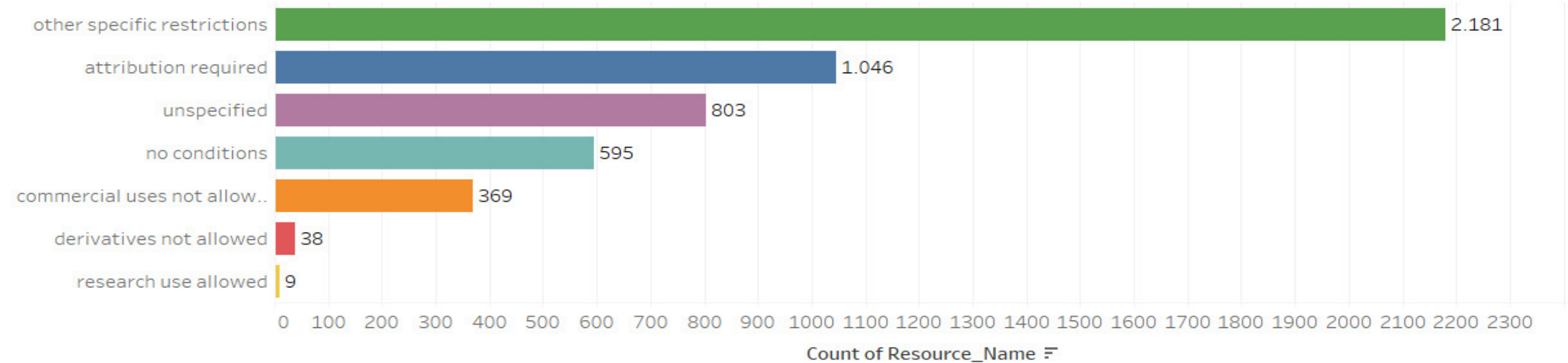
FUNDING

Funding type



Licences

Conditions of Use

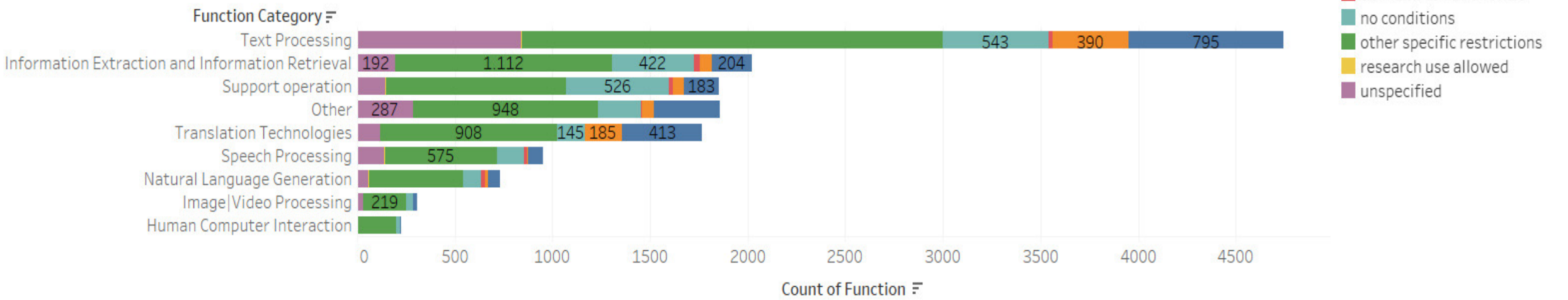


Conditions of Use	Access Rights	Count
other specific restrictions	Null	1.292
	Licensed with a fee	457
	Licensed without a fee for specific uses	266
	Open Access	68
	GNU General Public License	
	CLARIN_ACA	9
	BSD	
	Licensed without a fee for all uses	16
	proprietary	
	GNU Library General Public License	5
	Creative Commons Attribution Share Alike	
	Proprietary software/Free	2
	other	
	closed source	2
	Closed Access	
	open source	1
	Mozilla Public License	
	GNU Lesser General Public License	1
	Free for research, but primarily commercial.	
	European Union Public License	1
	Croatian Academic Network users and privat..	
	Copyright of Tartu Helikool	1
	Copyright of Eesti Keele Instituut and V H ro..	
	Copyright 2011-2026 The Arctic University o..	1
	CLARIN_RES	
	Available - Restricted Use	1

Conditions of Use	Access Rights	Count
attribution required	Null	1.011
	Open Access	26
	Creative Commons Attribution Share Alike	3
	Creative Commons Attribution	1
	Licensed without a fee for specific uses	1
	European Union Public License	
	Creative Commons Attribution Non Commerc..	1
unspecified	Null	695
	unknown	76
	under negotiation	16
	Open Access	
	Licensed without a fee for specific uses	6
	Licensed with a fee	
	Copyright of DCU	1
no conditions	Licensed without a fee for all uses	570
	Free	13
	Null	
	Proprietary software/Free	2
	open source	
	Copyright of Eesti Keele Instituut and V H ro..	1
commercial uses not allowed	Null	365
	Open Access	2
	Licensed without a fee for specific uses	
	Creative Commons Attribution Non Commerc..	1
derivatives not allowed	Null	
research use allowed	CLARIN_ACA	

Function Category COLLECTIONS

broken down by *Conditions of use*



Function Ca..	Function (Function)	Count
Text Processing	Part-of-Speech Tagging	258
	Named Entity Recognition	208
	Morphological annotation	196
	Lemmatization	196
	Tokenization	196
	Parsing	129
	Text and Data Analytics	115
	Text categorization	115
	Sentiment analysis	98
	Spell checking	98
	Dependency parsing	72
	Grammar checking	72
	Information Extraction	65
	Sentence splitting	51
	Semantic annotation	51
	Annotation	51
	Language identification	39
	Topic Detection	39
	Text and Data Mining	39
	Entity linking	39
Term extraction	38	
Authoring support	38	

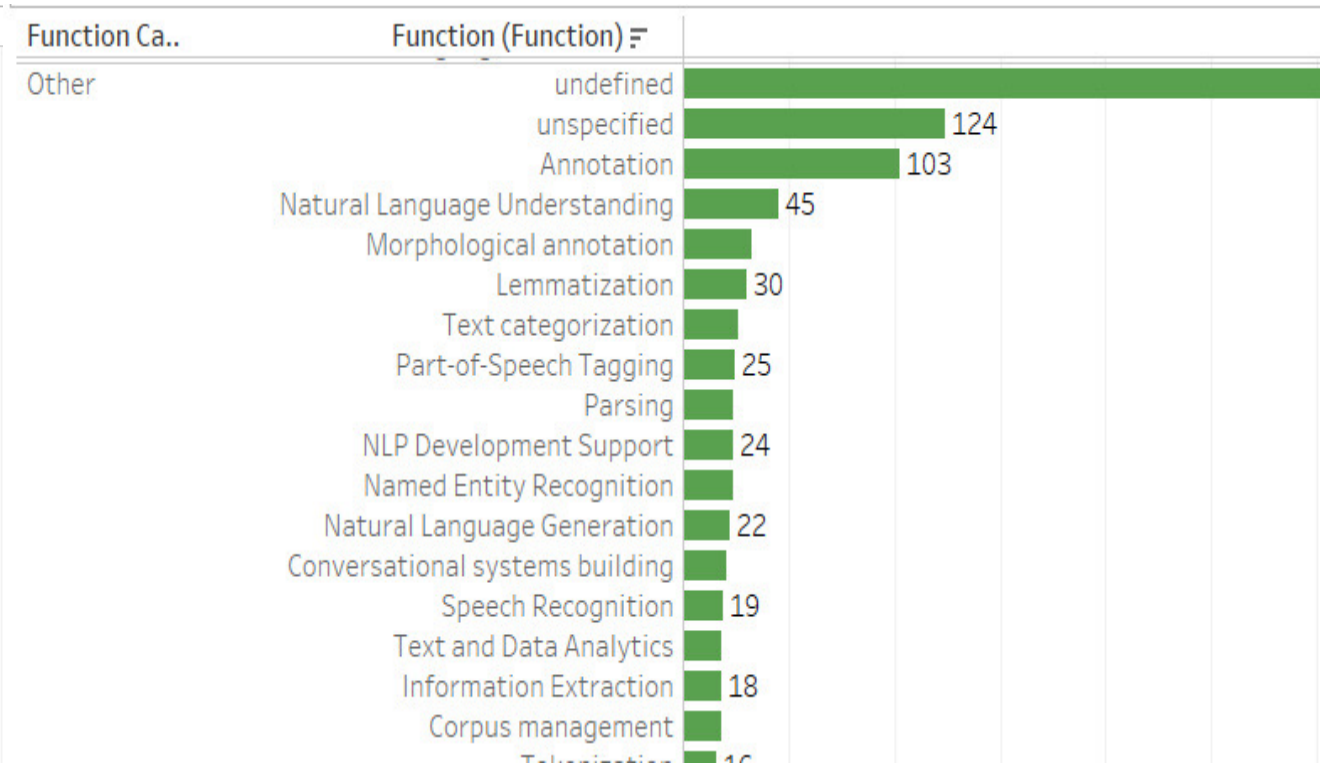
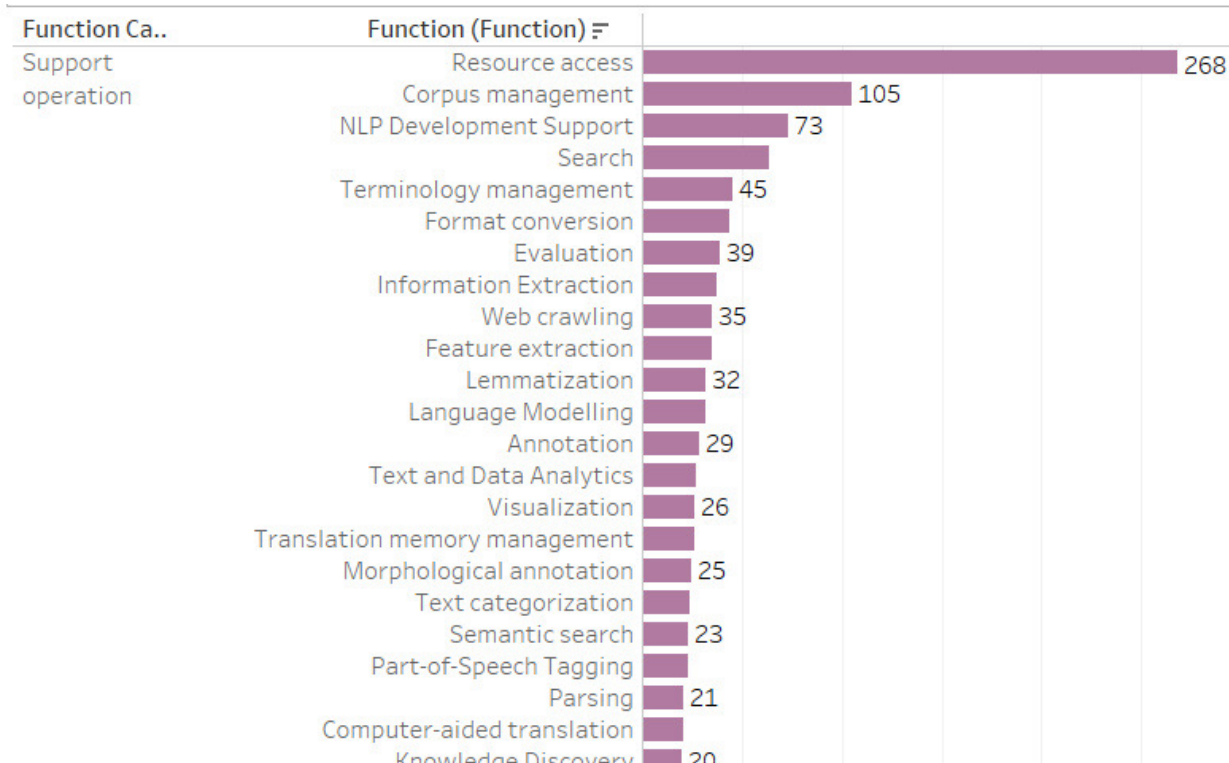
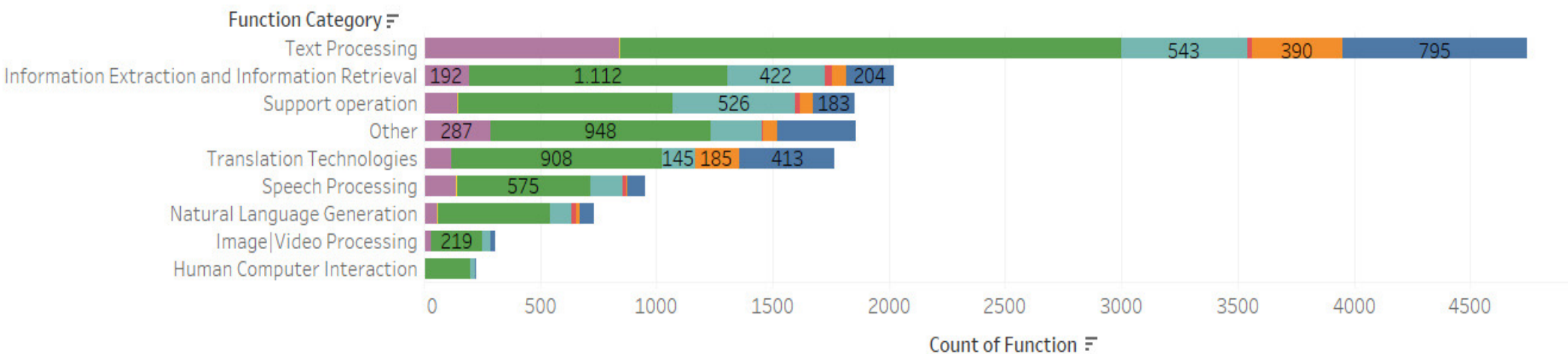
Function Ca..	Function (Function)	Count
Information Extraction and Information Retrieval	Information Extraction	142
	Text and Data Analytics	142
	Term extraction	89
	Resource access	89
	Semantic search	74
	Search	74
	Text categorization	63
	Corpus management	63
	Text and Data Mining	52
	Knowledge Discovery	52
	Named Entity Recognition	41
	Knowledge Representation	41
	Speech analytics	40
	Sentiment analysis	40
	Topic Detection	26
	Multimedia Information Extraction	26
	Lemmatization	26
	Annotation	25
	Relation Extraction	25
	Web crawling	23
Parsing	23	

Function Category COLLECTIONS

broken down by *Conditions of use*

Conditions of Use

- attribution required
- commercial uses not allowed
- derivatives not allowed
- no conditions
- other specific restrictions
- research use allowed
- unspecified

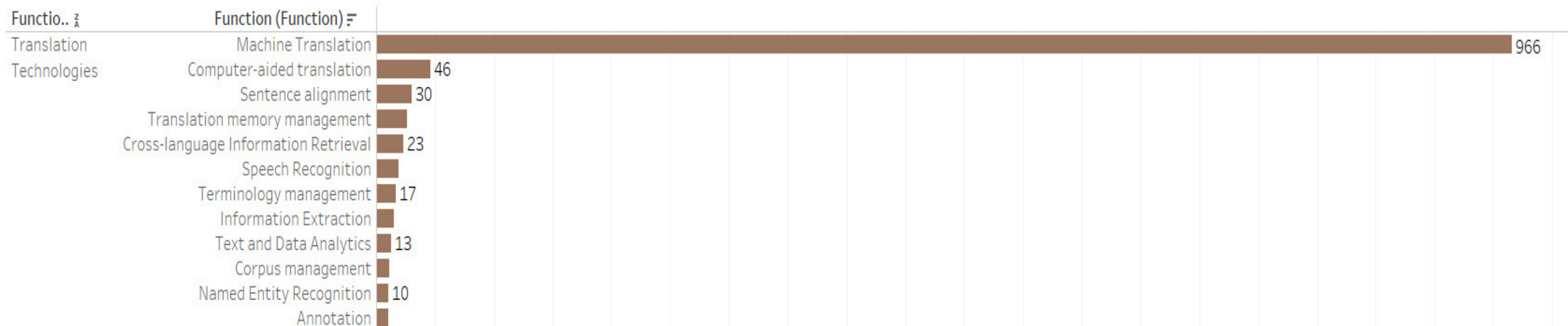
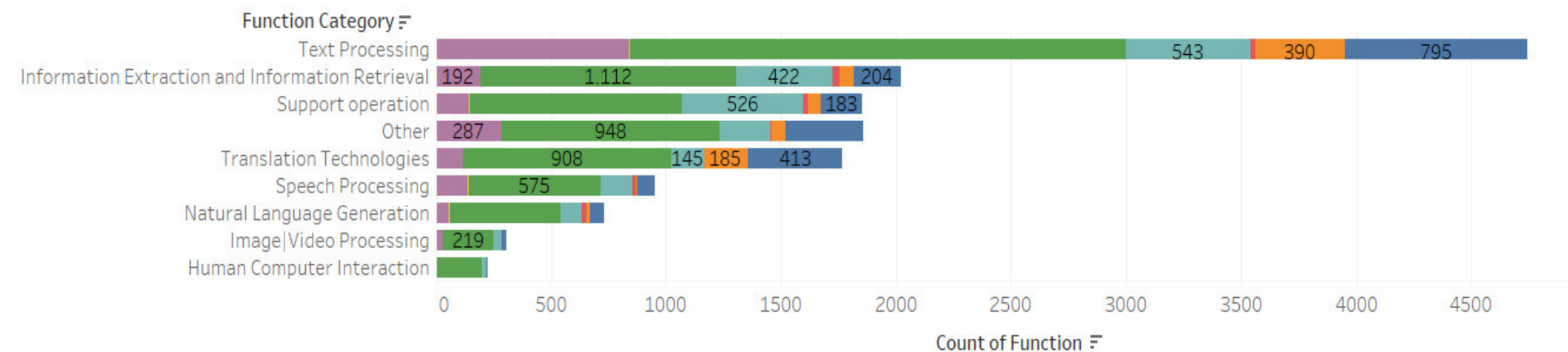


Function Category COLLECTIONS

broken down by *Conditions of use*

Conditions of Use

- attribution required
- commercial uses not allowed
- derivatives not allowed
- no conditions
- other specific restrictions
- research use allowed
- unspecified

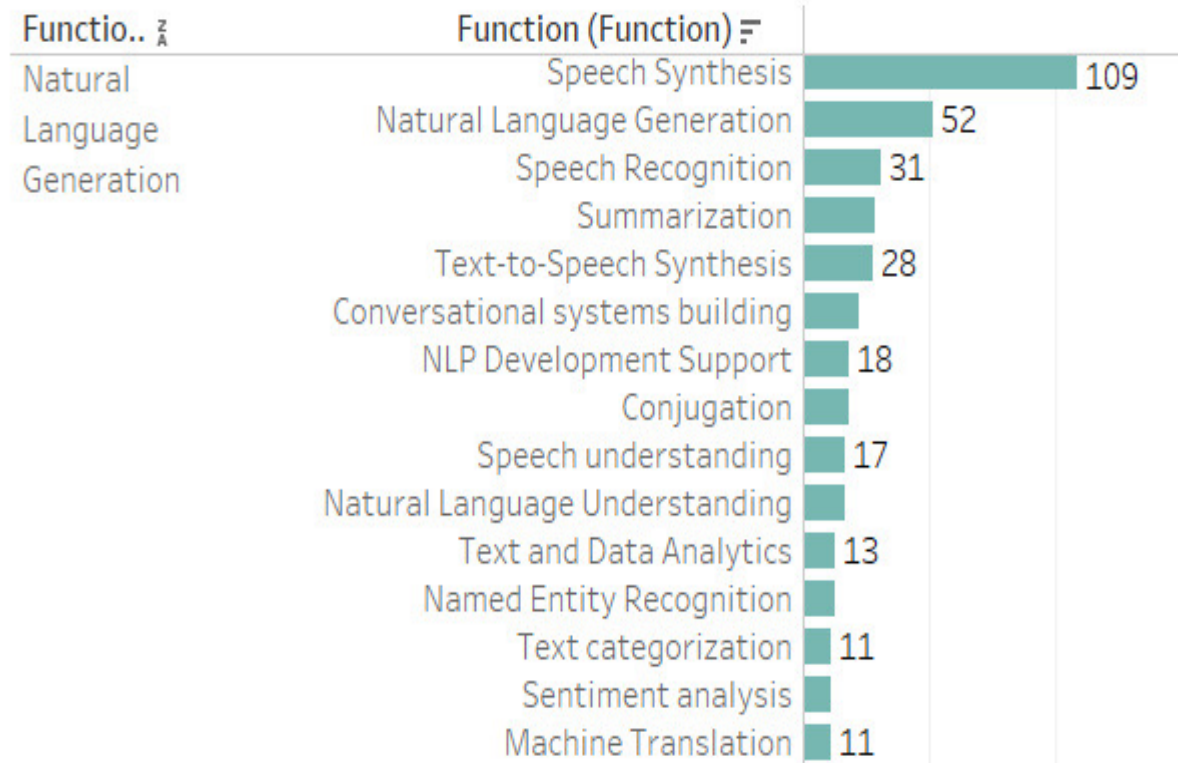
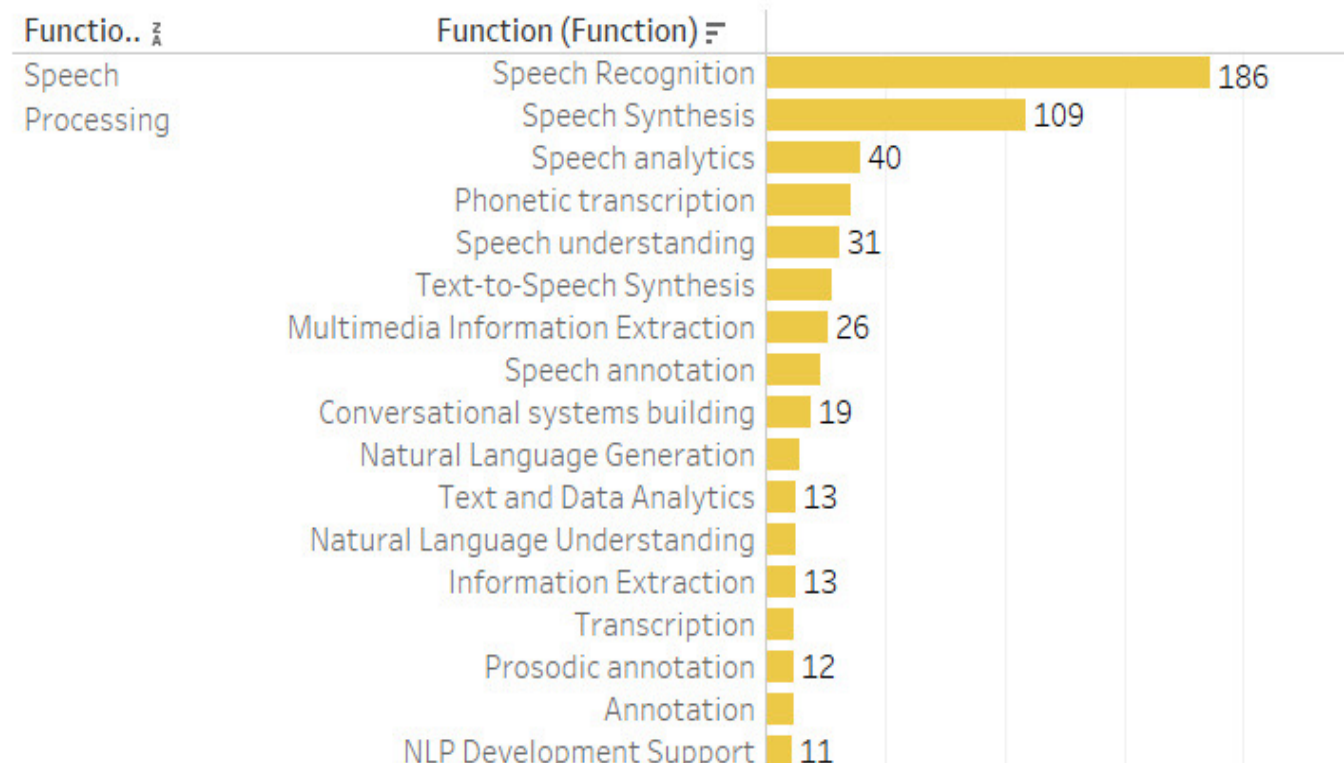
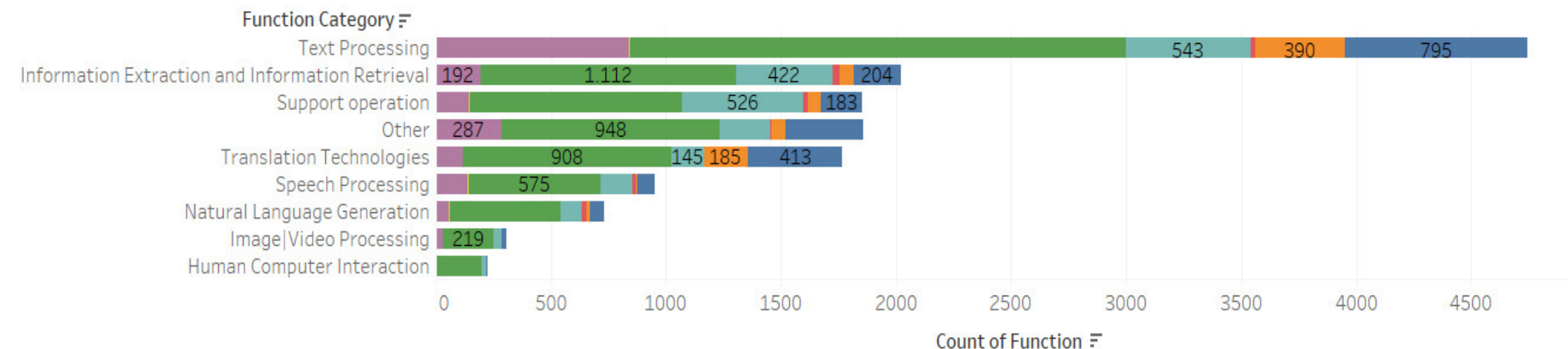


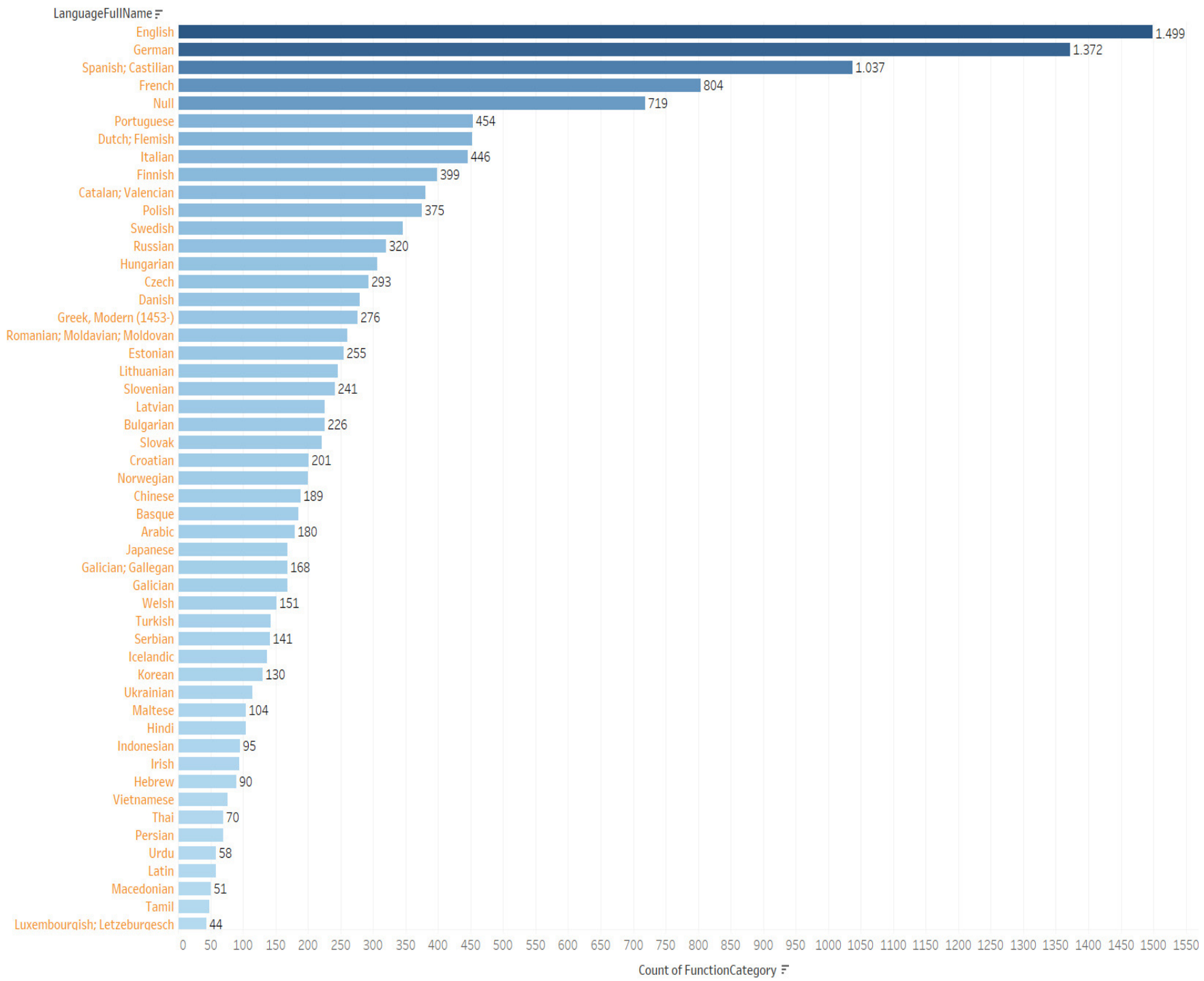
Function Category COLLECTIONS

broken down by *Conditions of use*

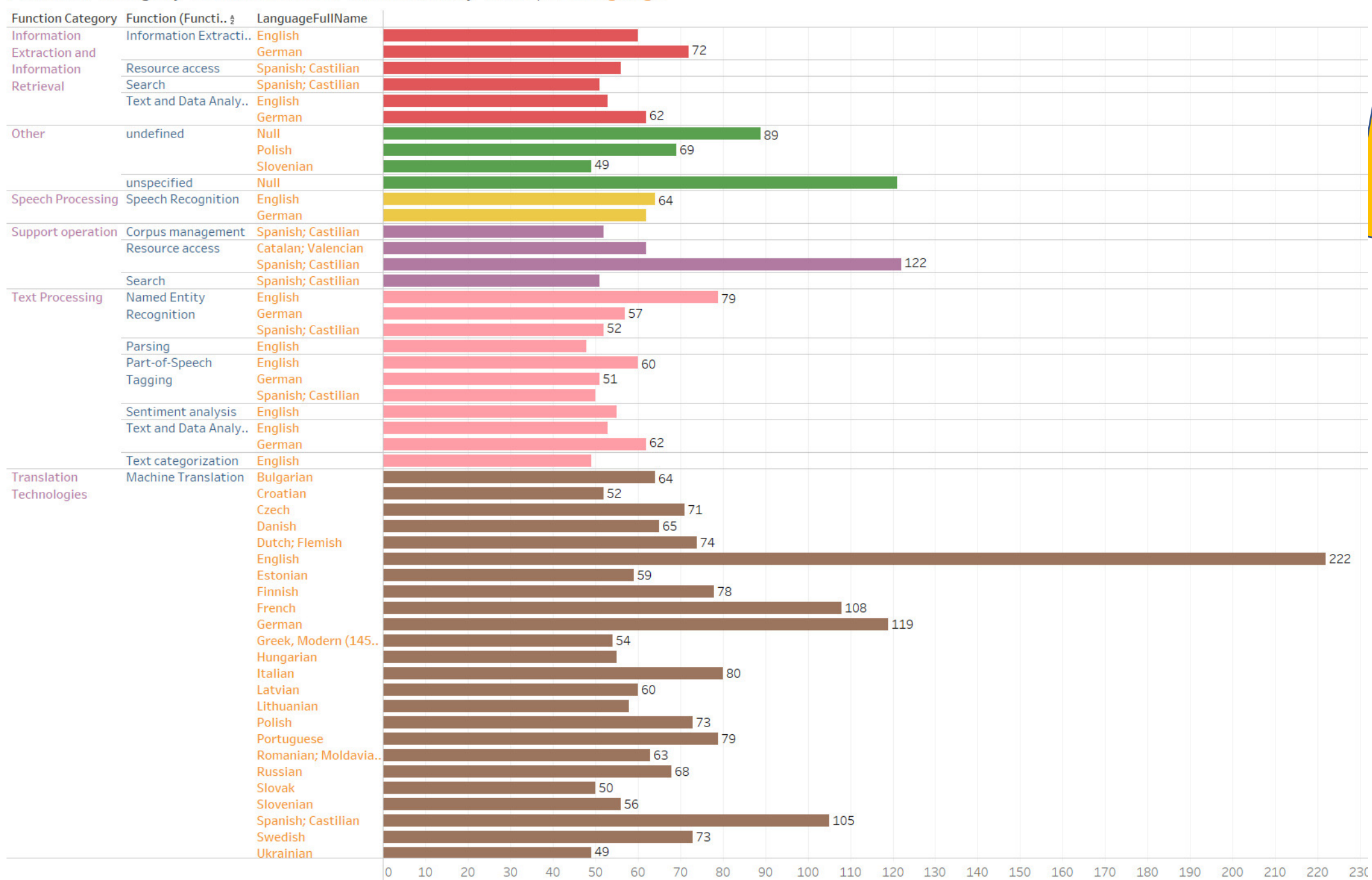
Conditions of Use

- attribution required
- commercial uses not allowed
- derivatives not allowed
- no conditions
- other specific restrictions
- research use allowed
- unspecified



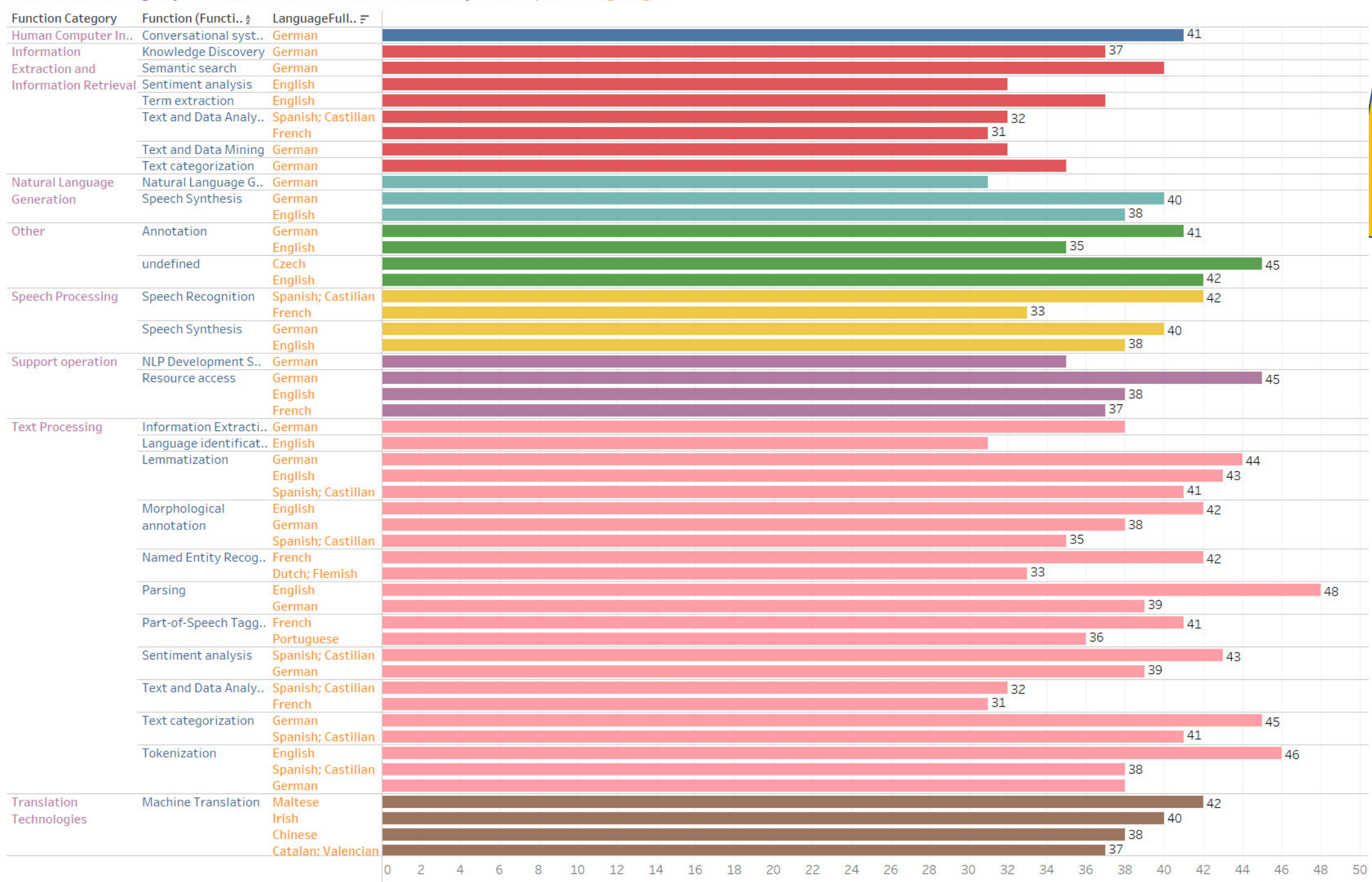


Function Category COLLECTIONS broken down by Tools per Language



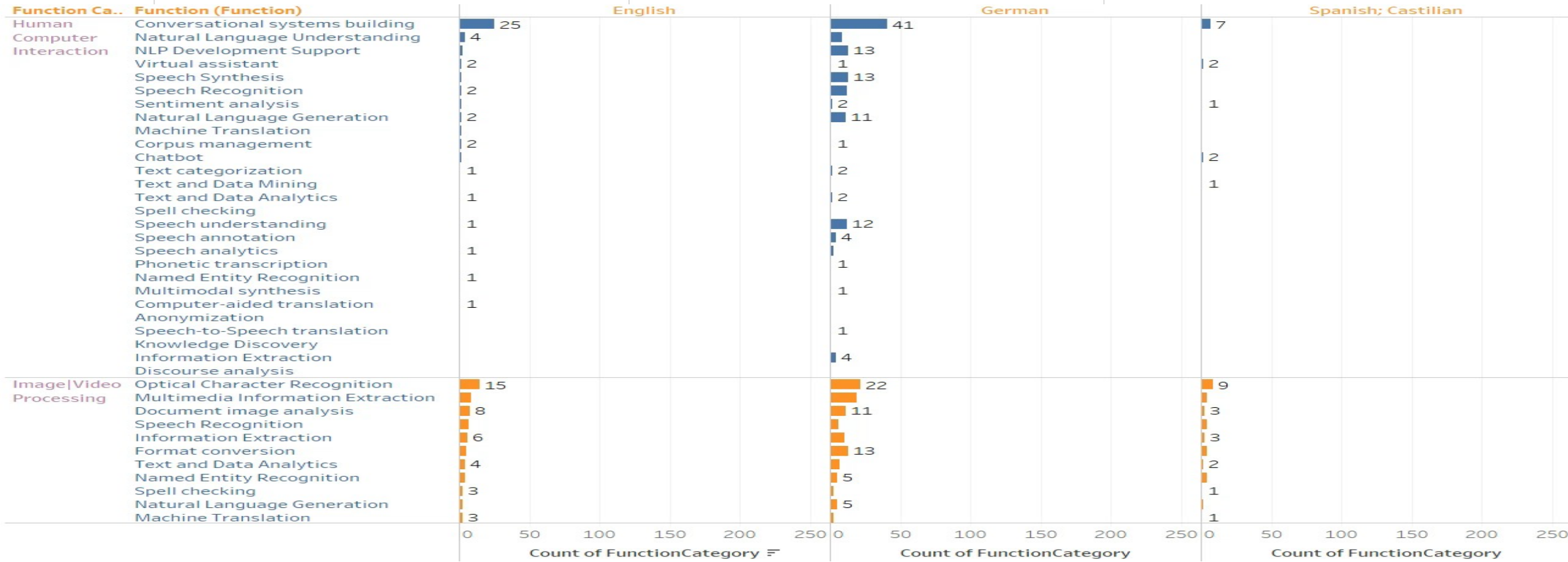
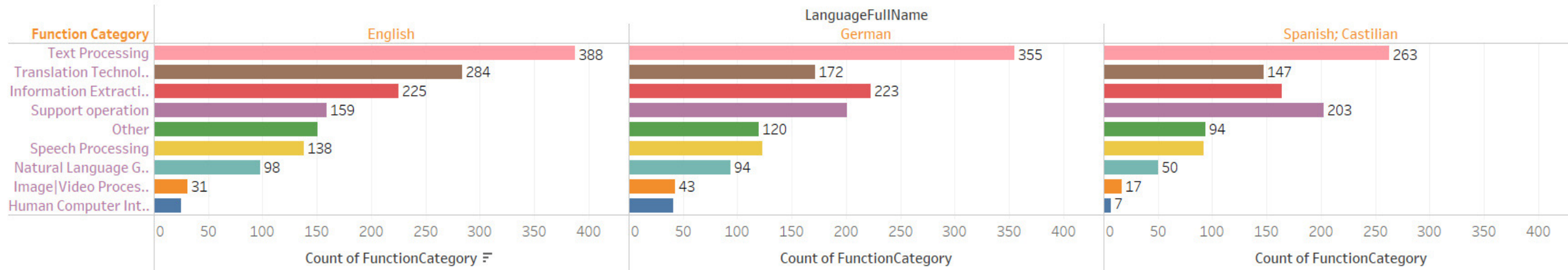
Languages with 49 and more resources for a Tool

Function Category COLLECTIONS broken down by Tools per Language

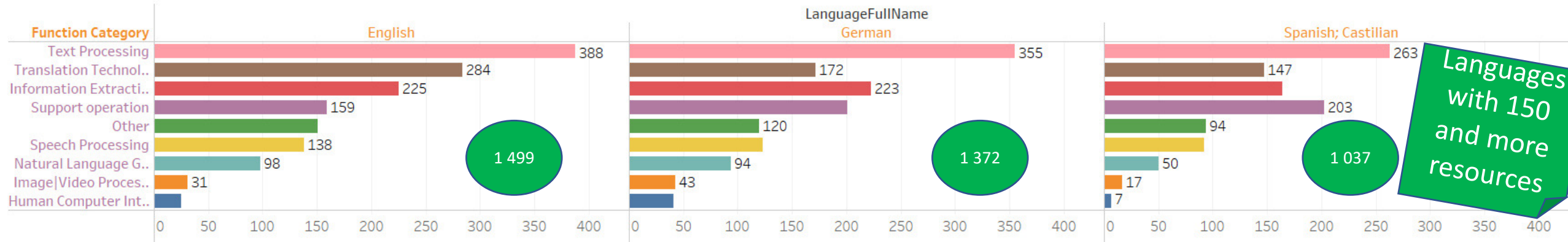


Languages with 31 to 48 resources for a Tool

Function Category COLLECTIONS broken down by Tools per Language

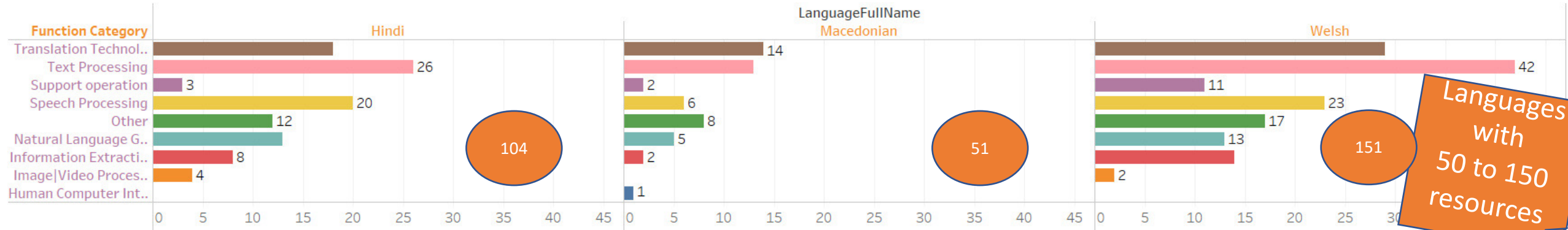


Function Category COLLECTIONS broken down by Tools per Language



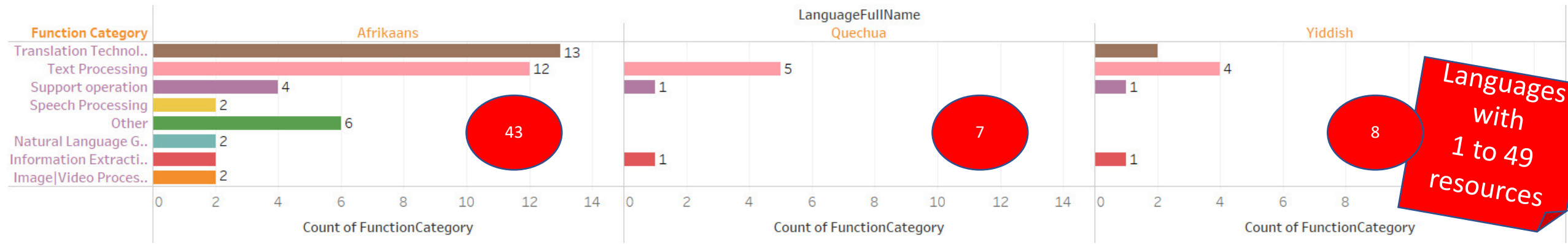
Languages with 150 and more resources

Function Category COLLECTIONS broken down by Tools per Language



Languages with 50 to 150 resources

Function Category COLLECTIONS broken down by Tools per Language



Languages with 1 to 49 resources



UniDive: WG3

TASK 2

Learning about **ELG** and
data behind it