Complementary Material TACL Submission 5931

Anonymous

Contents

Chapter 1

Research / State of the Art Dimension

1.1 Indicator D.1 Publication of Scientific Papers

Indicator D.1 Publication of scientific papers

It represents the percentage difference in terms of scientific publications in Spanish. It will be obtained as follows:

$$\mathrm{D.1} = \frac{|P_I| - |P_E|}{|P_E \cup P_I|} \cdot 100$$

where $|P_E|$ and $|P_I|$ represent the number of publications that cover Spanish and those that cover English only.

1.1.1 Data for calculations

Table 1.1 shows the data on which Indicator D.1 has been computed.

Table 1.1: Number of publications that describe experimentation for the languages considered at international conferences, by language and year, from 2018 to 2022.

	2018	2019	2020	2021	2022	Gap
			ACL			
Spanish	4	4	3	0	8	99%
French	1	3	2	0	2	99%
German	6	16	5	6	5	98%
Chinese	26	19	18	36	25	92%
English	346	618	751	668	662	
			SIGIR			
Spanish	0	0	0	0	0	100%
French	0	0	0	0	0	100%
German	0	0	0	0	0	100%
Chinese	0	0	10	2	6	98%
English	209	224	330	380	334	
			CONLL			
Spanish	1	0	2	0	-	97%
German	4	6	1	3	-	89%
French	0	1	1	2	-	97%
Chinese	5	3	0	3	-	91%
English	51	88	50	45	-	
			EACL			
Spanish	11	8	-	-	-	90%
German	18	19	-	-	-	81%
French	8	8	-	-	-	91%
Chinese	8	7	-	-	-	92%
English	75	285	-	-	-	
			NAACL			
Spanish	13	15	21	-	-	90%
German	22	19	17	-	-	88%
French	20	12	10	-	-	91%
Chinese	18	30	16	-	-	87%
English	133	348	414	-	-	
			Aggregated			
Spanish	29	27				95%
French	45	47				93%
German	34	37				94%
Chinese	57	59				91%
English	814	1563				0%

1.2 Indicator D.2 Funded Projects

Indicator D.2 Funded projects

It represents the percentage difference in terms of NLP projects subsidized in Spanish. It will be obtained as follows:

$$\mathrm{D.2} = \frac{|P_I| - |P_E|}{|P_E \cup P_I|} \cdot 100$$

where $|P_E|$ represents the number of funded projects for Spanish and $|P_I|$ represents the number of funded projects in English.

1.2.1 Data for calculations

Table 1.2 shows the data used to calculate Indicator D.2.

Table 1.2: Number of projects funded by CORDIS and NSF in English and Spanish from 2018 to 2022. Databases accessed on 30.10.2022.

	Spanish	English	Gap
CORDIS	3	52	89%
NSF	7	100	87%
Average			88 %

1.3 R.0 Indicator Text availability on the Internet

Indicador R.0 Text availability on the Internet

It represents the percentage difference in availability of text corpora in the respective languages. Let F be the set of sources (wikipedia, etc.), and let P_f be the weight assigned to source:

$$\mathrm{R.0} = \sum_{f \in F} P_f \frac{T_I^f - T_E^f}{T_I^f + T_E^f} \cdot 100$$

where T_I^f y $_E^f$ represent the volume of text in source f in English and Spanish respectively.

1.3.1 Data for calculations

Table 1.3 shows the data used to calculate Indicator R.0.

Table 1.3: Text volume on the Internet in English and Spanish. Consulted on 17.03.2023.

Collection	Size	Search Sp.	Spanish	Serch En.	English	Source	R.0
Wikipedia	# articles		1846819		6630966	https://es.wikipedia.org/wiki/Wikipedia_en_espa%C3%B1ol	77%
Internet	% pages		4.8		56.10	https://w3techs.com/technologies/history_overview/content_language	91%
Internet Archive	# texts	https://archive.org/search?query=Language%3A%28spanish%29	49193	https://archive.org/search?query=Language%3A%28english%29	7679285	https://archive.org/	99%
PubMed	# texts	https://pubmed.ncbi.nlm.nih.gov/?term=Spanish%5BLanguage%5D&sort=	380656	https://pubmed.ncbi.nlm.nih.gov/?term=English%5BLanguage%5D	30615956	https://pubmed.ncbi.nlm.nih.gov/	99%
Common Crawl: CC-MAIN-2023-06	% pages		4,5977		46,2733	https://commoncrawl.github.io/cc-crawl-statistics/plots/languages	90%

1.4 Indicator R.1 Language Models Availability

Indicator R.1 Language models availability

It represents the percentage difference in language model availability. Let D be the set of tasks or domains considered, and let P_d be the weight assigned to each task or domain:

$$\mathrm{R.1} = \sum_{d \in D} P_d \frac{|M_I| - |M_E|}{|M_I \cup M_E|} \cdot 100$$

where M_I and M_E represent the sets of models trained on English and Spanish text, respectively. Multilingual models will be considered as belonging to both sets.

1.4.1 Data for calculations

Table 1.4 shows the results obtained from the search described in the paper, by language and by task (searched on 6.11.23). Bilingual models (English/Spanish) are not included in the count of models by individual language. Therefore, the indicator variables have been set as follows: $|M_I|$ and $|M_E|$ have been calculated as the first or the second plus the third column respectively for each language. The variable $|M_I \bigcup M_E|$ has been estimated as the sum of the first two columns minus the third.

As the results show, taking into account the total number of models a gap of 86.03% is obtained. If we average the gaps obtained per NLP task, we obtain 76.02%. Specifically, as expected, the smallest gap is obtained in the case of machine translation, which always involves several languages (36%). If we exclude machine translation, we obtain an average indicator per task of 79.28%. We assign the same weight P_d to all task categories in Hugging Face. As the final value of the R.1 indicator, we consider the average per task, obtaining a gap of 76.02%.

Table 1.4: Results of Indicator R.1 based on data from Hugging Face.

	English	Spanish	En/Sp	R.1
Total	12282	1044	263	86,03
Per NLP Task				
Text generation	4052	59	35	96,31
Text clasification	1671	112	20	86,47
Text2Text	1203	77	39	85,37
Machine translation	578	257	36	36,85
Fill-mask	412	85	36	61,35
Token classification	340	77	18	60,46
Question answering	243	38	4	71,93
Summarization	190	27	8	72,44
Sentence similarity	81	10	2	76,34
Conversational	64	0	0	100,00
Zero-shot classification	39	8	2	63,27
Table question answering	37	1	0	94,74
Average per task				76,02
Average without machine translation				79,28

1.5 Indicator R.2.a Annotated Data in Repositories

R.2.a Annotated data in repositories

It represents the percentage difference in annotated corpus availability for training and testing in international forums (competitive evaluation campaigns and repositories). Let R be the set of repositories and P_r be the weight assigned to each repository:

$$\mathrm{R.2.a} = \sum_{r \in \mathcal{R}} P_r \frac{|D_I| - |D_E|}{|D_I \bigcup D_E|} \cdot 100$$

where D_I y D_E represent the sets of datasets in English and Spanish respectively in public repositories.

1.5.1 Data for calculations

Table 1.6 shows the data used to calculate Indicator R.2.a.

Table 1.5: Number of annotated datasets in repositories for Spanish and English.

	English	Spanish	En./Sp.	R.2.a
LRE				
Without date	1733	229		**76,66**
2018-2022	395	36		83,29
LCD				
Total	48	9		**68.42**
2019	10	3		53,85
2020	19	2		80,95
2021	12	4		50,00
2022	7	0		100,00
Average per year				64,68
Hugging Face				
Total	2307	268	171	84,82
By NLP Task:				
Text classification	490	60	33	83,17
Question answering	258	27	22	87,83
Text generation	247	37	30	82,68
Text2Text	171	17	15	89,02
Machine translation	153	66	59	54,38
Token classification	138	36	26	68,92
Text retrieval	134	17	15	86,03
Summarization	132	19	16	83,70
Fill-mask	108	34	31	66,67
Multiple choice	103	11	11	89,32
Zero-shot classification	82	9	9	89,02
Table question answering	73	9	9	87,67
Conversational	49	9	9	81,63
Sentence similarity	36	9	9	75,00
Table2text	28	11	11	60,71
Average per task				**79,05**
Average per repository				**74,71**

1.6 Indicator R.2.b Annotated Data in Evaluation Campaigns

Indicator R.2.b Annotated data in evaluation campaigns

It represents the percentage difference in availability of corpus annotated for training and test in competitive evaluation campaigns. Let $mathcalC$ be the set of evaluation campaigns and P_c be the weight assigned to each campaign:

$$\mathrm{R.2.b} = \sum_{c \in \mathcal{C}} P_c \frac{|T_I| - |T_E|}{|T_I \bigcup T_E|} \cdot 100$$

where T_I y T_E represent the set of tasks in the evaluation campaigns with data annotated in English and Spanish respectively.

1.6.1 Data for calculations

Table 1.6 shows the data used to calculate Indicator R.2.b. Only the data for English and Spanish have been used.

Table 1.6: Number of datasets annotated in different languages for the evaluation campaigns SemEval, CLEF and IberLEF from 2018 to 2022.

	2018	2019	2020	2021	2022	R.2.c
	SemEval-tasks					
Spanish	2	1	1	0	1	81%
German	0	1	1	0	1	88%
French	0	1	0	1	2	85%
Chinese	0	0	0	1	1	92%
English	10	8	10	9	12	
	CLEF					
Spanish	1	0	2	3	2	75%
French	2	0	1	0	1	87%
German	1	1	1	1	2	81%
Chinese	0	0	0	0	0	100%
English	9	9	12	12	14	
	IberLEF					
Spanish	6	8	7	11	9	-86%
English	1	0	0	1	1	
	Total					
Spanish	9	9	10	14	12	-33%
English	20	17	22	22	27	
Average gap English/Spanish in campaigns						23%

1.7 Indicator E.1 Effectiveness in NLP Tasks

Indicator E.1 Effectiveness in NLP tasks

It represents the difference between languages between percentage improvements over a non-linguistic baseline system. Let $\mathcal{D}$ be the set of domains, P_d be the weight assigned to each domain, and $\mathcal{H}_d$ be the set of applications selected for that domain, it is computed:

$$\mathrm{E.1} = \sum_{d \in \mathcal{D}} P_d \sum_{h \in \mathcal{H}_d} \frac{s_I^h - b_I^h}{|b_I^h - r^h|} - \frac{s_E^h - b_E^h}{|b_E^h - r_0^h|}$$

where s_I^h, b_I^h and r^h represent the effectiveness of tool h, of the bases system and the point of reference in English. The notation for Spanish is analogous.

1.7.1 Evaluated models

The models that have been evaluated are the following ones, all of them available in HuggingFace:

English models

- **Bert-base-cased** Bert (Bidirectional Encoder Representations from Transformers) (Devlin et al., 2018).
- **Bert-base-uncased**.
- **RoBERTa-base** RoBERTa (Liu et al., 2019).
- **Roberta-large**.
- **Distilbert-base-uncased** (Sanh et al., 2019)..

Spanish models

- **Roberta-base-bne**.
- **Roberta-large-bne**.
- **Bertin-roberta-base-spanish**.
- **Bert-base-spanish-wwm-cased**.
- **Distillbert-base-spanish-uncased**.

Bilingual models

- **Xlm-roberta-base** (Conneau et al., 2019).
- **Xlm-roberta-large**.
- **Bert-base-multilingual-cased**.
- **Distilbert-base-multilingual-cased**.

1.7.2 Evaluated tasks

The tasks selected are intended to cover a variety of Natural Language problems (classification, sequence labelling, question answering, regression) and tasks (toxic content, biomedical information extraction, disinformation, reading comprehension, news classification, ...).

- **EXIST Task 1** on sexism detection[1]. This is a binary classification task of tweets consisting on determining whether a tweet is sexist or not.

[1] http://nlp.uned.es/exist2022/

- **EXIST Task 2**: Once a message has been classified as sexist, the second task aims to categorize the message according to the type of sexism: ideological and inequality, stereotyping and dominance, objectification, sexual violence, misogyny and non-sexual violence. This is a multiclass classification task.
- **DIPROMATS Task 1**[2], a binary classification task that consists on identifying whether messages from diplomats and authorities from USA, China, Russia and the EU contain propaganda techniques.
- **DIPROMATS Task 2** consists in categorizing propagandistic messages in four types of propaganda plus a negative class. This is a hierarchical, multiclass, multilabel problem.
- **DIPROMATS Task 3**: this task is similar to the previous one, but adding a fine-grained distinction of 15 types of propaganda techniques.
- **DIANN Task 1** [3] is a bilingual Named Entity Recognition task on biomedical scientific abstracts.
- **DIANN Task 2** on detecting the scope of negation in biomedical abstracts.
- **MLDoc** The Multilingual Document Classification Corpus is a standard classification dataset for news in several languages (Schwenk and Li, 2018).
- **MultiCONER 2022** (Malmasi et al., 2022) is a multilingual dataset for complex named entity recognition.
- **STS-2017** (Cer et al., 2017) is a multilingual dataset for textual similarity. It is the only regression task in our experimental design, and also the only one that does not use a variant of F_1 as evaluation metric, but the Pearson correlation between system predictions and manual annotations.
- **SQUAD/SQAC**: These are two datasets build independently, but with the same methodology. SQAC (Spanish Question Answering Corpus) (Gutiérrez-Fandiño et al., 2021) is an extractive QA dataset where, given a question and an associated paragraph, the system must locate the shortest text span that contains the answer. It has been designed with the same methodology of its predecessor, SQuAD v1.1 (Rajpurkar et al., 2016).

1.7.3 Data for calculations

In Table 1.7 the results of our experimentation on the eleven tasks are presented. We have eleven gap measurements on the eleven tasks considered; in each of them, the baselines have been calculated as the average of several supervised learning algorithms without linguistic knowledge, as explained above. The columns *best ES* and *best EN* show the effectiveness of the best performing language model in each task in Spanish and English, respectively.

[2] https://sites.google.com/view/dipromats2023/home
[3] https://nlp.uned.es/diann/

Task	**baseline SP**	**Baseline EN**	**best SP**	**best EN**	**Gap %**
EXIST Task 1	0,693	0,674	0,751	0,752	4,99
EXIST Task 2	0,463	0,437	0,573	0,603	9,26
DIPROMATS Task 1	0,750	0,707	0,815	0,808	8,60
DIPROMATS Task 2	0,220	0,210	0,430	0,510	47,40
DIPROMATS Task 3	0,090	0,080	0,250	0,390	209,72
DIANN Task 1	0,747	0,665	0,812	0,811	17,67
DIANN Task 2	0,878	0,575	0,957	0,917	16,03
MLDoc	0,930	0,883	0,970	0,978	23,61
MultiCoNER 2022	0,800	0,803	0,852	0,885	15,45
STS-2017	0,680	0,707	0,828	0,862	6,71
SQUAD/SQAC	0,533	0,528	0,801	0,921	25,91
Efectiveness gap					**18 $\pm$ 04**

Table 1.7: Calculation of the effectiveness gap of language models on Spanish and English datasets. Baselines have been calculated as the average of several supervised learning algorithms without linguistic knowledge. LLMs are the effectiveness of the best performing language model on each task.

Chapter 2

Functionalities of Market Solutions Dimension

2.1 Indicator S.1 Functionalities of Market Solutions

For every family of applications Indicator I.S.1 is calculated as follows. Be F_I^p y F_E^p the set of functionalities present in the commercial product p developed for English and Spanish respectively::

$$\mathrm{I.\,S.\,1}(p) = \frac{|F_I^p \setminus F_E^p| - |F_E^p \setminus F_I^p|}{|F_I^p \cup F_E^p|} \cdot 100$$

This definition fulfills the following properties. The gap is zero if both languages offer the same functionalities:

$$F_I^p = F_E^p \Longrightarrow \mathrm{I.\,S.\,1}(p) = 0$$

The indicator is symmetrical with respect to languages. In addition, the gap is 100 percent only if none of the functionalities are covered in English.

$$F_E^p = \emptyset \Longleftrightarrow \mathrm{I.\,S.\,1}(p) = 100\%$$

Given a fixed difference in functionalities in both directions, the gap is inversely proportional to the number of functionalities present in either language:

$$|F_I^p \setminus F_E^p| = k \wedge |F_E^p \setminus F_I^p| = k' \Longrightarrow \mathrm{I.\,S.\,1}(h) \propto \frac{1}{|F_I^p \cup F_E^p|}$$

In case the product covers all the functionalities in English, then the indicator will be proportional to the amount of functionalities not covered in Spanish:

$$F_I^p \supset F_E^p \Longrightarrow \mathrm{I.\,S.\,1}(p) \propto |F_I^p| - |F_E^p|$$

Indicator S.1: Functionalities of Market Solutions

It represents the difference in terms of functionalities offered by the products in both languages. Let H be the set of application families considered in the project, let W_h be the weight assigned to one of the application families, and let P_h be the set of products identified for that family of tools:

$$\mathrm{I.\,S.\,1} = \sum_{h \in H} W_h \frac{1}{|P_h|} \sum_{p \in P_h} \frac{|F_I^p \setminus F_E^p| - |F_E^p \setminus F_I^p|}{|F_E^p \cup F_I^p|}$$

F_E^p and F_I^p represent the set of functionalities present in the product p in English and Spanish respectively.

2.2 List of Functionalities per NLP Area

In this section we list the functionalities that we have analyzed for each of the NLP areas. Those functionalities in which a (+1) or (+2) appears next to the name contain, under that same definition, one or two additional functionalities respectively, bringing the total to two or three. If a (+n) appears, it means that the total number of additional functionalities under that definition is not limited. In each of the descriptions the additional functionalities are detailed.

2.2.1 Opinion Analysis

- Sentiment classification (+1): Ability to classify the message according to the polarity of the sentiment shown by the source. The ability to classify neutral messages as a specific functionality will be assessed.
- Reputational impact classification (+1): Ability to classify the message according to the impact it has on a specific entity. Ability to classify neutral messages as a specific functionality will be assessed.
- Emotion classification (+n): Detection of the type of emotion displayed (anger, disgust, happiness etc.). Each of the classes it can detect will count as a functionality.
- Conversation topic detection: Detection of the conversation topic.
- Detection of inappropriate messages (+n): Detection of messages that may be harmful or not suitable for all audiences: hate messages, harassment, pornographic messages, etc. Each of the detected sections will count as a functionality.
- Entity detection: Ability to detect entities (names of brands, products or people).
- Detection of motivations (psychological drivers): Ability to detect motivations or stimuli of the message senders, intentions or possible habits.
- Attribute sentiment scoring: Sentiment analysis on an aspect or attribute of the product mentioned in the text or indicated beforehand.
- Sentiment by entities (+1): Sentiment analysis on each of the entities identified in the text. Like the reputational classification, it implies directionality, but in this case the detection of entities is done automatically.
- Possibility of defining classes (+n): Possibility for the user to define the classes to be detected. They could be emotions, sentiment or reputational polarity, topics, etc.
- Possibility of adjusting the model: Possibility of adjusting or training the model to adapt it to a certain theme or criterion.

2.2.2 Virtual Assistants

- Voice recognition: Ability to recognize who is speaking by voice recognition.
- Ability to add skills (+n): Ability to interact with other services such as diaries, applications. Each of the relevant services detected will count as a functionality.
- Regional accents (+n): Ability to communicate with regional accents and vocabularies. Each of the regional accents will count as a functionality.
- Accepted commands (+n): The commands understood by the wizard in each of the languages. For example, set the alarm or know what the weather is going to be. Each relevant command will count as a functionality.
- Text typing capability: Ability to translate the dictated conversation into text.
- Other functionalities (+n): Other non-general functionalities such as entity recognition, stemming, auto-correction, etc.

2.2.3 Machine translation

- Grammar correction: If the translator has grammar autocorrection.
- Languages from which it can translate (+n): When translating to English or Spanish, the number of languages from which it can translate. Each language will count as one feature.
- Language detection: When the source language is English or Spanish, if it is able to detect it.
- Possibility to modify translations (+1): If it allows to modify the translated text a posteriori for personal use or if it allows to modify the text as a suggestion for the translator. Each one counts as a feature.
- Translation versions: If the translator suggests different versions for the translation.
- File translation: Ability to translate entire files without altering file formats.
- Web translation: Ability to translate entire web pages without altering the layout of the web pages.
- Text to text: Text to text translation.
- Text-to-speech: Text-to-speech translation.
- Speech to text (+1): Speech to text translation. Counts as additional functionality if you can do it in real time.
- Speech to speech: Speech to speech translation. Word to text images (+1): Ability to detect text in images and translate it. Counts as additional functionality if it can do it in real time.
- Domain adaptation: If the translator can automatically adapt to different domains.
- Domain customization: Possibility of adding terms and training data to adapt the translator to specific domains.
- Regional variants (+n): Whether it offers different translations for regional variants. It will count as one feature per variant.

2.2.4 Predictive keyboards

- Grammar correction: Whether the predictive keyboard has grammar auto-correction.
- Personalized prediction: If the prediction of the predictive keyboard is adjusted to the user's way of writing.
- Language detection: When the starting language is English or Spanish, if it is able to detect the same.
- Word auto-complete: Ability to auto-complete the initiated word.
- Speech to text: Speech to text typing.
- Word suggestions: Ability to suggest the next word to type.
- Text generation (+2): If it can suggest, not only words, but complete texts or snippets. It will be one functionality if it can suggest expressions, another if it can suggest complete sentences and another if it can suggest paragraphs or longer texts.

2.2.5 Web search

- Grammar correction: If the search engine has grammar autocorrection. Meaning detection: If the search engine has NLP processes that search by meaning of the phrase and not by words.
- Topic classification (+n): Ability to classify documents by topic. If it is able to perform classifications in different areas (subject, purpose etc.), they will count as an additional functionality. For example, if it is able to identify the type of page (forum, blog post, news, etc.) it is an area. If, in turn, it is able to identify the topic of the text (sports, politics etc.) it is another area of classification. As each area is worth 1, in this case I would put a 2 because it classifies in two areas or dimensions.
- Entity detection: Whether it is able to perform entity detection (names, sites, companies, etc.) in documents.
- Synonym search: Ability to use synonyms in addition to the words that have been included in the search.
- Image search: Ability to search for images based on a text entered in the search engine.
- Search for text in images: Ability to search for text in documents in image format or in images with text.
- Video search: Ability to search for videos based on a text entered in the search engine.
- Audio search: Ability to search in audio files based on a text entered in the search engine.
- Answer search: Ability to directly return an answer instead of a list of URLs.
- Information search: Ability to return structured information instead of a list of URLs (eg Google knowledge panel).

2.3 List of Commercial Products per Area

The list of the products that have been analyzed is presented below.

2.3.1 Analysis of opinions

Source: The Forrester New Wave^{TM} AI-Enabled Consumer Intelligence Platforms, Q3 2021. https://www.talkwalker.com/case-studies/forrester-new-wave-report-2021.

- Sprinklr (revenue: 387 M$)
- Khoros (revenue: 200 M$)
- SNetBase Quid (revenue: 38.2 M$)
- SBrandwatch (revenue: 22.6 M$)
- SLinkfluence (revenue: 20 M$)
- SSynthesio (revenue: 27 M$)
- STalkwalker (revenue: 25.4 M$)
- SDigimind (revenue: 18 M$)
- SResonate (revenue: 17 M$)
- SMeltwater (p.k.a Sysomos) (revenue: 22.6 M$)

2.3.2 Virtual Assistants

- Google Assistant (2,500M mobiles + Smart Speakers (20% share) https://www.businessofapps.com/data/android-statistics/ https://www.statista.com/statistics/792604/worldwide-smart-speaker-market-share/)
- Siri (18% mobiles in the world https://www.counterpointresearch.com/global-smartphone-share/ (1,000 M users))
- Alexa (+100M users https://expandedramblings.com/index.php/amazon-alexa-statistics/)
- Bixby (+200M users https://www.samsungmobilepress.com/press-releases/samsung-further-develops-bixby-introducing-a-new-language-and-setting-a-foun)
- Cortana (+500M users, +18,000M consults https://news.microsoft.com/bythenumbers/en/cortana)

Source: Cloud based voice service/chatbots/conversational AI platforms. https://www.boost.ai/reports/gartner-magic-quadrant-for-enterprise-conversational-ai-platforms.

- Kore.ai (revenue: 270M USD)
- IBM Watson Assistant (leader per market quota in AI platforms https://www.ibm.com/blogs/journey-to-ai/2020/09/ibm-secures-fifth-consecutive-year-of-ai-software-platform-market-share-lead
- Amazon Lex
- Google Dialogflow
- Amelia (revenue: 61M USD)

2.3.3 Machine translation

Source: https://greatcontent.com/machine-ai-translation-tools/(statistics from https://www.similarweb.com/).

- Google Translate (729M visits/month)
- DeepL (258M visits/month)
- Microsoft Translator — Bing Translator (100M visits/month)
- Systran Translate (404K visits/month)
- Amazon Translate
- Reverso Translation (107M visits/month)
- memoQ Translator PRO (367K visits/month)
- Smartling (605K visits/month)
- Crowdin (1,9M visits/month)
- TextUnited (18,2K visits/month)

2.3.4 Predictive keyboards

- Microsoft SwiftKey (95,3K ratings in the App Store, 3,83M ratings in Google Play)
- GBoard (38,8K App Store, 10,9M Google Play)
- Grammarly (functionality of sentence prediction 42,8K App Store, 147K Google Play)
- Fleksy (670 App Store, 270K Google Play)
- Phone's default keyword (18% of mobiles in the worl https://www.counterpointresearch.com/global-smartphone-share/ (1,000 M users)).
- Phraseboard (50K Google Play).
- GMail (predictive functions in writing. +1,800M users https://financesonline.com/number-of-active-gmail-users/).
- Google Workspaces (predictive functions in writing. +126M paying users https://en.wikipedia.org/wiki/Google_Workspace , +3,000M users https://developers.googleblog.com/2022/01/year-in-review-google-workspace.html).
- Microsoft Outlook (predictive functions in writing, +500M users https://en.wikipedia.org/wiki/Outlook.com).
- Microsoft Office 365 (predictive functions in writing. +345M paying users https://office365itpros.com/2022/04/28/office-365-number-of-users/).

2.3.5 Web search

Source: https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines.

- Google Search (83,8%)
- Bing (8,9%)
- Yahoo Search (2,6%)
- DuckDuckGo
- Brave Search

Source: the two best positioned in Gartner's quadrant. https://www.coveo.com/en/resources/reports/gartner-magic-quadrant-for-insight-engines.

- Elasticsearch
- Mindbreeze
- Apache Solr

2.4 Data for calculations

The tables below contain the data that have been used to calculate the value of the indicator. For every application area the functionalities are listed as well as the number of functionalities per commercial product and language.

Table 2.1: Opinion analysis. Number of functionalities per product and language.

	netbase-quid		brandwatch		linkfluence		synthesio		talkwalker		digimind		meltwater	
	ES	EN	ES	EN	ES	EN	ES	EN	ES	EN	ES	EN	ES	EN
Sentiment classification	2	2	1	1	2	2	2	2	2	2	2	2	2	2
Reputation impact classification	1	1	0	0	1	1	1	1	1	1	1	1	0	0
Emotion classification	8	8	0	0	1	1	5	5	1	1	25	25	6	6
Bots detection	0	0	0	0	0	0	0	0	1	1	0	0	1	1
Entity detection	1	1	1	1	1	1	1	1	1	1	0	0	1	1
Inappropriate message detection	1	1	1	1	0	0	1	1	1	1	0	0	0	0
Motivation detection	0	0	1	1	1	1	1	1	1	1	1	1	1	1
Conversation topic detection	0	0	1	1	1	1	1	1	1	1	1	1	1	1
Adjusting the model	0	0	0	0	0	0	1	1	1	1	1	1	0	0
Defining classes	0	0	0	0	0	0	0	0	1	1	0	0	0	0
Aspect based sentiment	0	0	0	0	0	0	1	1	1	1	0	0	0	0
Sentiment per entity	0	0	0	0	1	1	0	0	1	1	0	0	1	1

Table 2.2: Virtual assistants. Number of functionalities per product and language.

	google ass.		siri		alexa		bixby		cortana		chatgpt		koreai		ibm-watson-ass.		amazon-lex		google-dialogflow		amelia	
	ES	EN	ES	EN	ES	EN	ES	EN	ES	EN	ES	EN	ES	EN	ES	EN	ES	EN	ES	EN	ES	EN
Regiional accents	6	10	4	9	3	5	2	3	2	5	0	0			4	13	3	5	3	6		
Capacity to write text	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1
Accepted commands	32	32			46	71	97	111	0	0	0	0					0	0	0	0		
Other functionalities															9	12	0	0	2	2		
Possibility to add capabilities					1	1	1	1	0	0	0	0	19	19	1	1	0	0	1	1		
Voice recognition	1	1	1	1	1	1	1	1	0	0	0	0	1	1	0	0	0	0	0	0	1	1

Table 2.3: Machine translation. Number of functionalities per product and language.

	google-trans		deepl		bing-trans		systran-trans		amazon-trans		reverso		memoq		smartling		crowdin		textunited	
	ES	EN	ES	EN	ES	EN	ES	EN	ES	EN	ES	EN	ES	EN	ES	EN	ES	EN	ES	EN
Domain adaptation			1	1			1	1	0	0	1	1	0	0	0	0	1	1	0	0
Grammar correction	1	1	0	0	1	1	0	0	0	0	1	1	1	1	0	0	1	1	1	1
Word images to text	1	1	1	1	2	2	1	1	0	0	0	0	0	0	0	0	1	1	0	0
Text to text	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Text to voice	1	1	0	0	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0	0
Voice to text	2	2	1	1	2	2	1	1	0	0	1	1	0	0	0	0	0	0	1	1
Voice to voice	1	1	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0
Language detection	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	1
Source languages	132	132	28	28	110	110	52	52	74	74	23	23	254	254	523	523	313	313	170	170
Personalization to domains	0	0	1	1	1	1	1	1	0	0	0	0	1	1	1	1	1	1	0	0
Optionn to modify translations	0	0	2	2	1	1	0	0	1	1	1	1	2	2	0	0	1	1	0	0
File translation	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Web translation	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	1	1	1	0	0
Regional variants									2	1							22	26	20	4
Versions of translation	0	0	1	1	0	0	0	0	0	0	1	1	1	1	0	0	1	1	0	0

Table 2.4: Predictive keyboards. Number of functionalities per product and language.

	microsoft-swiftkey		gboard		grammarly		fleksy		iphone-keyboard		gmail		google-workspaces		microsoft-outlook		microsoft-office-365	
	ES	EN	ES	EN	ES	EN	ES	EN	ES	EN	ES	EN	ES	EN	ES	EN	ES	EN
Word autocompletion	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
Grammar correction	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
Voice to text	1	1	1	1	0	1	1	1	1	1	0	0	1	1	1	1	1	1
Language detection	1	1	1	1	0	1	0	0	1	1	1	1	1	1	1	1	1	1
Text generation	1	1	1	1	0	2	2	2	0	0	2	2	2	2	2	2	2	2
Personalized prediction	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	0	0	0
Word suggestion	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 2.5: Search engines.Number of functionalities per product and language.

	google-search		bing		yahoo-search		duckduckgo		brave-search		elasticsearch		mindbreeze		apache-solr	
	ES	EN	ES	EN	ES	EN	ES	EN	ES	EN	ES	EN	ES	EN	ES	EN
Audio search	1	1	1	1	0	0	0	0	1	1	0	0	0	0	0	0
Image search	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	0
Information search	1	1	1	1			1	1	1	1	0	0	0	0	0	0
Answers search	1	1	1	1	1	2	1	1	1	1	0	0	1	1	0	0
Synonyms search	1	1	1	1	0	1	0	1	0	0	0	0	0	0	0	0
Search of text in images			1	1	0	0	0	0	1	1	0	0	0	0	0	0
Video search	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
Topic classification					2	3			2	2	1	1	0	0	0	0
Grammar correction	1	1	1	1	0	1	1	1	1	1	0	0	0	0	1	1
Entity detection	1	1	1	1	1	1			0	0	0	0	1	1	0	0
Meaning detection	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1

Chapter 3

Adoption Level Dimension

3.1 List of Companies

Table 3.1 shows the companies for which the level of adoption will be measured (indicators A.1, A.2 and A.3) and the selection criteria used. The companies with the largest market capitalization in the IBEX-35 and S&P 500 indices have been selected. The aim is to ensure that large companies in Spain and the USA, respectively, are represented in the study.

Table 3.1: List of companies to measure the level of adoption.

	IBEX 35	S&P 500
1	Inditex	Apple
2	Iberdrola	Microsoft
3	Santander	Alphabet
4	BBVA	Amazon
5	Caixabank	Berkshire Hathaway
6	Naturgy	UnitedHealth Group
7	Amadeus	Johnson & Johnson
8	Cellnex	Exxon Mobil
9	ArcelorMittal	Tesla
10	Telefonica	Visa
11	Repsol	NVIDIA
12	Endesa	Walmart
13	Ferrovial	JPMorgan Chase
14	Aena	Lilly
15	Siemens Gamesa	Procter & Gamble
16	Acciona	Chevron
17	Red Eléctrica	MasterCard
18	IAG	Home Depot
19	Grifols	Meta
20	ACS	Pfizer

3.2 Search Terms

The search terms used to find mentions are listed in Table 3.2.

Table 3.2: Terms used to search for mentions of NLP technologies.

Technology	English	Spanish	Source
Linguística de corpus	annotated corpora	corpus anotados	LREC Keywords
	annotated corpus	corpus anotado	LREC Keywords

	corpora annotation	anotacion de corpus	LREC Keywords
	corpus annotation		LREC Keywords
	corpus linguistics	linguistica de corpus	LREC Keywords
	english corpora	corpus en ingles	LREC Keywords
	english corpus		LREC Keywords
	spanish corpora	corpus en español	LREC Keywords
	spanish corpus		LREC Keywords
	text corpora	corpus de texto	LREC Keywords
	text corpus		LREC Keywords
	treebank		(Bikel and Zitouni, 2012)
	treebanks		(Bikel and Zitouni, 2012)
Sistemas de diálogo	spoken dialog system	sistema de dialogo hablado	(Bikel and Zitouni, 2012)
	spoken dialog systems	sistemas de dialogo hablado	(Bikel and Zitouni, 2012)
	chatbot		(Jurafsky and Martin, 2023)
	chatbots		(Jurafsky and Martin, 2023)
	dialogue modeling	modelado de dialogo	
		modelado de dialogos	TACL Areas of Interest
Semántica del discurso	argument mining	mineria de argumentos	Wikipedia - Common NLP tasks
	implicit semantic role labelling	etiquetado implicito de roles semanticos	Wikipedia - Common NLP tasks
	topic boundary detection	deteccion de limite de tema	(Bikel and Zitouni, 2012)
Menciones de PLN no específicas	computational linguistics	linguistica computacional	Wikipedia
	language technologies	tecnologias de lenguaje	
		tecnologias del lenguaje	Wikipedia
	language technology	tecnologia de lenguaje	
		tecnologia del lenguaje	Wikipedia
	natural language processing	procesamiento de lenguaje natural	
		procesamiento del lenguaje natural	Wikipedia
Sistemas generativos	natural language generation	generacion de lenguaje natural	
		generación de lenguajes naturales	LREC Keywords
	text to image generation	generacion de imagenes a partir de texto	
		generacion de imagenes a partir de textos	Wikipedia - Common NLP tasks
	text to scene generation	generacion de escenas a partir de texto	
		generacion de escenas a partir de textos	Wikipedia - Common NLP tasks
	text to video generation	generacion de video a partir de texto	
	generacion de videos a partir de texto		
	generacion de video a partir de textos		
		generacion de videos a partir de textos	Wikipedia - Common NLP tasks
Modelado de lenguaje	fill mask model	modelo fill mask	Hugging Face Tasks
	fill mask models	modelos fill mask	Hugging Face Tasks
	sentence similarity model	modelo de similitud de oraciones	
		modelo de similitud de frases	Hugging Face Tasks
	sentence similarity models	modelos de similitud de oraciones	
		modelos de similitud de frases	Hugging Face Tasks
	language based ai	ai basado en lenguaje	TACL Areas of Interest
	language modeling	modelado de lenguaje	
		modelado del lenguaje	TACL Areas of Interest
	word embedding		TACL Areas of Interest
	word embeddings		TACL Areas of Interest
	language model	modelo de lenguaje	LREC Keywords
	language models	modelos de lenguaje	LREC Keywords
Semántica léxica	terminology extraction	extractores de terminologia	Wikipedia - Common NLP tasks
	word sense disambiguation	desambiguacion linguistica	Wikipedia - Common NLP tasks
	morphological induction		B(Bikel and Zitouni, 2012)
Análisis morfológico	morphology induction	inducción morfologica	(Bikel and Zitouni, 2012)
	lemmatization	lematizacion	Wikipedia - Common NLP tasks
	lemmatisation		Wikipedia - Common NLP tasks
	morphological segmentation	segmentacion morfologica	Wikipedia - Common NLP tasks
	part of speech tagging	etiquetado de parte del discurso	
		etiquetado lexico	Wikipedia - Common NLP tasks
	pos tagging		Wikipedia - Common NLP tasks
Reconocimiento de entidades	entity linking	enlazamiento de entidades	Wikipedia - Common NLP tasks
	named entity recognition	reconocimiento de entidades nombradas	Wikipedia - Common NLP tasks
Comprensión del lenguaje natural	deep linguistic processing	procesamiento linguistico profundo	Wikipedia
	natural language programming	programacion de lenguaje natural	

		programacion del lenguaje natural	Wikipedia
	natural language understanding	comprension de lenguaje natural	
		comprension del lenguaje natural	Wikipedia
Reconocimiento óptico de caracteres	optical character recognition	reconocimiento optico de caracteres	Wikipedia - Common NLP tasks
Respuesta a preguntas	question answering model	modelo de respuesta a preguntas	Hugging Face Tasks
	question answering models	modelos de respuesta a preguntas	Hugging Face Tasks
	automatic question answering	respuesta automatica a preguntas	
		respuesta automatica de preguntas	Wikipedia - Common NLP tasks
Semántica relacional	computational semantics	semántica computacional	(Jurafsky and Martin, 2023)
	semantic parsing	parseo semantico	Wikipedia - Common NLP tasks
	semantic role labelling	etiquetado de roles semanticos	Wikipedia - Common NLP tasks
Análisis de sentimientos	opinion mining	mineria de opinion	
		mineria de opiniones	TACL Areas of Interest
	sentiment analysis	analisis de sentimiento	Wikipedia - Common NLP tasks
	sentiment classification	clasificacion de sentimiento	Wikipedia - Common NLP tasks
Lingüística estadística	statistical linguistics	linguistica estadistica	LREC Keywords
	quantitative linguistics	linguistica cuantitativa	Wikipedia
Generación de resúmenes	summarization model	modelo de sumarizacion	Hugging Face Tasks
	summarization models	modelos de sumarizacion	Hugging Face Tasks
	automatic summarization	resumen automatico	Wikipedia - Common NLP tasks
Análisis sintáctico	sentence boundary detection	deteccion de limimtes de oraciones	
		deteccion de limites de frases	(Bikel and Zitouni, 2012)
	grammar induction	induccion gramatical	Wikipedia - Common NLP tasks
	grammar inference	inferencia gramatical	Wikipedia - Common NLP tasks
	sentence boundary disambiguation	desambiguacion de limites de oraciones	
		desambiguacion de limites de frases	Wikipedia - Common NLP tasks
	syntactic parsing	parseo sintactico	Wikipedia - Common NLP tasks
Análisis de texto	text classification model	modelo de clasificacion de texto	
		modelo de clasificacion de textos	Hugging Face Tasks
	text classification models	modelos de clasificacion de texto	
		modelos de clasificacion de textos	Hugging Face Tasks
	automatic text analytics	analisis de texto automatico	Wikipedia
	text mining	mineria de texto	Wikipedia
Procesamiento de voz	speech processing	procesamiento de habla	
		procesamiento del habla	Wikipedia - Common NLP tasks
	speech recognition	reconocimiento de habla	
		reconocimiento del habla	Wikipedia - Common NLP tasks
	speech segmenatation	segmentacion de habla	
		segmentacion del habla	Wikipedia - Common NLP tasks
	text to speech	de texto a voz	Wikipedia - Common NLP tasks
Traducción	translation model	modelo de traduccion	Hugging Face Tasks
	translation models	modelos de traduccion	Hugging Face Tasks
	automatic translation	traducción automatica	Wikipedia - Common NLP tasks
	machine translation		Wikipedia - Common NLP tasks

3.3 Generic Definition of Indicators Based on Mentions

Let M_I and M_E be the number of mentions found in the respective languages, the indicators of mentions in media and reports will be computed as:

$$\text{Ind}(I, E) = \frac{M_I - M_E}{max(M_I, M_E)}$$

This indicator has the following properties. First, it is symmetrical with respect to the languages ($\text{Ind}(I, E) = -\text{Ind}(E, I)$). Moreover, the indicator will be zero if the same number of mentions appear in both languages.

$$M_I = M_E \Longrightarrow \text{Ind}(I, E) = 0\%$$

The indicator will be 100 percent if in both languages the mentions in English are double the mentions in Spanish.

$$M_I = 2 \cdot M_E \Longrightarrow \text{Ind}(I, E) = 100\%$$

In case of a fixed difference between mentions and between mentions for English, the indicator will be inversely proportional to the total number of mentions in English.

$$M_I - M_E = k > 0 \Longrightarrow \text{Ind}(I, E) \propto \frac{1}{M_I}$$

We define different indicators for reports and media respectively. Given the low frequency of mentions in reports, they are not be weighted by products, families of applications or domains.

Indicator A.1 Product mentions in corporate reports

Let Inf_I and Inf_E be the number of mentions of any of the products and application families in reports:

$$\text{I. A. 1} = \frac{Inf_I - Inf_E}{max(Inf_I, Inf_E)} \cdot 100$$

Indicator A.2: NLP technology mentions in corporate reports

Let Inf_I and Inf_E be the number of mentions of any of the products and application families in reports:

$$\text{I. A. 2} = \frac{Inf_I - Inf_E}{max(Inf_I, Inf_E)} \cdot 100$$

Let med_I^i and med_E^i be the number of total mentions found of product or technology i in the respective language media mention indicators, and med_I and med_E be the number of total mentions found (all news stories mentioning the representative companies in each language). The normalized mentions Med_I^i and Med_E^i will be computed as:

$$Med_I^i = \frac{med_I^i}{med_I}$$

$$Med_E^i = \frac{med_E^i}{med_E}$$

Indicador A.3: Product mentions in social media

Let H be the set of application families considered, let W_h be the weight assigned to one of the application families, and let P_h be the set of products identified for that application family:

$$\text{I. A. 3} = \sum_{h \in H} W_h \frac{1}{|P_h|} \sum_{p \in P_h} \frac{Med_I^p - Med_E^p}{max(Med_E^p, Med_I^p)}$$

Med_E^p y Med_I^p represent the number of mentions of product p found in the media.

Indicator A.4: NLP technology mentions in social media.

Let H be the set of application families considered, let W_h be the weight assigned to one of the application families, and let T_h be the set of mentions identified for that application family:

$$\text{I. A. 4} = \sum_{h \in H} W_h \frac{1}{|A_h|} \sum_{t \in T_h} \frac{Med_I^t - Med_E^t}{max(Med_E^t, Med_I^t)}$$

Med_E^t y Med_I^t represent the number of mentions of technology t found in media.

Each of the technologies identified has been assigned the impact shown in the table 3.3.

Table 3.3: Impact assign to each family of NLP applications.

Application family	Impact category
Sentiment analysis	high
Syntactic parsing	low
Textual analysis	low
Morphological Analysis	low
Natural language comprehension	very high
Summary generation	medium
Corpus linguistics	very low
Statistical Linguistics	low
Non-specific PLN mentions	medium
Language modeling	low
Speech processing	high
Entity recognition	medium
Optical character recognition	Medium
Question answering	high
Discourse semantics	medium
Lexical semantics	low
Relational semantics	low
Dialogue systems	high
Generative systems	high
Translation	high

Indicator A.5: NLP technology impact in industry

Let M_I and M_E be the set of companies established in English and Spanish speaking territories respectively. Let Imp_m be the impact produced by the use of language technologies in the company. m:

$$\text{I. A. 5} = \frac{Imp_I - Imp_E}{max(Imp_I, Imp_E)}$$

where

$$Imp_I = \frac{1}{|M_I|} \sum_{m \in M_I} Imp_m \qquad Imp_E = \frac{1}{|M_E|} \sum_{m \in M_E} Imp_m$$

3.3.1 Data for calculations

The number of mentions of products in social media are shown in Table 3.4.

Table 3.4: Mentions in social media by language, application family and product.

Opinion analysis	SP	EN
Sprinklr	152	9482
Khoros	0	2409
NetBase Quid	7	1660
Brandwatch	93	8701
Linkfluence	5	396
Synthesio	1	594
Talkwalker	34	2173
Digimind	115	945
Resonate	574	48495
Meltwater	87	5377
Virtual assistants	SP	EN
Google Assistant	1588	373329
Siri	3944	407456
Alexa	4830	975200
Bixby	397	39817
Cortana	340	52991
Kore.ai	14	6032
IBM Watson Assistant	246	1484
Amazon Lex	6	2539
Google Dialogflow	47	979
Amelia	3	206
ChatGPT	13	2743
Machine translation	SP	EN
Google Translate	980	22217
DeepL	157	3616
Bing Translator o Microsoft Translator	5	1966
Amazon Translate	5	1966
Systran Translate	11	2820
Reverso Translator	0	6
memoQ Translator PRO	0	99
Smartling	18	237
Crowdin	0	67
TextUnited	0	50
Predictive keyboards	SP	EN
Microsoft SwiftKey	0	972
GBoard	86	2890
Grammarly Keyboard	0	335
Fleksy	24	678
iPhone Keyboard	11	6677
GMail	0	25
Google Workspace	0	15
Microsoft Outlook	476	66377
Microsoft Office 365	0	4
Search engines	SP	EN
Google Search	1288	129139
Bing	48	18071
Yahoo Search	35	2836
DuckDuckGo	216	22472
Brave Search	0	555
Elasticsearch	78	7279
Mindbreeze	20	965
Apache Solr	1	825

3.4 Indicators Based on Adoption Surveys

This indicator will be obtained by means of surveys, the design of which is detailed in Appendix. A.

Indicator A.6: Professional technology adoption.

Let H be the set of application families considered in the project, let W_h be the weight assigned to one of the application families, and let P_h be the set of products identified for that application family:

$$\text{I. A. 6} = \sum_{h \in H} W_h \frac{1}{|P_h|} \sum_{p \in P_h} \frac{GAE_I^p - GAE_E^p}{max(GAE_E^p, GAE_I^p)}$$

where GAE_E^p and GAE_I^p represent the adoption for professional use of the product p in both languages.

Indicator A.7.Personal technology adoption.

Let H be the set of application families considered, let W_h be the weight assigned to one of the application families, and let P_h be the set of products identified for that application family:

$$\text{I. A. 6} = \sum_{h \in H} W_h \frac{1}{|P_h|} \sum_{p \in P_h} \frac{GAC_I^p - GAC_E^p}{max(GAC_E^p, GAC_I^p)}$$

where GAC_E^p y GAC_I^p represent the adoption for personal use of the p product in both languages.

3.4.1 Data for calculations

The nummerical data used to calculate the values of indicators A.6 and A.7 are shown in Table 3.5.

Table 3.5: Results of the surveys on adoption of NLP solutions for personal and professional use.

Opinion analysis	Spanish			English		
	personal	professional	both	personal	professional	both
Sprinklr	9	5	9	42	39	27
Khoros	16	10	5	46	28	26
NetBase Quid	9	4	2	44	29	20
Brandwatch	8	6	8	60	27	28
Linkfluence	12	10	4	54	31	40
Synthesio	10	3	4	47	35	21
Talkwalker	20	7	8	58	34	25
Digimind	9	6	7	43	37	22
Resonate	10	6	5	46	26	32
Meltwater	12	4	4	39	36	24
Virtual assistants	Spanish			English		
	personal	professional	both	personal	professional	both
Google Assistant	271	16	43	231	30	74
Siri	206	10	32	234	30	70
Alexa	268	11	25	260	31	53
Bixby	50	10	6	104	34	25
Cortana	151	14	27	136	40	33
Kore.ai	4	5	4	36	22	20
IBM Watson Assistant	5	4	5	37	38	21
Amazon Lex	18	7	7	50	30	27
Google Dialogflow	18	10	10	40	38	32
Amelia	7	5	1	36	32	20
ChatGPT	41	16	18	57	38	25
Machine translation	Spanish			English		
	personal	professional	both	personal	professional	both
Google Translate	369	42	289	257	48	124
DeepL	24	23	59	30	33	30
Bing Translator o Microsoft Translator	56	17	23	82	47	30
Amazon Translate	38	8	15	74	38	36
Systran Translate	10	7	5	41	29	25
Reverso Translator	48	16	28	33	35	30
memoQ Translator PRO	3	5	9	33	35	20
Smartling	7	5	9	39	31	21
Crowdin	3	6	13	33	33	22
TextUnited	11	7	9	39	31	30
Predictive keyboards	Spanish			English		
	personal	professional	both	personal	professional	both
Microsoft SwiftKey	79	8	24	66	37	49
GBoard	120	7	36	85	41	40
Grammarly Keyboard	16	18	12	121	49	90
Fleksy	6	5	4	31	26	21
iPhone Keyboard	127	14	67	204	45	112
GMail	183	26	123	193	49	131
Google Workspace	33	19	26	82	54	56
Microsoft Outlook	105	54	88	121	112	85
Microsoft Office 365	79	50	99	104	91	119
Search engines	Spanish			English		
	personal	professional	both	personal	professional	both
Google Search	313	19	486	434	62	323
Bing	193	31	80	346	62	109
Yahoo Search	197	13	56	343	46	120
DuckDuckGo	61	2	21	180	41	38
Brave Search	26	6	13	66	18	36
Elasticsearch	5	4	6	35	26	28
Mindbreeze	2	3	4	35	22	24
Apache Solr	5	5	5	27	36	22

Table 3.6 contains the results of the adoption surveys by gender.

Table 3.6: Adoption survey results by gender group for NLP application families.

	Spanish			**English**		
	Total	**Masculine**	**Femenine**	**Total**	**Masculine**	**Femenine**
Opinion analysis	18.8%	19.1%	18.4%	34.4%	**39.4%**	29.4%
Virtual assistants	49.1%	46.6%	51.3%	51.6%	55.6%	47.7%
Machine translation	83.2%	81.8%	84.4%	55.3%	**60.8%**	50.1%
Predictive keyboards	68.4%	63.6%	**72.9%**	71.6%	71.8%	71.3%
Search engines	**92.0%**	91.4%	92.7%	96.3%	95.5%	97.1%
Average	**62.3%**	**60.5%**	**63.9%**	**61.8%**	**64.6%**	**59.1%**

Table 3.7 contains the results of the adoption surveys by age group.

Table 3.7: Adoption survey results by age group for NLP application families.

	Spanish				**English**			
	Total	**18 to 29**	**30 to 45**	**46 or more**	**Total**	**18 to 29**	**30 to 45**	**46 oor more**
Opinion analysis	18.8%	20.8%	18.9%	16.8%	**34.4%**	**53.3%**	**35.5%**	14.3%
Virtual assistants	49.1%	**53.5%**	**53.2%**	40.8%	51.6%	**65.5%**	**56.5%**	33.2%
Machine translation	83.2%	85.8%	84.2%	79.6%	**55.3%**	**72.0%**	**58.1%**	35.9%
Predictive keyboards	68.4%	71.3%	**68.7%**	65.1%	71.6%	**86.3%**	**73.4%**	55.1%
Search engines	**92.0%**	92.1%	91.2%	92.8%	96.3%	95.0%	97.0%	97.0%
Average	**62.3%**	**64.7%**	**63.2%**	**59.0%**	**61.8%**	**74.4%**	**64.1%**	**47.1%**

Chapter 4

User Experience Dimension

4.1 Indicators Based on Opinion Analysis

Indicator E.1 Reputational polarity will follow the same scheme as previous indicators, considering the ratio of the difference between languages with respect to the maximum between the two.

$$\frac{\mathrm{Pol}_I^h - \mathrm{Pol}_E^h}{max(\mathrm{Pol}_I^h, \mathrm{Pol}_E^h)}$$

The polarity for each language will be computed as the ratio of positive opinions against the total, considering the neutral opinion samples as elements of uncertainty with a weight of 1/2 for each negative and positive polarity.

$$\mathrm{Pol}_E^d = \frac{\mathrm{Pol}_E^+ + \frac{1}{2}\,\mathrm{Pol}_E^N}{\mathrm{Pol}_E^+ + \mathrm{Pol}_E^- + \mathrm{Pol}_E^N}$$

Only those applications that have a significant volume of opinions will be considered.

Indicator E.1: Reputational polarity

Let H be the set of families of applications considered, let W_h be the weight assigned to each of them and let P_h be the set of products identified for that family of applications:

$$\mathrm{I.\,E.\,1} = \sum_{h \in H} W_h \frac{1}{|P_h|} \sum_{p \in P_h} \frac{\mathrm{Pol}_I^p - \mathrm{Pol}_E^p}{max(\mathrm{Pol}_I^p, \mathrm{Pol}_E^p)}\ ,$$

being:

$$\mathrm{Pol}_E^p = \frac{\mathrm{Pol}_E^+ + \frac{1}{2}\,\mathrm{Pol}_E^N}{\mathrm{Pol}_E^+ + \mathrm{Pol}_E^- + \mathrm{Pol}_E^N} \qquad \mathrm{Pol}_I^p = \frac{\mathrm{Pol}_I^+ + \frac{1}{2}\,\mathrm{Pol}_I^N}{\mathrm{Pol}_I^+ + \mathrm{Pol}_I^- + \mathrm{Pol}_I^N}$$

where Pol_E^+, Pol_E^-, Pol_E^N, Pol_I^+, Pol_I^- y Pol_I^N represent the number of positive, negative and neutral product entries in English or Spanish respectively for the product p.

Indicator E.2 Value curves of product attributes quantifies the gap in terms of product attributes. For this purpose, the reputational polarity of product mentions related to each of the attributes are be considered. These mentions are be categorized into product attributes based on the occurrence of certain key terms that have been identified by means of regular expressions found in Table 4.1 below.

Indicator E.2 Value curves of product attributes

Let H be the set of families of applications considered, let W_h be the weight assigned to each of them, and let C be the set of attributes considered in the value curves:

$$\mathrm{I.\,E.\,2} = \sum_{h \in H} W_h \frac{1}{|C|} \sum_{c \in C} \frac{\mathrm{Pol}_I^{h,c} - \mathrm{Pol}_E^{h,c}}{max(\mathrm{Pol}_I^{h,c}, \mathrm{Pol}_E^{h,c})} \ ,$$

being:

$$\mathrm{Pol}_E^{h,c} = \frac{\mathrm{Pol}_E^{+} + \frac{1}{2}\,\mathrm{Pol}_E^{N}}{\mathrm{Pol}_E^{+} + \mathrm{Pol}_E^{-} + \mathrm{Pol}_E^{N}} \qquad \mathrm{Pol}_I^{h,c} = \frac{\mathrm{Pol}_I^{+} + \frac{1}{2}\,\mathrm{Pol}_I^{N}}{\mathrm{Pol}_I^{+} + \mathrm{Pol}_I^{-} + \mathrm{Pol}_I^{N}}$$

where Pol_E^{+}, Pol_E^{-}, Pol_E^{N}, Pol_I^{+}, Pol_I^{-} y Pol_I^{N} represent the number of positive, negative and neutral product entries in English or Spanish respectively for the h family of applications and associated with the c attribute.

Table 4.1: Terms used to detect product attributes.

<table>
<tr><th>Attribute</th><th>English</th><th>Spanish</th></tr>
<tr><td>Effectiveness</td><td></td><td>precis[oa]</td></tr>
<tr><td></td><td></td><td>correct[oa]</td></tr>
<tr><td></td><td></td><td>falla</td></tr>
<tr><td></td><td></td><td>imprecis[oa]</td></tr>
<tr><td></td><td></td><td>desempeñ[oa]</td></tr>
<tr><td></td><td>precise</td><td>precision</td></tr>
<tr><td></td><td>accurate</td><td>equivoca</td></tr>
<tr><td></td><td>correct</td><td>error</td></tr>
<tr><td></td><td>failure</td><td>errores</td></tr>
<tr><td></td><td>imprecise</td><td>rendimiento</td></tr>
<tr><td></td><td>precision</td><td>eficacia</td></tr>
<tr><td></td><td>wrong</td><td>eficiente</td></tr>
<tr><td></td><td>performance</td><td>eficiencia</td></tr>
<tr><td></td><td>efficiency</td><td>efectividad</td></tr>
<tr><td></td><td>efficient</td><td>procesador</td></tr>
<tr><td></td><td>efficacy</td><td>recursos</td></tr>
<tr><td></td><td>effectiveness</td><td>consum[oe]</td></tr>
<tr><td></td><td>processor</td><td>memoria</td></tr>
<tr><td></td><td>resources</td><td>energia</td></tr>
<tr><td></td><td>memory</td><td>almacenamiento</td></tr>
<tr><td></td><td>energy</td><td>ancho de banda</td></tr>
<tr><td></td><td>storage</td><td>lent[ao]</td></tr>
<tr><td></td><td>bandwidth</td><td>rapid[ao]</td></tr>
<tr><td></td><td>slow</td><td>funciona (?:bien|mal|regular|peor|mejor)</td></tr>
<tr><td></td><td>fast</td><td></td></tr>
<tr><td></td><td>works (?:well|bad|fine|worse|better)</td><td></td></tr>
<tr><td>Usability</td><td>security</td><td>seguridad</td></tr>
<tr><td></td><td>privacy</td><td>privacidad</td></tr>
<tr><td></td><td>login</td><td>login</td></tr>
<tr><td></td><td>session</td><td>sesion</td></tr>
<tr><td></td><td>password</td><td>contrasena</td></tr>
<tr><td></td><td>key</td><td>password</td></tr>
<tr><td></td><td>breach</td><td>violacion</td></tr>
<tr><td></td><td>infringement</td><td>infraccion</td></tr>
<tr><td></td><td>credentials</td><td>credenciales</td></tr>
<tr><td></td><td>permission</td><td>permiso</td></tr>
<tr><td></td><td>cookies</td><td>cookies</td></tr>
<tr><td></td><td>insecure</td><td>acceder</td></tr>
<tr><td></td><td>-</td><td>insegur[oa]</td></tr>
<tr><td>Price</td><td>cheap</td><td>car[oa]</td></tr>
<tr><td></td><td>cheaper</td><td>carisim[oa]</td></tr>
<tr><td></td><td>expensive</td><td>barat[oa]</td></tr>
<tr><td></td><td>free of charge</td><td>baratisim[oa]</td></tr>
<tr><td></td><td>(?:completely|totally|for) free</td><td>gratis</td></tr>
<tr><td></td><td>gratuitous</td><td>economic[oa]</td></tr>
<tr><td></td><td>economical</td><td>de pago</td></tr>
<tr><td></td><td>paid</td><td>licencia</td></tr>
<tr><td></td><td>license</td><td>abonar</td></tr>
<tr><td></td><td>subscribe</td><td>suscribir</td></tr>
<tr><td></td><td>suscripcion</td><td>suscripcion</td></tr>
</table>

4.1.1 Data for calculations

The data used to calculate the value curve are presented in Table 4.2.

Table 4.2: Value curves.

Area	**Attribute**	**Spanish**	**English**	**Spanish**	**English**
Opinion analysis	Effectiveness	0,95	0,76	245	2.296
	Usability	0,87	0,67	57	2.355
	Safety and privacy	0,99	0,76	357	1.752
	Price	0,89	0,95	64	1.943
		Spanish	**English**	**Spanish**	**English**
Virtual assistants	Effectiveness	0,65	0,62	3.168	7.258
	Usability	0,73	0,74	1.140	2.456
	Safety and privacy	0,63	0,65	1.098	4.021
	Price	0,72	0,63	6.528	6.725
		Spanish	**English**	**Spanish**	**English**
Traducción Automática	Effectiveness	0,57	0,54	1.946	5.235
	Usability	0,61	0,52	577	745
	Safety and privacy	0,44	0,43	213	1.857
	Price	0,62	0,58	1.336	3.349
		Spanish	**English**	**Spanish**	**English**
Predictive keyboards	Effectiveness	0,45	0,46	635	3.152
	Usability	0,77	0,55	80	489
	Safety and privacy	0,28	0,42	99	4.916
	Price	0,41	0,60	394	2.582
		Spanish	**English**	**Spanish**	**English**
Search engines	Effectiveness	0,50	0,55	773	4.217
	Usability	0,68	0,64	362	1.802
	Safety and privacy	0,62	0,56	1.468	8.297
	Price	0,73	0,59	431	2.395
		Spanish	**English**	**Spanish**	**English**
Total	Effectiveness	0,62	0,59	1.353	4.432
	Usability	0,73	0,62	443	1.569
	Safety and privacy	0,59	0,57	647	4.169
	Price	0,67	0,67	1.751	3.399

4.2 Indicators Based on Surveys of User Experience

The survey questions are to be found in Appendix A.

Indicator E.3: User satisfaction

Let H be the set of families of applications considered, let W_h be the weight assigned to each of them: Let H be the set families of applications considered in the project, let W_h be the weight assigned to each of them, and let P_h be the set of products identified for that family of applications:

$$\text{I. E. 3} = \sum_{h \in H} W_h \frac{1}{|P_h|} \sum_{p \in P_h} \frac{\text{GS}_I^p - \text{GS}_E^p}{max(\text{GS}_I^p, \text{GS}_E^p)},$$

where GS_E^p, y GS_I^p represents the degree of satisfaction derived from the surveys for the family of applications h in Spanish and English respectively for product p.

Indicator E.4: Limitations

Let H be the set of families of applications considered, let W_h be the weight assigned to each of them, and let C be the set of attributes considered in this indicator:

$$\mathrm{I.\,E.\,4} = \sum_{h \in H} W_h \frac{1}{|P_h|} \sum_{p \in P_h} \frac{1}{|L|} \sum_{l \in L} \frac{\mathrm{GL}_E^{p,l} - \mathrm{GL}_I^{p,l}}{max(\mathrm{GL}_I^{p,l}, \mathrm{GL}_E^{p,l})} ,$$

where $\mathrm{GL}_E^{p,l}$ and $\mathrm{GL}_I^{p,l}$, are the ratio of users who have observed the l limitation of the surveys for the p product in Spanish and English respectively.

4.2.1 Data for calculations

Table 4.3 contains the results of the surveys on user satisfaction.

Table 4.3: Results of the surveys of user experience.

Opinion analysis	Spanish	English
Sprinklr	3,74	3,63
Khoros	3,61	3,72
NetBase Quid	3,87	3,80
Brandwatch	3,55	3,71
Linkfluence	3,77	3,72
Synthesio	3,47	3,80
Talkwalker	3,57	3,91
Digimind	3,68	3,81
Resonate	3,71	3,91
Meltwater	3,70	3,75
Virtual assistants	Spanish	English
Google Assistant	3,72	4,08
Siri	3,80	3,97
Alexa	3,84	4,04
Bixby	3,11	3,31
Cortana	3,15	3,50
Kore.ai	3,69	3,42
IBM Watson Assistant	3,50	3,50
Amazon Lex	3,69	3,53
Google Dialogflow	3,21	3,50
Amelia	3,46	3,40
ChatGPT	3,76	3,77
Machine translation	Spanish	English
Google Translate	3,80	4,13
DeepL	4,41	3,53
Bing Translator o Microsoft Translator	3,48	3,79
Amazon Translate	3,70	3,80
Systran Translate	3,27	3,46
Reverso Translator	3,57	3,45
memoQ Translator PRO	3,65	3,53
Smartling	3,67	3,43
Crowdin	3,64	3,43
TextUnited	3,56	3,53
Predictive keyboards	Spanish	English
Microsoft SwiftKey	3,75	3,80
GBoard	3,82	3,89
Grammarly Keyboard	3,78	4,07
Fleksy	3,67	3,67
iPhone Keyboard	3,87	3,98
GMail	3,83	4,01
Google Workspace	3,73	3,93
Microsoft Outlook	3,80	3,86
Microsoft Office 365	3,91	4,04
Search engines	Spanish	English
Google Search	4,34	4,31
Bing	3,02	3,56
Yahoo Search	3,05	3,69
DuckDuckGo	3,70	3,68
Brave Search	3,69	3,73
Elasticsearch	3,60	3,28
Mindbreeze	3,67	3,42
Apache Solr	3,53	3,51

Table 4.4 contains the results of the surveys on the limitations of NLP applications.

Table 4.4: Results of the surveys on limitations of NLP applications.

	Area	Id_q	Spanish	English	Id_s

0	opinion analysis	P4_1#1	0,22	0,30	P4_1
0	opinion analysis	P4_1#2	0,17	0,29	P4_1
0	opinion analysis	P4_1#3	0,04	0,13	P4_1
0	opinion analysis	P4_1#4	0,26	0,20	P4_1
0	opinion analysis	P4_1#5	0,17	0,23	P4_1
0	opinion analysis	P4_1#6	0,22	0,27	P4_1
0	opinion analysis	P4_1#7	0,17	0,17	P4_1
0	opinion analysis	P4_1#98	0,04	0,11	P4_1
0	opinion analysis	P4_2#1	0,13	0,33	P4_2
0	opinion analysis	P4_2#2	0,16	0,28	P4_2
0	opinion analysis	P4_2#3	0,19	0,16	P4_2
0	opinion analysis	P4_2#4	0,03	0,24	P4_2
0	opinion analysis	P4_2#5	0,16	0,17	P4_2
0	opinion analysis	P4_2#6	0,13	0,19	P4_2
0	opinion analysis	P4_2#7	0,16	0,13	P4_2
0	opinion analysis	P4_2#98	0,00	0,10	P4_2
0	opinion analysis	P4_3#1	0,13	0,30	P4_3
0	opinion analysis	P4_3#2	0,20	0,19	P4_3
0	opinion analysis	P4_3#3	0,20	0,17	P4_3
0	opinion analysis	P4_3#4	0,13	0,19	P4_3
0	opinion analysis	P4_3#5	0,20	0,17	P4_3
0	opinion analysis	P4_3#6	0,13	0,17	P4_3
0	opinion analysis	P4_3#7	0,27	0,18	P4_3
0	opinion analysis	P4_3#98	0,07	0,09	P4_3
0	opinion analysis	P4_4#1	0,05	0,27	P4_4
0	opinion analysis	P4_4#2	0,14	0,26	P4_4
0	opinion analysis	P4_4#3	0,14	0,17	P4_4
0	opinion analysis	P4_4#4	0,27	0,25	P4_4
0	opinion analysis	P4_4#5	0,09	0,18	P4_4
0	opinion analysis	P4_4#6	0,00	0,15	P4_4
0	opinion analysis	P4_4#7	0,09	0,14	P4_4
0	opinion analysis	P4_4#98	0,14	0,14	P4_4
0	opinion analysis	P4_5#1	0,08	0,25	P4_5
0	opinion analysis	P4_5#2	0,27	0,26	P4_5
0	opinion analysis	P4_5#3	0,12	0,18	P4_5
0	opinion analysis	P4_5#4	0,15	0,22	P4_5
0	opinion analysis	P4_5#5	0,12	0,22	P4_5
0	opinion analysis	P4_5#6	0,15	0,16	P4_5
0	opinion analysis	P4_5#7	0,15	0,20	P4_5
0	opinion analysis	P4_5#98	0,08	0,14	P4_5
0	opinion analysis	P4_6#1	0,18	0,31	P4_6
0	opinion analysis	P4_6#2	0,29	0,25	P4_6
0	opinion analysis	P4_6#3	0,12	0,24	P4_6
0	opinion analysis	P4_6#4	0,12	0,20	P4_6
0	opinion analysis	P4_6#5	0,06	0,17	P4_6
0	opinion analysis	P4_6#6	0,06	0,17	P4_6
0	opinion analysis	P4_6#7	0,24	0,13	P4_6
0	opinion analysis	P4_6#98	0,18	0,11	P4_6
0	opinion analysis	P4_7#1	0,11	0,32	P4_7

0	opinion analysis	P4_7#2	0,17	0,23	P4_7
0	opinion analysis	P4_7#3	0,17	0,16	P4_7
0	opinion analysis	P4_7#4	0,11	0,25	P4_7
0	opinion analysis	P4_7#5	0,14	0,18	P4_7
0	opinion analysis	P4_7#6	0,26	0,22	P4_7
0	opinion analysis	P4_7#7	0,14	0,17	P4_7
0	opinion analysis	P4_7#98	0,03	0,13	P4_7
0	opinion analysis	P4_8#1	0,18	0,24	P4_8
0	opinion analysis	P4_8#2	0,27	0,34	P4_8
0	opinion analysis	P4_8#3	0,14	0,20	P4_8
0	opinion analysis	P4_8#4	0,27	0,22	P4_8
0	opinion analysis	P4_8#5	0,23	0,21	P4_8
0	opinion analysis	P4_8#6	0,23	0,23	P4_8
0	opinion analysis	P4_8#7	0,05	0,16	P4_8
0	opinion analysis	P4_8#98	0,14	0,10	P4_8
0	opinion analysis	P4_9#1	0,29	0,29	P4_9
0	opinion analysis	P4_9#2	0,10	0,26	P4_9
0	opinion analysis	P4_9#3	0,05	0,21	P4_9
0	opinion analysis	P4_9#4	0,19	0,16	P4_9
0	opinion analysis	P4_9#5	0,05	0,21	P4_9
0	opinion analysis	P4_9#6	0,05	0,14	P4_9
0	opinion analysis	P4_9#7	0,14	0,22	P4_9
0	opinion analysis	P4_9#98	0,10	0,13	P4_9
0	opinion analysis	P4_10#1	0,15	0,35	P4_10
0	opinion analysis	P4_10#2	0,30	0,30	P4_10
0	opinion analysis	P4_10#3	0,20	0,10	P4_10
0	opinion analysis	P4_10#4	0,15	0,17	P4_10
0	opinion analysis	P4_10#5	0,15	0,20	P4_10
0	opinion analysis	P4_10#6	0,10	0,24	P4_10
0	opinion analysis	P4_10#7	0,10	0,13	P4_10
0	opinion analysis	P4_10#98	0,05	0,10	P4_10
0	virtual assistants	P7_1#1	0,15	0,22	P7_1
0	virtual assistants	P7_1#2	0,15	0,17	P7_1
0	virtual assistants	P7_1#3	0,12	0,08	P7_1
0	virtual assistants	P7_1#4	0,11	0,13	P7_1
0	virtual assistants	P7_1#5	0,16	0,15	P7_1
0	virtual assistants	P7_1#6	0,06	0,09	P7_1
0	virtual assistants	P7_1#7	0,22	0,14	P7_1
0	virtual assistants	P7_1#98	0,09	0,07	P7_1
0	virtual assistants	P7_2#1	0,13	0,27	P7_2
0	virtual assistants	P7_2#2	0,14	0,25	P7_2
0	virtual assistants	P7_2#3	0,14	0,14	P7_2
0	virtual assistants	P7_2#4	0,10	0,10	P7_2
0	virtual assistants	P7_2#5	0,16	0,14	P7_2
0	virtual assistants	P7_2#6	0,10	0,11	P7_2
0	virtual assistants	P7_2#7	0,18	0,16	P7_2
0	virtual assistants	P7_2#98	0,10	0,08	P7_2
0	virtual assistants	P7_3#1	0,10	0,24	P7_3
0	virtual assistants	P7_3#2	0,17	0,19	P7_3

0	virtual assistants	P7_3#3	0,13	0,13	P7_3
0	virtual assistants	P7_3#4	0,11	0,10	P7_3
0	virtual assistants	P7_3#5	0,19	0,15	P7_3
0	virtual assistants	P7_3#6	0,12	0,11	P7_3
0	virtual assistants	P7_3#7	0,16	0,16	P7_3
0	virtual assistants	P7_3#98	0,08	0,08	P7_3
0	virtual assistants	P7_4#1	0,23	0,33	P7_4
0	virtual assistants	P7_4#2	0,24	0,26	P7_4
0	virtual assistants	P7_4#3	0,09	0,17	P7_4
0	virtual assistants	P7_4#4	0,08	0,12	P7_4
0	virtual assistants	P7_4#5	0,12	0,13	P7_4
0	virtual assistants	P7_4#6	0,06	0,12	P7_4
0	virtual assistants	P7_4#7	0,17	0,14	P7_4
0	virtual assistants	P7_4#98	0,15	0,09	P7_4
0	virtual assistants	P7_5#1	0,21	0,31	P7_5
0	virtual assistants	P7_5#2	0,23	0,27	P7_5
0	virtual assistants	P7_5#3	0,11	0,15	P7_5
0	virtual assistants	P7_5#4	0,12	0,09	P7_5
0	virtual assistants	P7_5#5	0,14	0,20	P7_5
0	virtual assistants	P7_5#6	0,06	0,10	P7_5
0	virtual assistants	P7_5#7	0,17	0,15	P7_5
0	virtual assistants	P7_5#98	0,11	0,07	P7_5
0	virtual assistants	P7_6#1	0,00	0,40	P7_6
0	virtual assistants	P7_6#2	0,31	0,15	P7_6
0	virtual assistants	P7_6#3	0,23	0,22	P7_6
0	virtual assistants	P7_6#4	0,08	0,22	P7_6
0	virtual assistants	P7_6#5	0,08	0,26	P7_6
0	virtual assistants	P7_6#6	0,08	0,14	P7_6
0	virtual assistants	P7_6#7	0,00	0,14	P7_6
0	virtual assistants	P7_6#98	0,08	0,10	P7_6
0	virtual assistants	P7_7#1	0,07	0,32	P7_7
0	virtual assistants	P7_7#2	0,07	0,28	P7_7
0	virtual assistants	P7_7#3	0,29	0,18	P7_7
0	virtual assistants	P7_7#4	0,14	0,21	P7_7
0	virtual assistants	P7_7#5	0,14	0,23	P7_7
0	virtual assistants	P7_7#6	0,00	0,17	P7_7
0	virtual assistants	P7_7#7	0,07	0,14	P7_7
0	virtual assistants	P7_7#98	0,07	0,09	P7_7
0	virtual assistants	P7_8#1	0,16	0,31	P7_8
0	virtual assistants	P7_8#2	0,16	0,29	P7_8
0	virtual assistants	P7_8#3	0,13	0,16	P7_8
0	virtual assistants	P7_8#4	0,16	0,15	P7_8
0	virtual assistants	P7_8#5	0,09	0,14	P7_8
0	virtual assistants	P7_8#6	0,06	0,12	P7_8
0	virtual assistants	P7_8#7	0,06	0,10	P7_8
0	virtual assistants	P7_8#98	0,06	0,07	P7_8
0	virtual assistants	P7_9#1	0,08	0,28	P7_9
0	virtual assistants	P7_9#2	0,29	0,20	P7_9
0	virtual assistants	P7_9#3	0,13	0,14	P7_9

0	virtual assistants	P7_9#4	0,26	0,15	P7_9
0	virtual assistants	P7_9#5	0,21	0,19	P7_9
0	virtual assistants	P7_9#6	0,08	0,20	P7_9
0	virtual assistants	P7_9#7	0,13	0,15	P7_9
0	virtual assistants	P7_9#98	0,11	0,06	P7_9
0	virtual assistants	P7_10#1	0,00	0,36	P7_10
0	virtual assistants	P7_10#2	0,31	0,16	P7_10
0	virtual assistants	P7_10#3	0,00	0,24	P7_10
0	virtual assistants	P7_10#4	0,00	0,15	P7_10
0	virtual assistants	P7_10#5	0,08	0,24	P7_10
0	virtual assistants	P7_10#6	0,31	0,22	P7_10
0	virtual assistants	P7_10#7	0,15	0,11	P7_10
0	virtual assistants	P7_10#98	0,08	0,10	P7_10
0	virtual assistants	P7_11#1	0,25	0,32	P7_11
0	virtual assistants	P7_11#2	0,16	0,27	P7_11
0	virtual assistants	P7_11#3	0,13	0,18	P7_11
0	virtual assistants	P7_11#4	0,11	0,16	P7_11
0	virtual assistants	P7_11#5	0,16	0,22	P7_11
0	virtual assistants	P7_11#6	0,03	0,18	P7_11
0	virtual assistants	P7_11#7	0,12	0,12	P7_11
0	virtual assistants	P7_11#98	0,15	0,13	P7_11
0	machine translation	P10_1#1	0,22	0,24	P10_1
0	machine translation	P10_1#2	0,10	0,17	P10_1
0	machine translation	P10_1#3	0,03	0,06	P10_1
0	machine translation	P10_1#4	0,05	0,09	P10_1
0	machine translation	P10_1#5	0,05	0,10	P10_1
0	machine translation	P10_1#6	0,02	0,08	P10_1
0	machine translation	P10_1#7	0,16	0,16	P10_1
0	machine translation	P10_1#98	0,10	0,06	P10_1
0	machine translation	P10_2#1	0,10	0,28	P10_2
0	machine translation	P10_2#2	0,14	0,24	P10_2
0	machine translation	P10_2#3	0,07	0,13	P10_2
0	machine translation	P10_2#4	0,07	0,22	P10_2
0	machine translation	P10_2#5	0,03	0,24	P10_2
0	machine translation	P10_2#6	0,06	0,19	P10_2
0	machine translation	P10_2#7	0,10	0,12	P10_2
0	machine translation	P10_2#98	0,08	0,04	P10_2
0	machine translation	P10_3#1	0,14	0,26	P10_3
0	machine translation	P10_3#2	0,14	0,25	P10_3
0	machine translation	P10_3#3	0,06	0,14	P10_3
0	machine translation	P10_3#4	0,08	0,13	P10_3
0	machine translation	P10_3#5	0,03	0,10	P10_3
0	machine translation	P10_3#6	0,05	0,09	P10_3
0	machine translation	P10_3#7	0,13	0,11	P10_3
0	machine translation	P10_3#98	0,08	0,05	P10_3
0	machine translation	P10_5#1	0,14	0,24	P10_5
0	machine translation	P10_5#2	0,18	0,25	P10_5
0	machine translation	P10_5#3	0,18	0,23	P10_5
0	machine translation	P10_5#4	0,00	0,19	P10_5

0	machine translation	P10_5#5	0,23	0,18	P10_5
0	machine translation	P10_5#6	0,05	0,18	P10_5
0	machine translation	P10_5#7	0,09	0,11	P10_5
0	machine translation	P10_5#98	0,14	0,07	P10_5
0	machine translation	P10_4#1	0,13	0,24	P10_4
0	machine translation	P10_4#2	0,25	0,23	P10_4
0	machine translation	P10_4#3	0,10	0,09	P10_4
0	machine translation	P10_4#4	0,18	0,15	P10_4
0	machine translation	P10_4#5	0,13	0,14	P10_4
0	machine translation	P10_4#6	0,10	0,11	P10_4
0	machine translation	P10_4#7	0,08	0,13	P10_4
0	machine translation	P10_4#98	0,07	0,07	P10_4
0	machine translation	P10_6#1	0,20	0,27	P10_6
0	machine translation	P10_6#2	0,22	0,27	P10_6
0	machine translation	P10_6#3	0,07	0,12	P10_6
0	machine translation	P10_6#4	0,05	0,18	P10_6
0	machine translation	P10_6#5	0,07	0,13	P10_6
0	machine translation	P10_6#6	0,04	0,15	P10_6
0	machine translation	P10_6#7	0,22	0,09	P10_6
0	machine translation	P10_6#98	0,09	0,10	P10_6
0	machine translation	P10_7#1	0,06	0,26	P10_7
0	machine translation	P10_7#2	0,12	0,27	P10_7
0	machine translation	P10_7#3	0,18	0,22	P10_7
0	machine translation	P10_7#4	0,35	0,17	P10_7
0	machine translation	P10_7#5	0,12	0,14	P10_7
0	machine translation	P10_7#6	0,00	0,18	P10_7
0	machine translation	P10_7#7	0,12	0,13	P10_7
0	machine translation	P10_7#98	0,18	0,09	P10_7
0	machine translation	P10_8#1	0,19	0,33	P10_8
0	machine translation	P10_8#2	0,24	0,27	P10_8
0	machine translation	P10_8#3	0,24	0,13	P10_8
0	machine translation	P10_8#4	0,00	0,13	P10_8
0	machine translation	P10_8#5	0,05	0,21	P10_8
0	machine translation	P10_8#6	0,10	0,20	P10_8
0	machine translation	P10_8#7	0,14	0,13	P10_8
0	machine translation	P10_8#98	0,14	0,10	P10_8
0	machine translation	P10_9#1	0,14	0,32	P10_9
0	machine translation	P10_9#2	0,27	0,30	P10_9
0	machine translation	P10_9#3	0,09	0,17	P10_9
0	machine translation	P10_9#4	0,00	0,17	P10_9
0	machine translation	P10_9#5	0,41	0,17	P10_9
0	machine translation	P10_9#6	0,09	0,22	P10_9
0	machine translation	P10_9#7	0,18	0,15	P10_9
0	machine translation	P10_9#98	0,05	0,08	P10_9
0	machine translation	P10_10#1	0,11	0,33	P10_10
0	machine translation	P10_10#2	0,11	0,27	P10_10
0	machine translation	P10_10#3	0,15	0,14	P10_10
0	machine translation	P10_10#4	0,22	0,19	P10_10
0	machine translation	P10_10#5	0,15	0,19	P10_10

0	machine translation	P10_10#6	0,07	0,13	P10_10
0	machine translation	P10_10#7	0,15	0,11	P10_10
0	machine translation	P10_10#98	0,07	0,11	P10_10
0	predictive keyboards	P13_1#1	0,11	0,24	P13_1
0	predictive keyboards	P13_1#2	0,16	0,18	P13_1
0	predictive keyboards	P13_1#3	0,07	0,16	P13_1
0	predictive keyboards	P13_1#4	0,06	0,15	P13_1
0	predictive keyboards	P13_1#5	0,06	0,13	P13_1
0	predictive keyboards	P13_1#6	0,04	0,11	P13_1
0	predictive keyboards	P13_1#98	0,13	0,09	P13_1
0	predictive keyboards	P13_2#1	0,14	0,29	P13_2
0	predictive keyboards	P13_2#2	0,11	0,27	P13_2
0	predictive keyboards	P13_2#3	0,04	0,14	P13_2
0	predictive keyboards	P13_2#4	0,10	0,14	P13_2
0	predictive keyboards	P13_2#5	0,10	0,12	P13_2
0	predictive keyboards	P13_2#6	0,03	0,13	P13_2
0	predictive keyboards	P13_2#98	0,09	0,08	P13_2
0	predictive keyboards	P13_3#1	0,26	0,23	P13_3
0	predictive keyboards	P13_3#2	0,26	0,24	P13_3
0	predictive keyboards	P13_3#3	0,09	0,15	P13_3
0	predictive keyboards	P13_3#4	0,17	0,11	P13_3
0	predictive keyboards	P13_3#5	0,09	0,12	P13_3
0	predictive keyboards	P13_3#6	0,15	0,18	P13_3
0	predictive keyboards	P13_3#98	0,04	0,11	P13_3
0	predictive keyboards	P13_4#1	0,20	0,32	P13_4
0	predictive keyboards	P13_4#2	0,00	0,37	P13_4
0	predictive keyboards	P13_4#3	0,20	0,13	P13_4
0	predictive keyboards	P13_4#4	0,07	0,23	P13_4
0	predictive keyboards	P13_4#5	0,33	0,13	P13_4
0	predictive keyboards	P13_4#6	0,00	0,15	P13_4
0	predictive keyboards	P13_4#98	0,07	0,08	P13_4
0	predictive keyboards	P13_5#1	0,22	0,24	P13_5
0	predictive keyboards	P13_5#2	0,08	0,24	P13_5
0	predictive keyboards	P13_5#3	0,04	0,10	P13_5
0	predictive keyboards	P13_5#4	0,04	0,11	P13_5
0	predictive keyboards	P13_5#5	0,04	0,12	P13_5
0	predictive keyboards	P13_5#6	0,05	0,09	P13_5
0	predictive keyboards	P13_5#98	0,10	0,07	P13_5
0	predictive keyboards	P13_7#1	0,12	0,24	P13_7
0	predictive keyboards	P13_7#2	0,11	0,18	P13_7
0	predictive keyboards	P13_7#3	0,04	0,10	P13_7
0	predictive keyboards	P13_7#4	0,05	0,11	P13_7
0	predictive keyboards	P13_7#5	0,07	0,12	P13_7
0	predictive keyboards	P13_7#6	0,02	0,08	P13_7
0	predictive keyboards	P13_7#98	0,07	0,05	P13_7
0	predictive keyboards	P13_8#1	0,22	0,26	P13_8
0	predictive keyboards	P13_8#2	0,21	0,21	P13_8
0	predictive keyboards	P13_8#3	0,03	0,11	P13_8
0	predictive keyboards	P13_8#4	0,06	0,16	P13_8

0	predictive keyboards	P13_8#5	0,12	0,14	P13_8
0	predictive keyboards	P13_8#6	0,05	0,10	P13_8
0	predictive keyboards	P13_8#98	0,15	0,08	P13_8
0	predictive keyboards	P13_9#1	0,14	0,23	P13_9
0	predictive keyboards	P13_9#2	0,11	0,19	P13_9
0	predictive keyboards	P13_9#3	0,06	0,09	P13_9
0	predictive keyboards	P13_9#4	0,06	0,08	P13_9
0	predictive keyboards	P13_9#5	0,04	0,08	P13_9
0	predictive keyboards	P13_9#6	0,03	0,09	P13_9
0	predictive keyboards	P13_9#98	0,06	0,06	P13_9
0	predictive keyboards	P13_10#1	0,11	0,22	P13_10
0	predictive keyboards	P13_10#2	0,13	0,19	P13_10
0	predictive keyboards	P13_10#3	0,04	0,11	P13_10
0	predictive keyboards	P13_10#4	0,04	0,13	P13_10
0	predictive keyboards	P13_10#5	0,04	0,10	P13_10
0	predictive keyboards	P13_10#6	0,10	0,11	P13_10
0	predictive keyboards	P13_10#98	0,09	0,07	P13_10
0	search engines	P16_1#1	0,06	0,17	P16_1
0	search engines	P16_1#2	0,05	0,13	P16_1
0	search engines	P16_1#3	0,04	0,06	P16_1
0	search engines	P16_1#4	0,11	0,10	P16_1
0	search engines	P16_1#5	0,18	0,16	P16_1
0	search engines	P16_1#6	0,02	0,06	P16_1
0	search engines	P16_1#98	0,05	0,05	P16_1
0	search engines	P16_2#1	0,32	0,24	P16_2
0	search engines	P16_2#2	0,28	0,21	P16_2
0	search engines	P16_2#3	0,09	0,08	P16_2
0	search engines	P16_2#4	0,12	0,08	P16_2
0	search engines	P16_2#5	0,14	0,12	P16_2
0	search engines	P16_2#6	0,02	0,05	P16_2
0	search engines	P16_2#98	0,11	0,07	P16_2
0	search engines	P16_3#1	0,28	0,23	P16_3
0	search engines	P16_3#2	0,24	0,17	P16_3
0	search engines	P16_3#3	0,08	0,07	P16_3
0	search engines	P16_3#4	0,11	0,10	P16_3
0	search engines	P16_3#5	0,14	0,11	P16_3
0	search engines	P16_3#6	0,01	0,05	P16_3
0	search engines	P16_3#98	0,10	0,07	P16_3
0	search engines	P16_4#1	0,17	0,26	P16_4
0	search engines	P16_4#2	0,11	0,21	P16_4
0	search engines	P16_4#3	0,08	0,10	P16_4
0	search engines	P16_4#4	0,05	0,16	P16_4
0	search engines	P16_4#5	0,07	0,14	P16_4
0	search engines	P16_4#6	0,04	0,10	P16_4
0	search engines	P16_4#98	0,06	0,04	P16_4
0	search engines	P16_5#1	0,16	0,30	P16_5
0	search engines	P16_5#2	0,22	0,26	P16_5
0	search engines	P16_5#3	0,07	0,12	P16_5
0	search engines	P16_5#4	0,09	0,17	P16_5

0	search engines	P16_5#5	0,09	0,19	P16_5
0	search engines	P16_5#6	0,04	0,14	P16_5
0	search engines	P16_5#98	0,11	0,11	P16_5
0	search engines	P16_6#1	0,20	0,30	P16_6
0	search engines	P16_6#2	0,13	0,30	P16_6
0	search engines	P16_6#3	0,13	0,18	P16_6
0	search engines	P16_6#4	0,27	0,21	P16_6
0	search engines	P16_6#5	0,07	0,15	P16_6
0	search engines	P16_6#6	0,13	0,18	P16_6
0	search engines	P16_6#98	0,07	0,10	P16_6
0	search engines	P16_7#1	0,00	0,36	P16_7
0	search engines	P16_7#2	0,22	0,21	P16_7
0	search engines	P16_7#3	0,33	0,19	P16_7
0	search engines	P16_7#4	0,22	0,21	P16_7
0	search engines	P16_7#5	0,22	0,25	P16_7
0	search engines	P16_7#6	0,33	0,15	P16_7
0	search engines	P16_7#98	0,00	0,11	P16_7
0	search engines	P16_8#1	0,20	0,29	P16_8
0	search engines	P16_8#2	0,00	0,25	P16_8
0	search engines	P16_8#3	0,20	0,21	P16_8
0	search engines	P16_8#4	0,13	0,20	P16_8
0	search engines	P16_8#5	0,13	0,19	P16_8
0	search engines	P16_8#6	0,20	0,20	P16_8
0	search engines	P16_8#98	0,07	0,08	P16_8

References

Daniel M. Bikel and Imed Zitouni. 2012. *Multilingual Natural Language Processing Applications: From Theory to Practice*. IBM Press.

Daniel Cer, Mona Diab, Eneko Agirre, Iñigo Lopez-Gazpio, and Lucia Specia. 2017. SemEval-2017 task 1: Semantic textual similarity multilingual and crosslingual focused evaluation. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 1–14, Vancouver, Canada. Association for Computational Linguistics.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised cross-lingual representation learning at scale. *CoRR*, abs/1911.02116.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805.

Asier Gutiérrez-Fandiño, Jordi Armengol-Estapé, Marc Pàmies, Joan Llop-Palao, Joaquin Silveira-Ocampo, Casimiro Pio Carrino, Aitor Gonzalez-Agirre, Carme Armentano-Oller, Carlos Rodriguez-Penagos, and Marta Villegas. 2021. Maria: Spanish language models. *arXiv preprint arXiv:2107.07253*.

Daniel Jurafsky and James H. Martin. 2023. *Speech and Language Processing. Third Edition draft*.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach.

Shervin Malmasi, Anjie Fang, Besnik Fetahu, Sudipta Kar, and Oleg Rokhlenko. 2022. Semeval-2022 task 11: Multilingual complex named entity recognition (multiconer). In *Proceedings of the 16th international workshop on semantic evaluation (SemEval-2022)*, pages 1412–1437.

Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. Squad: 100,000+ questions for machine comprehension of text. *arXiv preprint arXiv:1606.05250*.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

Holger Schwenk and Xian Li. 2018. A corpus for multilingual document classification in eight languages. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Paris, France. European Language Resources Association (ELRA).

Appendix A

Surveys

ANEXO I: Cuestionario adopción y satisfacción en España

SECCIÓN A. INTRODUCCIÓN

Gracias por aceptar la invitación para participar en este estudio.

SECCIÓN B. DATOS DEL ENTREVISTADO

Edad. ¿Cuál es tu edad?

Por protocolo se excluyen menores de 18 y mayores de 70.

Género. ¿Cuál es tu género?

Masculino	
Femenino	
Prefiero no contestar	

Ubicación: No se pregunta viene directamente de la base de datos del panel. Pueden participar de todas los territorios, distribución esperada proporcional a la población.

Nivel socioeconómico: No se pregunta, incluye todos los segmentos de a, b y c directamente de la base de datos del panel.

SECCIÓN C. ADOPCIÓN

P1. Has utilizado alguna vez:

		Si	No
P1.1	Alguna herramienta para análisis de opiniones en español		
P1.2	Algún asistente virtual en español		
P1.3	Alguna herramienta para la traducción automática de textos en español o al español		
P1.4	Alguna herramienta de teclado predictivo en español (los teclados predictivos te sugieren o corrigen palabras a medida que escribes texto en el teclado).		
P1.5	Algún buscador web en español		

Se rota el orden de presentación de las herramientas entre los encuestados.

Sólo se continúa con las secciones correspondientes a las P.1 donde haya respondido afirmativamente.

Si una persona responde no a todas las P.1, es decir si la persona no ha utilizado ninguna herramienta, finaliza la encuesta.

SECCIÓN D. ANÁLISIS DE OPINIONES

P2. ¿Has utilizado alguna vez una de estas soluciones para el análisis de opiniones en español?

		Sí, para uso personal	Sí, para uso profesional	Sí, para ambas	No

P2.1	Sprinklr	1	2	3	4
P2.2	Khoros	1	2	3	4
P2.3	NetBase Quid	1	2	3	4
P2.4	Brandwatch	1	2	3	4
P2.5	Linkfluence	1	2	3	4
P2.6	Synthesio	1	2	3	4
P2.7	Talkwalker	1	2	3	4
P2.8	Digimind	1	2	3	4
P2.9	Resonate	1	2	3	4
P2.10	Sysomos	1	2	3	4

Se rota el orden de presentación de las herramientas entre los encuestados.

Las herramientas que en P2 tienen código 1, 2 o 3 pasan a P3, las que tienen código 4 se omiten, si en todas hay código 4 se pasa a la siguiente sección.

P3. ¿Cuál es el grado de satisfacción con estas soluciones de análisis de opiniones en español?

		⭐	⭐⭐	⭐⭐⭐	⭐⭐⭐⭐	⭐⭐⭐⭐⭐
P3.1	Sprinklr	1	2	3	4	5
P3.2	Khoros	1	2	3	4	5

P3.3	NetBase Quid	1	2	3	4	5
P3.4	Brandwatch	1	2	3	4	5
P3.5	Linkfluence	1	2	3	4	5
P3.6	Synthesio	1	2	3	4	5
P3.7	Talkwalker	1	2	3	4	5
P3.8	Digimind	1	2	3	4	5
P3.9	Resonate	1	2	3	4	5
P3.10	Sysomos	1	2	3	4	5

Se mantiene la rotación usada en la pregunta anterior.

P4. ¿Has observado alguna de estas limitaciones a la hora de utilizar estas herramientas en español? Si no ha observado limitaciones, marque “ninguna limitación”.

Tabla con las herramientas que en P2 tienen código 1, 2 o 3. Se pueden incluir 2 o más hojas si se necesitan para que evalúe muchas herramientas.

		Herramienta x	Herramienta x	Herramienta x	Herramienta x
P4.1	En el desempeño o rendimiento				
P4.2	En las funcionalidades				

P4.3	En la compatibilidad con otros sistemas o con el equipo o dispositivo				
P4.4	De seguridad				
P4.5	De privacidad				
P4.6	De precio				
P4.7	En la comprensión del español				
P4.8	Otras limitaciones				
P4.9	Ninguna limitación				

El encuestado puede marcar varias limitaciones por herramienta, pero si usa la última opción de "ninguna", debe ser única y no podrá seleccionar otra.

SECCIÓN E. ASISTENTES VIRTUALES

P5. ¿Has utilizado alguna vez una de estas soluciones de asistencia virtual en español?

		Sí, para uso personal	Sí, para uso profesional	Sí, para ambas	No
P5.1	Google Assistant	1	2	3	4

P5.2	Siri	1	2	3	4
P5.3	Alexa	1	2	3	4
P5.4	Bixby	1	2	3	4
P5.5	Cortana	1	2	3	4
P5.6	Kore.ai	1	2	3	4
P5.7	IBM Watson Assistant	1	2	3	4
P5.8	Amazon Lex	1	2	3	4
P5.9	Google Dialogflow	1	2	3	4
P5.10	Amelia	1	2	3	4
P5.11	ChatGPT	1	2	3	4

Se rota el orden de presentación de las herramientas entre los encuestados.

Las herramientas que en P2 tienen código 1, 2 o 3 pasan a P3, las que tienen código 4 se omiten, si en todas hay código 4 se pasa a la siguiente sección.

P6. ¿Cuál es el grado de satisfacción con estos asistentes virtuales en español?

		⭐	⭐⭐	⭐⭐⭐	⭐⭐⭐⭐	⭐⭐⭐⭐⭐
P6.1	Google Assistant	1	2	3	4	5
P6.2	Siri	1	2	3	4	5

P6.3	Alexa	1	2	3	4	5
P6.4	Bixby	1	2	3	4	5
P6.5	Cortana	1	2	3	4	5
P6.6	Kore.ai	1	2	3	4	5
P6.7	IBM Watson Assistant	1	2	3	4	5
P6.8	Amazon Lex	1	2	3	4	5
P6.9	Google Dialogflow	1	2	3	4	5
P6.10	Amelia	1	2	3	4	5

Se mantiene la rotación usada en la pregunta anterior.

P7. ¿Has observado alguna de estas limitaciones a la hora de utilizar estos asistentes en español? Si no ha observado limitaciones marque en "ninguna limitación".

Tabla con las herramientas que en P2 tienen código 1, 2 o 3. Se pueden incluir 2 o más hojas si se necesitan para que evalúe muchas herramientas.

		Herramienta x	Herramienta x	Herramienta x	Herramienta x
P7.1	En el desempeño o rendimiento				

P7.2	En las funcionalidades				
P7.3	En la compatibilidad con otros sistemas o con el equipo o dispositivo				
P7.4	De seguridad				
P7.5	De privacidad				
P7.6	De precio				
P7.7	En la comprensión del español				
P7.8	Otras limitaciones				
P7.9	Ninguna limitación				

El encuestado puede marcar varias limitaciones por herramienta, pero si usa la última opción de "ninguna", debe ser única y no podrá seleccionar otra.

SECCIÓN F. TRADUCCIÓN AUTOMÁTICA

P8. ¿Has utilizado alguna vez alguna de estas soluciones para la traducción automática de textos en español o al español?

		Sí, para uso personal	Sí, para uso profesional	Sí, para ambas	No

P8.1	Google Translate	1	2	3	4
P8.2	DeepL	1	2	3	4
P8.3	Bing Translator o Microsoft Translator	1	2	3	4
P8.4	Amazon Translate	1	2	3	4
P8.5	Systran Translate	1	2	3	4
P8.6	Reverso Translator	1	2	3	4
P8.7	memoQ Translator PRO	1	2	3	4
P8.8	Smartling	1	2	3	4
P8.9	Crowdin	1	2	3	4
P8.10	TextUnited	1	2	3	4

Se rota el orden de presentación de las herramientas entre los encuestados.

Las herramientas que en P2 tienen código 1, 2 o 3 pasan a P3, las que tienen código 4 se omiten, si en todas hay código 4 se pasa a la siguiente sección.

P9. ¿Cuál es el grado de satisfacción con estas soluciones de traducción automática en español?

		⭐	⭐⭐	⭐⭐⭐	⭐⭐⭐⭐	⭐⭐⭐⭐⭐
P9.1	Google Translate	1	2	3	4	5

P9.2	DeepL	1	2	3	4	5
P9.3	Bing Translator o Microsoft Translator	1	2	3	4	5
P9.4	Amazon Translate	1	2	3	4	5
P9.5	Systran Translate	1	2	3	4	5
P9.6	Reverso Translator	1	2	3	4	5
P9.7	memoQ Translator PRO	1	2	3	4	5
P9.8	Smartling	1	2	3	4	5
P9.9	Crowdin	1	2	3	4	5
P9.10	TextUnited	1	2	3	4	5

Se mantiene la rotación usada en la pregunta anterior.

P10. ¿Has observado alguna de estas limitaciones a la hora de utilizar estas herramientas de traducción automática en español? Si no ha observado limitaciones marque en “ninguna limitación”.

Tabla con las herramientas que en P2 tienen código 1, 2 o 3. Se pueden incluir 2 o más hojas si se necesitan para que evalúe muchas herramientas.

		Herramienta x	Herramienta x	Herramienta x	Herramienta x
P10.1	En el desempeño o rendimiento				
P10.2	En las funcionalidades				
P10.3	En la compatibilidad con otros sistemas o con el equipo o dispositivo				
P10.4	De seguridad				
P10.5	De privacidad				
P10.6	De precio				
P10.7	En un ámbito o sector específico (por ejemplo, traducción jurídica)				
P10.8	Otras limitaciones				
P10.9	Ninguna limitación				

El encuestado puede marcar varias limitaciones por herramienta, pero si usa la última opción de "ninguna", debe ser única y no podrá seleccionar otra.

SECCIÓN G. TECLADOS PREDICTIVOS

P11. ¿Has utilizado alguna vez alguno de estos teclados predictivos en español?

		Sí, para uso personal	Sí, para uso profesional	Sí, para ambas	No
P11.1	Microsoft SwiftKey	1	2	3	4
P11.2	GBoard	1	2	3	4
P11.3	Grammarly (funcionalidad de predicción de frases)	1	2	3	4
P11.4	Fleksy	1	2	3	4
P11.5	iPhone (teclado por defecto)	1	2	3	4
P11.6	Phraseboard	1	2	3	4
P11.7	GMail (funciones predictivas en la redacción)	1	2	3	4
P11.8	Google Workspaces (funciones predictivas en la redacción)	1	2	3	4
P11.9	Microsoft Outlook (funciones predictivas en la redacción)	1	2	3	4
P11.10	Microsoft Office 365 (funciones predictivas en la redacción)	1	2	3	4

Se rota el orden de presentación de las herramientas entre los encuestados.

Las herramientas que en P2 tienen código 1, 2 o 3 pasan a P3, las que tienen código 4 se omiten, si en todas hay código 4 se pasa a la siguiente sección.

P12. ¿Cuál es el grado de satisfacción con estos teclados predictivos en español?

		⭐	⭐⭐	⭐⭐⭐	⭐⭐⭐⭐	⭐⭐⭐⭐⭐
P12.1	Microsoft SwiftKey	1	2	3	4	5
P12.2	GBoard	1	2	3	4	5
P12.3	Grammarly (funcionalidad de predicción de frases)	1	2	3	4	5
P12.4	Fleksy	1	2	3	4	5
P12.5	iPhone (teclado por defecto)	1	2	3	4	5
P12.6	Phraseboard	1	2	3	4	5
P12.7	GMail (funciones predictivas en la redacción)	1	2	3	4	5

P12.8	Google Workspaces (funciones predictivas en la redacción)	1	2	3	4	5
P12.9	Microsoft Outlook (funciones predictivas en la redacción)	1	2	3	4	5
P12.10	Microsoft Office 365 (funciones predictivas en la redacción)	1	2	3	4	5

Se mantiene la rotación usada en la pregunta anterior.

P13. ¿Has observado alguna de estas limitaciones a la hora de utilizar estos teclados predictivos en español? Si no ha observado limitaciones marque en "ninguna limitación".

Tabla con las herramientas que en P2 tienen código 1, 2 o 3. Se pueden incluir 2 o más hojas si se necesitan para que evalúe muchas herramientas.

		Herramienta x	Herramienta x	Herramienta x	Herramienta x
P13.1	En el desempeño o rendimiento				
P13.2	En las funcionalidades				

P1 3.3	En la compatibilidad con otros sistemas o con el equipo o dispositivo				
P1 3.4	De seguridad				
P1 3.5	De privacidad				
P1 3.6	De precio				
P1 3.7	Otras limitaciones				
P1 3.8	Ninguna limitación				

El encuestado puede marcar varias limitaciones por herramienta, pero si usa la última opción de "ninguna", debe ser única y no podrá seleccionar otra.

SECCIÓN H. BUSCADORES WEB

P14. ¿Has utilizado alguna vez alguno de estos buscadores web en español?

		Sí, para uso personal	Sí, para uso profesional	Sí, para ambas	No
P14.1	Google Search	1	2	3	4
P14.2	Bing	1	2	3	4
P14.3	Yahoo Search	1	2	3	4

P14.4	DuckDuckGo	1	2	3	4
P14.5	Brave Search	1	2	3	4
P14.6	Elasticsearch	1	2	3	4
P14.7	Mindbreeze	1	2	3	4
P14.8	Apache Solr	1	2	3	4

Se rota el orden de presentación de las herramientas entre los encuestados.

Las herramientas que en P2 tienen código 1, 2 o 3 pasan a P3, las que tienen código 4 se omiten, si en todas hay código 4 se pasa a la siguiente sección.

P15. ¿Cuál es el grado de satisfacción con estos buscadores web en español?

		⭐	⭐⭐	⭐⭐⭐	⭐⭐⭐⭐	⭐⭐⭐⭐⭐
P15.1	Google Search	1	2	3	4	5
P15.2	Bing	1	2	3	4	5
P15.3	Yahoo Search	1	2	3	4	5
P15.4	DuckDuckGo	1	2	3	4	5
P15.5	Brave Search	1	2	3	4	5
P15.6	Elasticsearch	1	2	3	4	5

P15.7	Mindbreeze	1	2	3	4	5
P15.8	Apache Solr	1	2	3	4	5

Se mantiene la rotación usada en la pregunta anterior.

P16. ¿Has observado alguna de estas limitaciones a la hora de utilizar estos buscadores web en español? Si no ha observado limitaciones marque en “ninguna limitación”.

Tabla con las herramientas que en P2 tienen código 1, 2 o 3. Se pueden incluir 2 o más hojas si se necesitan para que evalúe muchas herramientas.

		Herramienta x	Herramienta x	Herramienta x	Herramienta x
P16.1	En el desempeño o rendimiento				
P16.2	En las funcionalidades				
P16.3	En la compatibilidad con otros sistemas o con el equipo o dispositivo				
P16.4	De seguridad				
P16.5	De privacidad				
P16.6	De precio				

P1 6.7	Otras limitaciones				
P1 6.8	Ninguna limitación				

El encuestado puede marcar varias limitaciones por herramienta, pero si usa la última opción de "ninguna", debe ser única y no podrá seleccionar otra.

CIERRE: MUCHAS GRACIAS POR PARTICIPAR EN ESTE ESTUDIO, QUE TENGAS EXCELENTE DÍA, TARDE, NOCHE.

ANEXO II: Questions adoption and satisfaction in the USA

A. INTRODUCTION

Thank you for accepting the invitation to participate in this study.

B. INTERVIEWEE DATA

Age. How old are you?

Gender. What is your gender?

Male	
Female	
I rather not answer	

C. ADOPTION

P1. Have you ever used:

		Yes	No
P1.1	Any tool for opinion analysis in English		
P1.2	Any virtual assistant in English		
P1.3	Any tool for automatic translation of texts in English or into English		
P1.4	Any predictive keyboard tool in English (predictive keyboards suggest or correct words as you type text on the keyboard).		
P1.5	Any web search engine in English		

D. OPINION ANALYSIS

P2. Have you ever used any of these solutions for opinion analysis in English?

		Yes, for personal use	Yes, for professional use	Yes, for both	No
P2.1	Sprinklr	1	2	3	4
P2.2	Khoros	1	2	3	4
P2.3	NetBase Quid	1	2	3	4
P2.4	Brandwatch	1	2	3	4
P2.5	Linkfluence	1	2	3	4

P2.6	Synthesio	1	2	3	4
P2.7	Talkwalker	1	2	3	4
P2.8	Digimind	1	2	3	4
P2.9	Resonate	1	2	3	4
P2.10	Sysomos	1	2	3	4

P3. How satisfied are you with these opinion analysis solutions in English?

		★	★★	★★★	★★★★	★★★★★
P3.1	Sprinklr	1	2	3	4	5
P3.2	Khoros	1	2	3	4	5
P3.3	NetBase Quid	1	2	3	4	5
P3.4	Brandwatch	1	2	3	4	5
P3.5	Linkfluence	1	2	3	4	5
P3.6	Synthesio	1	2	3	4	5
P3.7	Talkwalker	1	2	3	4	5
P3.8	Digimind	1	2	3	4	5
P3.9	Resonate	1	2	3	4	5

P3.10	Sysomos	1	2	3	4	5

P4. Have you observed any of these limitations when using these tools? If you have not observed any limitations, check "no limitations".

		Tool x	Tool x	Tool x	Tool x
P4.1	In performance				
P4.2	In functionalities				
P4.3	On compatibility with other systems or with the equipment or device				
P4.4	In security				
P4.5	In privacy				
P4.6	In price				
P4.7	In understanding English				
P4.8	Other limitations				
P4.9	No limitations				

E. VIRTUAL ASSISTANTS

P5. Have you ever used any of these virtual assistants in English?

		Yes, for personal use	Yes, for professional use	Yes, for both	No
P5.1	Google Assistant	1	2	3	4
P5.2	Siri	1	2	3	4
P5.3	Alexa	1	2	3	4
P5.4	Bixby	1	2	3	4
P5.5	Cortana	1	2	3	4
P5.6	Kore.ai	1	2	3	4
P5.7	IBM Watson Assistant	1	2	3	4
P5.8	Amazon Lex	1	2	3	4
P5.9	Google Dialogflow	1	2	3	4
P5.10	Amelia	1	2	3	4
P5.11	ChatGPT	1	2	3	4

P6. How satisfied are you with these virtual assistants in English?

		⭐	⭐⭐	⭐⭐⭐	⭐⭐⭐⭐	⭐⭐⭐⭐⭐

P6.1	Google Assistant	1	2	3	4	5
P6.2	Siri	1	2	3	4	5
P6.3	Alexa	1	2	3	4	5
P6.4	Bixby	1	2	3	4	5
P6.5	Cortana	1	2	3	4	5
P6.6	Kore.ai	1	2	3	4	5
P6.7	IBM Watson Assistant	1	2	3	4	5
P6.8	Amazon Lex	1	2	3	4	5
P6.9	Google Dialogflow	1	2	3	4	5
P6.10	Amelia	1	2	3	4	5

P7. Have you observed any of these limitations when using these virtual assistants in English? If you have not observed any limitations, please check "no limitations".

		Tool x	Tool x	Tool x	Tool x
P7.1	In performance				

P7.2	In functionalities				
P7.3	On compatibility with other systems or with the equipment or device				
P7.4	In security				
P7.5	In privacy				
P7.6	In price				
P7.7	In understanding English				
P7.8	Other limitations				
P7.9	No limitations				

F. AUTOMATIC TRANSLATION

P8. Have you ever used any of these solutions for the automatic translation of texts in or into English?

		Yes, for personal use	Yes, for professional use	Yes, for both	No
P8.1	Google Translate	1	2	3	4

P8.2	DeepL	1	2	3	4
P8.3	Bing Microsoft Translator	1	2	3	4
P8.4	Amazon Translate	1	2	3	4
P8.5	Systran Translate	1	2	3	4
P8.6	Reverso Translator	1	2	3	4
P8.7	memoQ Translator PRO	1	2	3	4
P8.8	Smartling	1	2	3	4
P8.9	Crowdin	1	2	3	4
P8.10	TextUnited	1	2	3	4

P9. How satisfied are you with these solutions for the automatic translation of texts into English?

		⭐	⭐⭐	⭐⭐⭐	⭐⭐⭐⭐	⭐⭐⭐⭐⭐
P9.1	Google Translate	1	2	3	4	5
P9.2	DeepL	1	2	3	4	5
P9.3	Bing Translator or Microsoft Translator	1	2	3	4	5

P9.4	Amazon Translate	1	2	3	4	5
P9.5	Systran Translate	1	2	3	4	5
P9.6	Reverso Translator	1	2	3	4	5
P9.7	memoQ Translator PRO	1	2	3	4	5
P9.8	Smartling	1	2	3	4	5
P9.9	Crowdin	1	2	3	4	5
P9.10	TextUnited	1	2	3	4	5

P10. Have you observed any of these limitations when using these tools for automatic translation in or into English? If you have not observed any limitations, please check "no limitations".

		Tool x	Tool x	Tool x	Tool x
P10.1	In performance				
P10.2	In functionalities				
P10.3	On compatibility with other systems or with the equipment or device				

P1 0.4	In security				
P1 0.5	In privacy				
P1 0.6	In price				
P1 0.7	In a specific sector or área (for instance, legal texts translation)				
P1 0.8	Other limitations				
P1 0.9	No limitations				

G. PREDICTIVE KEYBOARDS

P11. Have you ever used any of these predictive keyboards in English?

		Yes, for personal use	Yes, for professional use	Yes, for both	No
P11.1	Microsoft SwiftKey	1	2	3	4
P11.2	GBoard	1	2	3	4
P11.3	Grammarly (phrasal predictions feature)	1	2	3	4
P11.4	Fleksy	1	2	3	4

P11.5	iPhone's default keyboard	1	2	3	4
P11.6	Phraseboard	1	2	3	4
P11.7	GMail (predictive functions in writing)	1	2	3	4
P11.8	Google Workspaces (predictive functions in writing)	1	2	3	4
P11.9	Microsoft Outlook (predictive functions in writing)	1	2	3	4
P11.10	Microsoft Office 365 (predictive functions in writing)	1	2	3	4

P12. How satisfied are you with these predictive keyboards in English?

		⭐	⭐⭐	⭐⭐⭐	⭐⭐⭐⭐	⭐⭐⭐⭐⭐
P12.1	Microsoft SwiftKey	1	2	3	4	5
P12.2	GBoard	1	2	3	4	5
P12.3	Grammarly (phrasal predictions feature)	1	2	3	4	5

P12.4	Fleksy	1	2	3	4	5
P12.5	iPhone’s default keyboard	1	2	3	4	5
P12.6	Phraseboard	1	2	3	4	5
P12.7	GMail (predictive functions in writing)	1	2	3	4	5
P12.8	Google Workspaces (predictive functions in writing)	1	2	3	4	5
P12.9	Microsoft Outlook (predictive functions in writing)	1	2	3	4	5
P12.10	Microsoft Office 365 (predictive functions in writing)	1	2	3	4	5

P13. Have you observed any of these limitations when using these predictive keyboards in English? If you have not observed any limitations, please check "no limitations".

		Tool x	Tool x	Tool x	Tool x

P1 3.1	In performance				
P1 3.2	In functionalities				
P1 3.3	On compatibility with other systems or with the equipment or device				
P1 3.4	In security				
P1 3.5	In privacy				
P1 3.6	In price				
P1 3.7	Other limitations				
P1 3.8	No limitations				

H. WEB SEARCH ENGINES

P14. Have you ever used any of these web search engines in English?

		Yes, for personal use	Yes, for professional use	Yes, for both	No
P14.1	Google	1	2	3	4
P14.2	Bing	1	2	3	4

P14.3	Yahoo	1	2	3	4
P14.4	DuckDuckGo	1	2	3	4
P14.5	brave.com	1	2	3	4
P14.6	Elasticsearch	1	2	3	4
P14.7	Mindbreeze	1	2	3	4
P14.8	Apache Solr	1	2	3	4

P15. How satisfied are you with these web search engines in English?

		⭐	⭐⭐	⭐⭐⭐	⭐⭐⭐⭐	⭐⭐⭐⭐⭐
P15.1	Google Search	1	2	3	4	5
P15.2	Bing	1	2	3	4	5
P15.3	Yahoo Search	1	2	3	4	5
P15.4	DuckDuckGo	1	2	3	4	5
P15.5	Brave Search	1	2	3	4	5
P15.6	Elasticsearch	1	2	3	4	5

P15.7	Mindbreeze	1	2	3	4	5
P15.8	Apache Solr	1	2	3	4	5

P16. Have you observed any of these limitations when using these web search engines in English? If you have not observed any limitations, please check "no limitations".

		Tool x	Tool x	Tool x	Tool x
P16.1	In performance				
P16.2	In functionalities				
P16.3	On compatibility with other systems or with the equipment or device				
P16.4	In security				
P16.5	In privacy				
P16.6	In price				
P16.7	Other limitations				
P16.8	No limitations				

CLOSING: THANK YOU VERY MUCH FOR PARTICIPATING IN THIS STUDY, HAVE A GREAT DAY, AFTERNOON, EVENING.