DOI: 10.5748/19CONTECSI/PSE/DSC/6978

PROPOSAL OF A SOFTWARE TRANSLATOR WITH INTERLANGUAGE TRANSLATION RESOURCES, BRASILIAN SIGN LANGUAGE (LIBRAS) – PORTUGUESE

Vagner Luiz Gava ; https://orcid.org/0000-0001-5965-957X
Instituto de Pesquisas Tecnológicas do Estado de São Paulo - IPT

Angelina da Silva Inacio ; https://orcid.org/0000-0003-4585-116X
Instituto de Pesquisas Tecnológicas do Estado de São Paulo - IPT

Felipe Augusto Frazão Kleine ; https://orcid.org/0000-0001-8372-3770
Instituto de Pesquisas Tecnológicas do Estado de São Paulo - IPT

Fabio Rogerio Lins Pereira de Souza ; https://orcid.org/0000-0002-1015-6880
Instituto de Pesquisas Tecnológicas do Estado de São Paulo - IPT

PROPOSAL OF A SOFTWARE TRANSLATOR WITH INTERLANGUAGE TRANSLATION RESOURCES, BRASILIAN SIGN LANGUAGE (LIBRAS) – PORTUGUESE

Abstract: Several works focus on the translation process involving sign and written-oral languages. Regardless of the translation order assumed by the process, whether from sign languages to written-oral languages or vice versa, an intermediate written representation is taken as a basis of processing, in which words are presented following the grammatical order established by the sign languages. This system of transcription is known as gloss. According to literature, gloss is a resource used by specialists only to represent signs through writing. In a way, deaf writing (cursive or typed on some input device) is very similar to gloss. Although this writing is often quite difficult for the deaf, this is the only way they can communicate in writing with hearing people. Our challenge is to propose a solution that is not only capable of helping the deaf in the production of their sentences in Libras (glosses) but at the same time translating them into Portuguese. We bring to the debate some relevant technical aspects of this translation software. This work proposed a translator based solely on morphological and syntactic classifications in the first version. Showing themselves as encouraging, the results prove useful when showing hints to UX development in building an interface to support the words typed by deaf users, although on the other side it proved unsuitable for dealing with more elaborate syntactic structures. As a future work, it will be developed a grammar with the necessary resources to over-come this limitation.

Keywords: Natural Language Processing, Brazilian Sign Language, Libras, Machine Translation, Rule-Based.

1. Introduction

According to the IBGE (2010) Census, about 10 million Brazilians have hearing loss, of which 2 million are severely impaired: they hear nothing or very little. Sign languages are the usual means of visuospatial communication used by the deaf, where interlocutors communicate through sequential signs that use various parts of the body (hands, arms, head, trunk, and facial expressions) (STOKOE, 1960; STOKOE; CASTERLINE; CRONEBERG, 1965).

Deaf individuals who are literate in Portuguese and have learned to communicate with hearing people through written language may still have difficulties interpreting sentences with complex structures or words (e.g., specific to a field of knowledge or ambiguous in meaning) (IBGE, 2010) and writing sentences using articles, punctuation, verb conjugation, etc., to create sentences closer to formal Portuguese.

In conventional means of Information and Communication Technology, deaf individuals prefer to use videos to communicate, especially when talking about more complex subjects. However, exchanging videos can be slow and costly due to the inherent delay in sending and downloading videos and the use of the devices' local storage memory capacity. Therefore, communication through texts can benefit the deaf community in several day-to-day situations.

In Brazil, there are few platforms with accessibility in Libras for the insertion of texts, especially regarding essential services, such as registration on government pages, scheduling and communication at medical and hospital care centers, consumer services, banks, and self-service terminals. In customer service channels, in particular, it is almost impossible to communicate with a deaf person through any of the usual means (i.e., call center, telephone, chat, etc.), so the provision of service to such customers is almost wholly thwarted. This points to the need to develop a specific tool to facilitate the insertion of texts in Portuguese.

Thus, the general objective of this article is to develop a tool to help deaf individuals whose primary language is Libras in the insertion of texts in written Portuguese in information and communication technology input devices like notebooks, smartphones, and tablets.

The article is organized as follows: Section 2 describes works related to the issue of translation from gloss to text. In Section 3, the solution proposed in this work is presented. Then, in Section 4, the results of the tests performed to evaluate the quality of the proposed solution are given. In Section 5, the main conclusions about this work are shown and finally in Section 6 future works are related.

2. Background

The development of systems aimed at Natural Language Processing (CHOWDHARY, 2020; NADKARNI; OHNO-MACHADO; CHAPMAN, 2011) is an area that for some time has been drawing a lot of attention, not only from researchers and language scholars but also from extensive technology and service providers (PIRIS; GAY, 2021; HOPP et all, 2021; PANDEY; PANDEY, 2019). But despite notable advances, especially in recent years, there are still many issues that need to be resolved.

The challenges faced by developers of this type of system are of different shades of complexity, not only in the field of computing but also, in a unique way, in linguistics, due to the demands for good performance and quality of the solutions. The linguistic issues present in these systems involve different processing modules: morphologic, phonetic-phonologic, syntactic, semantic, and, depending on the application, pragmatic-discursive (KLEIN; MANNING, 2005; DE LIMA; ABRAHÃO; FILHO, 1998).

The following are well-known computing solutions (websites, applications, etc.), and their main characteristics:

- Dictionaries and Glossaries: are limited to translating from the written to the sign language. They have word search, meanings, and videos with the corresponding sign. Examples: Libras Practical Guide (ACESSIBILIDADE BRASIL, 2011), Librazuka (LIBRASIL, 2022), WikiSigns (gathers sign languages from various locales) (WIKISIGNS, 2022), WikiLibras (collaborative platform for inclusion and correction of signs available in VLibras) (LIBRASOL, 2020). For English we have: Sign ASL (SIGNASL, 2022), Sign BSL (SIGNASL, 2022).
- Toolkit for use on computers, websites, and mobile devices: translate selected Portuguese texts or texts input into a device for Signal Language (SL) through an avatar – projected representation of a person or user within the virtual environment. They have the disadvantages of having little or no facial expression when performing the signs by the avatars (facial expression has a grammatical function in sign languages) and performing literal sign-word translation or spelling of the written word if there is no corresponding sign (LIBRASOL, 2022). Examples: VLibras (GOVERNO DIGITAL, 2022), Hand Talk (HAND TALK, 2022), and Rybená (RYBENÁ, 2022).
- Applications for translating Libras into written Portuguese: in 2019, a prototype of the Charles Translator application was launched (CANALTECH, 2019). However, its development appears to have been frozen or interrupted.
- Tools that work with SL to text translation: use Artificial Intelligence techniques such as Natural Language Processing and Computer Vision to translate signals into text or audio. The applications use video cameras, and the most advanced ones focus on the American SL (ASL). Examples: SignAll (SIGNALL, 2022), Signily by ASLized (ASLIZED!, 2022), five (LABROOTS, 2018), ASL Keyboard (ANDROIDBLIP, 2022) and The ASL App (THE ASL APP, 2022). Some projects were stopped or abandoned.

“Research has shown that for acquiring a second language, humans develop, un-der different influences, an internally structured cognitive grammar (intermediate language), which also allows us to anticipate that this grammar has elements of both the source language and the target language. From these studies, it is also evident that some significant absences in both the source and target languages tend to occur and that, in these cases, it is necessary to find strategies or other means that allow the translator to achieve an effective and efficient result, given the purpose of the translation (CORDER, 1992; ELLIS, 1985; NEMSER, 1971; SELINKER, 1992; BIDARRA, 2018).

Linguists/researchers established and adopted the term glosses to call the transcriptions of an SL using written words from the oral language. According to Ferreira-Brito and Langevin (1995), Klima and Bellugi (1979), Wilcox and Wilcox (1997) and Quadros e Karnopp (2004), the glosses and words of an oral language are used to loosely represent the signs in a written form. The deaf person educated in Libras, when writing, maintains a structure equivalent to glosses, inserting some components of the syntax of oral Portuguese depending on their degree of its knowledge. The resulting textual Portuguese is called interlingua in this work.

3. Gloss to Portuguese Translator

3.1 Initial Studies

Initially, the studies have been conducted through the newest environments of artificial intelligence, and in order to make a consistent road map of this work, a brief introduction has to be made.

In accordance with Koehn (2010), the machine translation is almost as old as the computer itself if we consider that the Great Britain had used the machines to crack the Enigma code in World War II. Since that many technics where evolved, ranging from direct translation, which maps inputs to outputs with simple and basic rules, to more sophisticated transfer methods that employ morphological, syntactic, and lexical analysis, as well as through the interlingua methods that use an abstract meaning representation.

Starting with rule-based, passing through statistical and neural machine translation, technics where revised, tested with the available deaf-to-listeners parallel corpus, and back to the beginning in order to define the best 'reachable in time' way to con-struct the translator and help the deaf in the way to communicate with listeners through digital means.

Deep learning gives us a very powerful tool with neural machine translation by the use of artificial neural networks (ANN). These networks (KOEHN, 2020) are similar to the biological ones, with output based on input and activation functions, yet in the same way being quite different in response. The artificial ones are connected in orderly architectures and grouped into layers, where each perceptron (artificial neuron) can 'learn' mathematical operations and functions, in order to compose, layer by layer up to the output, with a complex output signal representing the response, categorical or numerical, depending on the problem.

So on, the first tests were made with ANN were conducted with Long-Sort Term Memory (LSTM), which is a type of Recurrent Neural Network (RNN), that can correlate layers outputs, through time, and with this, correlate words with previous ones, computing the context of a text sample. These RNNs were chosen for their capacity for generalization and were mounted within the Encoder-Decoder structure, which are made for encoding the inputs into numerical multidimensional context vectors, and feeding the decoder that will make the inverse path to choose translated sentences into a complete output, respecting order and context.

The results of these types of solutions tend to be good. However, for these deep networks, enormous parallel corpora for best training are needed, and for this, the results of the training were far distant from good.

In this context, another path was in sight. The Statistical Machine Translation (SMT) was created before Neural Machine Translation (NMT). That does not have the generalization capacity of the NMT, only responding to the questions for the sentences previously trained, but can generate optimal results with a good corpus, far minor than the necessity of a These models are well known and used in the context of machine translation, using the n-gram model and Bayes theorem that correlates the probability of one event with the occurrence of another given event as follows (1):

$$P(meta_1 \mid foreign_2) = [P(foreign_2 \mid meta_1) * P(meta_1)] / P(foreign_2)(1)$$

Where $meta_i$ and $foreign_i$ are the sentences in the output and input languages respectively, $P(x \mid y)$ is the probability of x given that y occurred, and at last, $P(x)$ is the probability of x into the corpus. The probability of the sentences was given by the occurrence in the corpus and the related probability was given by the probability of an output given that an input, as it appears in the same parallel corpus.

In more complex models, the Markov Assumption is used, and the sentence is estimated by the previous ones. As it can be seem below, assuming that w_i is the sentence in the position i of an n length input (2):

$$P(w_1,w_2,w_3,\ldots,w_n) = P(w_1)*P(w_2|w_1),P(w_3|w_2)*...*P(w_n|w_{n-1}) \qquad (2)$$

These models were already introduced in NLTK pack for Python language (NLTK, 2022), that which was used in the tests. IBM Models 1, 2, and 3, were tested, in this order, with the theoretical corpus, and the deaf-to-listener corpus created for this research. In all models, the theoretical corpus converges for good results, though only in 1 and 2 converge for the other corpus. From the models tested, the IBM Model 3 is the one that has tools for null and multiple convergences, and for that reason, the corpus must be constructed in a way to be partially connected, line by line, giving means to the algorithm to create the alignments and then making the translation possible. These results and the necessity of grammatical rules corrections lead the research back to grammatical emphasis as described in the next section.

3.2 Proposed Solution

In order to meet the demands of a tool as described in Section 2, the proposed solution of this work involves a software component developed to perform automatic translation into oral Portuguese. This solution covers some morphological-syntactic aspects, such as the treatment of verb tenses, prepositions, and numerals.

According to Othero (2009), the study of the syntax of a language can be essential for its computational treatment in several other levels of linguistic description, and generally, works of computational implementation in a language involve the syntactic or morphological treatment of that language. Sign languages have autonomy regarding structural and geographical rules (NORTH, 1990) in such a way that even if an SL is based on an oral language (Libras and Brazilian Portuguese), it has its own morphological rules, constructions, semantics, and variations within a social group that differ from the oral language of origin.

This work is based on the premise that each typed sentence has an organization between the word level and the sentence (syntagma). Thus, each syntagma has a set of words, and the sentence will be understood as a set of syntagmas. In Fig. 1, the architecture of the proposed translation component is shown.

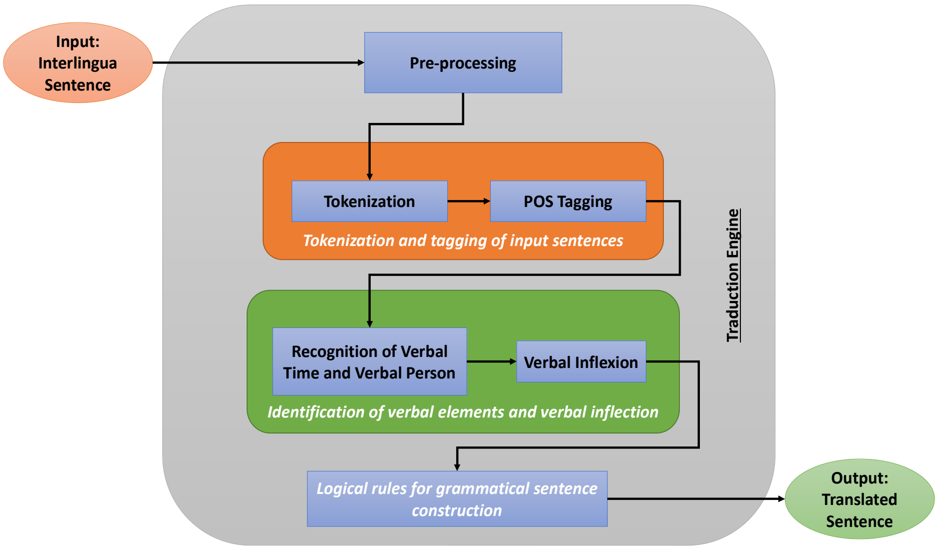

Fig. 1. Translation component architecture.

According to Fig. 1, the translation software receives an interlingua text as input, and, after a few processing steps, the result is a text close to spoken Portuguese. Specifically, the proposed solution flow works according to the steps described below:

Tokenization and tagging of input sentences. Initially, the translation software processes the input sentences, dividing them into the smallest feasible unit, such as words, numerals, and symbols. This smallest possible unit is commonly named, in PLN, as tokens, and this splitting process is called tokenization.

Next, the software classifies which types of elements each token of the sentence most likely represents. This is done by a statistical survey of tokens present in a corpus and a previous classification of these in the different elements. This tagging process (Part-Of-Speech Tagging) uses technology and corpus current in the SpaCy tool (SPACY, 2022). In this specific software, this tagging was performed with the Universal POS Tagging classification (UNIVERSAL DEPENDENCIES, 2022) applied to Portuguese, based on a corpus of oral Portuguese since there is no corpus for gloss.

Identification of verbal elements and verbal inflection. With the help of the labeling made by the previous step, the elements that indicate the verb tense and person of the input sentence are identified. For this purpose, initially, two files are used in ".csv" containing two established classification sets: one in which the person of the verb of the sentence is returned according to pronouns or input nouns; and another that returns the tense of the sentence given an input adverb. In this way, the tense and person of the verb in the sentence are stipulated using the adverbs, pronouns, and nouns identified by the sentence labeling.

The software's next step is to return the most appropriate verb inflection for the verb in the sentence with this information. For this, the third file in ".csv" is used, which returns the conjugated verb given the information provided: the verb in its infinitive (also named in PLN as lemma); the person of the verb; and the tense identified in the sentence.

The classification tables made in ".csv" help in future processes of development of this work by the team in two ways. Firstly, this file format facilitates the insertion of new information and corrections attending to the team's needs. Secondly, collecting new data by bots is made possible through Web Scraping due to the ease of writing in these files.

Logical rules for grammatical sentence construction. After the previous steps, the elements prepared are sorted to form a new sentence. The allocation order of these elements uses logical rules based on the rules of the grammatical construction of the gloss.

According to the token type, these logic rules classify the grammatical elements in order of priority. They are then allocated in the new sentence translated into oral Portuguese. After assigning the first tokens of each type present in the sentence, the allocation proceeds with the subsequent existing tokens for each type, following the priority order defined by the rules until the translated sentence is elaborated, going through all the construction rules.

One future improvement aspect for elaborating new sentences is due to the limitation of the number of construction rules, where several tokens of the same type are not considered. Therefore, if too many tokens of the same type are inserted in the sentence (e.g., too many verbs or too many nouns), some will not be considered in the returned translated sentence. Another aspect is that tokenization is performed for a corpus of written Portuguese and not for glosses, which may introduce syntactic classification errors in certain situations.

It is important to note that the strategy chosen for the oral interlingua Portuguese translation was based on rules. That is more suitable due to the SL context for not requiring a bilingual corpus since SL has practically no linguistic corpus. Given the visuospatial nature of SL, they do not have literate registers like oral languages, and even using the gloss doesn’t make that possible, as it still does not have well-defined rules to express all the semantic details existing in SL (e.g., facial expressions, body expressions).

The results of the proposed process's application and its application's context will be presented in the next section.

4. Experiments and Results

A preliminary gloss-to-text translator for Libras was configured in the context of banking situations faced by deaf users to validate the proposed approach.

Some examples of the application of these rules are presented in Table I. The sentences in this table were obtained with two deaf users with different knowledge of interlingua (see section 2), where one of them know only a little bit of oral Portuguese, closer to glosses, while the second deaf user is more proficient, closer to oral Portuguese. Despite many sentences were produced by these two users, only relevant ones to this work were chosen, avoiding information repetition.

Table 1. Translator Input/Translator Output/Expected Phrases

Translator input (Simple sentences)	Translator output (Simple sentences) **P: Portuguese, E: English**	Expected phrase (Simple sentences)
(P) eles depositar hoje dinheiro	(P) eles depositam dinheiro hoje	(P) Eles depositam dinheiro hoje.
(E) they to deposit money	(E) they deposit money today	(E) They deposit money today.
(P) ele devolver amanhã dinheiro	(P) ele devolverá dinheiro amanhã	(P) Ele devolverá dinheiro amanhã.
(E) he to return tomorrow money	(E) he will return money tomorrow	(E) He will return the money tomorrow.
(P) nós sacar amanhã dinheiro	(P) nós sacaremos dinheiro amanhã	(P) Nós sacaremos dinheiro amanhã.
(E) we to withdraw money	(E) we will withdraw money tomorrow	(E) We will withdraw money tomorrow.
(P) eu emprestar dinheiro	(P) Eu empresto dinheiro	(P) Eu empresto dinheiro.

(E) I to borrow money	(E) I borrow money	(E) I borrow money.
Translator input (compound sentences)	Translator output (compound sentences)	Correct sentences (compound sentences)
(P) eles depositar hoje dinheiro nós sacar amanhã dinheiro	(P) eles depositam hoje dinheiro sacaremos dinheiro	(P) Eles depositam dinheiro hoje. Nós sacaremos dinheiro amanhã.
(E) they to deposit today money we to withdraw tomorrow money	(E) they deposit today money we will withdraw money	(E) They deposit money today. We will withdraw money tomorrow.
(P) eu emprestar dinheiro ele devolver amanhã dinheiro	(P) eu emprestarei amanhã dinheiro devolve dinheiro	(P) Eu empresto dinheiro. Ele devolverá dinheiro amanhã.
(E) I to borrow money he to return tomorrow money	(E) I will borrow tomorrow money return money	(E) I borrow money. He will return money tomorrow.

For this version, only some morphological rules of glosses were modeled. When the application of the rules required a more specific treatment, in addition to the generic operations discussed in Section III, SpaCy resources (SPACY, 2022) and auxiliary dictionaries were also used. These resources were used, for example, to define the treatment of verbs and plural words and identify specific words.

It is possible to verify that the translation has satisfactory results compared to the written Portuguese expected response to simple sentences. However, translation becomes more difficult as the sentence increases in complexity, as seen at the bottom of Table I for compound sentences.

It can be verified from this table and from the other tests performed that:

- Longer sentences rarely have punctuation.
- Sentences with more than one paragraph are not correctly translated.
- Additional marking information is needed in sentences entered by the deaf for proper translation (e.g., adverbs of time).
- Sentence punctuation is not entered as: question mark, exclamation point and period.
- If this information is not provided, the “defaults” situations will be used, which may not match the reality of the construction of the sentence.
- The fact that the corpus of oral Portuguese was used for labeling also made translation more difficult.

However, it should be noted that the purpose of the proposed solution is not to replace or match human translators but to improve access to information, especially when interpreters are not available.

The following section presents the main conclusions obtained in the research so far, and the related future work will be discussed.

5. Conclusions

In this work, a new automatic translation strategy was proposed for an interlingua input text for listeners of Portuguese that incorporates morphological-syntactic aspects and a proof of concept.

However, the tests that have been carried out by a group of deaf collaborators of the project have pointed out some problems that still need to be solved. One of them directly addresses issues related to the syntax of both languages, provoked more specifically by the structural (syntactic) differences between them. To solve this problem, the specification and implementation of a grammar defined through a rule de-scription language will be needed, with the ability to translate written texts (typed) in Libras into Portuguese.

Although Libras already has a writing system that, despite being very little spread and known by the deaf, it is worth clarifying that using the solution proposed in this article should be relatively easy, requiring only the basic knowledge of the Portuguese language from its users. This knowledge is essential because, in the proposed solution, writing will be done using words (interlingua) that are equivalent to the intended signs but ordered according to the syntactic structures provided by the Libras gram-mar.

As contributions to this research, the following stand out:

- Proposal of a solution to the problem of texts typed by the deaf (interlingua) and their translation for the hearing person.
- The observation of the need to create a grammar of glosses as future work to improve the proposed process.
- Need for additional information for marking sentences inserted by the deaf for proper translation. This contribution is especially important when applied to UX development in building an interface to support the words typed by deaf users.

The main limitation of this work is that was not possible yet to apply the prototype in a large group of deaf users, but only in two opposite users: one of them with little knowledge of oral Portuguese and another one with a good knowledge of oral Portuguese. This is especially due to some still present limitations of the grammar rules, which is not yet being capable of considering other levels of interlingua used by the great variety of deaf users. Nonetheless, the grammar rules are being constantly re-fined, which will comprehend more and more users over time.

Another aspect is that tokenization is performed for a corpus of written Portuguese and not for glosses, which may introduce syntactic classification errors in certain situations.

6 Future Works

As pointed out, it will be needed to create a grammar that translates the interlingua defined in this work to the hearing Portuguese language with aspects that was proved to be important as described is this work as: additional marking information is need-ed in sentences entered by the deaf for proper translation (e.g., adverbs of time), or sentence punctuation as: question mark, exclamation point and period, as can be seen in section 4.

The main goal of this work in the future is to improve the scalability of this translator, constantly refining the grammar of the translator proposed, meeting the needs of a bigger margin of deaf users, each one with different level of knowledge in interlingua. To achieve this, a contribution of a wider range of users with different characteristic will be searched for dataset improvements and application tests.

Acknowledgment: This research was partially funded by Lenovo, as part of its R&D investment under Brazil's Informatics Law. The authors want to acknowledge the support of Lenovo R&D and Instituto de Pesquisas Tecnológicas do Estado de São Paulo (IPT).

References

ACESSIBILIDADE BRASIL. **Dicionário da Língua Brasileira de Sinais V3. 2011.** Disponível em: http://www.acessibilidadebrasil.org.br/libras_3/. Acesso em: 16 OUT. 2022.

ANDROIDBLIP. **ASL Keyboard for Android.** Disponível em: https://www.androidblip.com/android-apps/com.mindrocketsinc.signskeyboard.html. Acesso em: 16 OUT. 2022.

ASLIZED!. **Innovations.** Disponível em: http://aslized.org/innovations/. Acesso em: 16 OUT. 2022.

BIDARRA, J. PORLIBRAS: Fundamentos para a Implementação de Ferramentas Computacionais para suporte ao Desenvolvimento de um Sistema bilíngue de Tradução Automática Português-Libras. Projeto de Pesquisa. **Universidade Estadual do Oeste do Paraná (UNIOESTE)**, 2018(em curso).

CANALTECH. **Estudantes desenvolvem primeiro aplicativo tradutor de Libras para o português. 2019.** Disponível em: https://canaltech.com.br/apps/estudantes-desenvolvem-primeiro-aplicativo-tradutor-de-libras-para-o-portugues-138565/. Acesso em: 16 OUT. 2022.

CHOWDHARY, K. R. **Natural Language Processing.** New Delhi: Springer India, 2020, p. 603–649.

CORDER, S. P. **A Role for the Mother Tongue.** S. G. A. S. S. and L. SELINKER, Eds., Amsterdam: John Benjamins, 1992, p. 18.

DE LIMA, V. L. S; ABRAHÃO, P. R. C.; FILHO, I. S. M. Distributing linguistic knowledge in a multi-agent natural language processing system: remodelling the dictionary. **Procesamiento del lenguaje natural**, no. 23, p. 104–109, 1998.

ELLIS, R. **Understanding second language acquisition.** Oxford: Oxford University Press, 1985.

FERREIRA-BRITO, L.; LANGEVIN, R. **Sistema Ferreira Brito-Langevin de Transcrição de Sinais**, Rio de Janeiro: Tempo Brasileiro, 1995.

GOVERNO DIGITAL. **VLibras.** Disponível em: https://www.vlibras.gov.br/. Acesso em: 16 OUT. 2022.

HAND TALK. **Hand Talk Plugin.** Disponível em: https://www.handtalk.me/br/. Acesso em: 16 OUT. 2022.

HOPP, F. R et all. The extended Moral Foundations Dictionary (eMFD): Development and applications of a crowd-sourced approach to extracting moral intuitions from text. **Behavior Research Methods**, vol. 53, no. 1, p. 232–246, February 2021.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Censo demográfico do IBGE, 2010**. Disponível em: https://biblioteca.ibge.gov.br/visualizacao/periodicos/94/cd_2010_religiao_deficiencia.pdf. Acesso em: 16 OUT. 2022.

KLEIN, D.; MANNING, C. D. Natural language grammar induction with a generative constituent-context model. **Pattern Recognition**, vol. 38, no. 9, p. 1407–1419, September 2005.

KLIMA, E.; BELLUGI, U. **The signs of language.** Cambridge, Mass: Harvard University Press, 1979.

KOEHN, P. **Statistical Machine Translation.** Cambridge: Cambridge University Press, 2010.

KOEHN. P. **Neural Machine Translation.** Cambridge: Cambridge University Press, 2020.

LABROOTS. **Sign Language App Revolutionizes Messaging for the Deaf and Hard of Hearing Community. 2018.** Disponível em: https://www.labroots.com/trending/technology/7757/five-app-allows-deaf-people-send-animated-sign-language. Acesso em: 16 OUT. 2022.

LIBRASIL. **Librazuka.** Disponível em: https://www.librazuka.com.br/. Acesso em: 16 OUT. 2022.

LIBRASOL. **Divulgado resultado da consulta pública sobre ferramenta VLibras. 2020.** Disponível em: https://www.librasol.com.br/apos-consulta-sobre-vlibras-governo-promete-ampliar-acessibilidade-para-surdos/. Acesso em: 16 OUT. 2022.

NADKARNI, P. M.; OHNO-MACHADO, L.; CHAPMAN, W. W. Natural language processing: an introduction. **Journal of the American Medical Informatics Association**, vol. 18, no. 5, p. 544–551, September 2011.

NEMSER, W. Approximative systems of foreign language learners. **IRAL - International Review of Applied Linguistics in Language Teaching**, vol. 9, no. 2, p. 15–123, 1971.

NLTK. **Natural Language Toolkit.** Disponível em: https://www.nltk.org/. Acesso em: 16 OUT. 2022.

NORTH, W. **Handbook of Semiotics.** Bloomington, IN: Indiana University Press, 1990.

OTHERO, G. A. **A gramática da Frase em Português: algumas reflexões para a formalização da estrutura frasal em português.** Porto Alegre: EdiPUCRS, 2009.

PANDEY, S.; PANDEY S. K. Applying Natural Language Processing Capabilities in Computerized Textual Analysis to Measure Organizational Culture. **Organizational Research Methods**, vol. 22, no. 3, p. 765–797, July 2019.

PIRIS, Y.; GAY, A.-C. Customer satisfaction and natural language processing. **Journal of Business Research**, vol. 124, p. 264–271, January 2021.

QUADROS, R. M.; KARNOPP, L. B. **Língua de sinais brasileira: estudos linguísticos.** Artmed Editora, 2004.

LIBRASOL. **Divulgado resultado da consulta pública sobre ferramenta VLibras. 2020.** Disponível em: https://www.librasol.com.br/apos-consulta-sobre-vlibras-governo-promete-ampliar-acessibilidade-para-surdos/. Acesso em: 16 OUT. 2022.

RYBENÁ**. Rybená - Tradutor de Português-Libras.** Disponível em: https://portal.rybena.com.br/site-rybena/. Acesso em: 16 OUT. 2022.

SELINKER, L. La interlengua. **La adquisición de las lenguas extranjeras**, July 1992.

SIGNALL. **We translate sign language. Automatically.** Disponível em: https://www.signall.us/chat. Acesso em: 16 OUT. 2022.

SIGNASL**. American Sign Language Dictionary.** Disponível em: https://www.signasl.org/. Acesso em: 16 OUT. 2022.

SIGNASL**. British Sign Language Dictionary.** Disponível em: https://www.signbsl.com/. Acesso em: 16 OUT. 2022.

SPACY. **Industrial-Strength Natural Language Processing in Python.** Disponível em: https://spacy.io/. Acesso em: 16 OUT. 2022.

STOKOE, W. C. **Sign Language Structure.** Silver Spring. Linstok Press, 1960.

STOKOE, W. C; CASTERLINE, D. C; CRONEBERG, G. **A dictionary of American Sign Language on linguistic principles**, DC: Gallaudet College. Linstok Press, 1965.

THE ASL APP. **The ASL App.** Disponível em: https://theaslapp.com/. Acesso em: 16 OUT. 2022.

UNIVERSAL DEPENDENCIES. **Universal POS tags.** Disponível em: https://universaldependencies.org/u/pos/. Acesso em: 16 OUT. 2022.

WIKISIGNS. **Sign Language Dictionaries of the World.** Disponível em: http://www.wikisigns.org/. Acesso em: 16 OUT. 2022.

WILCOX, S.; WILCOX, P. P. **Learning to see: Teaching American Sign Language as a second language.** Gallaudet, Ed., University Press, 1997.