



Universidade de Brasília

Instituto de Letras (IL)

Departamento de Línguas Estrangeiras e Tradução (LET)

Bacharelado em Letras Tradução Espanhol

FLÁVIA COSTA BOMFIM

**MAPEAMENTO SISTEMÁTICO DE UNIDADES DE TRADUÇÃO:  
proposta de desenvolvimento de algoritmos de reconhecimento de padrões**

Brasília, 2020

FLÁVIA COSTA BOMFIM

**MAPEAMENTO SISTEMÁTICO DE UNIDADES DE TRADUÇÃO:  
proposta de desenvolvimento de algoritmos de reconhecimento de padrões**

Trabalho de Conclusão de Curso de graduação apresentado como requisito para obtenção do título de Bacharel em Letras Tradução Espanhol no Departamento de Línguas Estrangeiras e Tradução, Instituto de Letras da Universidade de Brasília.

Orientador: Prof. Dr.: Gleiton Malta

FLÁVIA COSTA BOMFIM

**MAPEAMENTO SISTEMÁTICO DE UNIDADES DE TRADUÇÃO:  
proposta de desenvolvimento de algoritmos de reconhecimento de padrões**

Trabalho de conclusão de curso de graduação apresentado ao Departamento de Línguas Estrangeiras e Tradução do Instituto de Letras da Universidade de Brasília como requisito parcial para a obtenção do título de Bacharel(a) em Letras Tradução Espanhol.

Aprovado em:

16 de dezembro de 2020.

BANCA EXAMINADORA

---

Prof. Dr. Gleiton Malta - UnB (orientador)

---

Prof. Dr. Luís Carlos Ramos Nogueira - UnB

---

Prof. Dr. Pablo Cardellino Soto - UnB

BRASÍLIA

Dezembro, 2020



Dedico este trabalho ao Prof. Dr. Gleiton Malta, que incentivou o desenvolvimento de ferramenta que possa apoiar tradutores e pesquisadores em Estudos da Tradução. Obrigada por apontar o caminho.

## AGRADECIMENTOS

A todos aqueles que, direta ou indiretamente, contribuíram para realização deste trabalho, os meus agradecimentos sinceros e emocionados.

- Ao professor Malta, pela orientação competente, presente e especial;
- À professora e coordenadora de curso Magali Pedro, que sempre muito dedicada, apoiou os alunos com esclarecimentos e ações positivas;
- Às professoras e aos professores por compartilharem comigo seus conhecimentos e suas experiências;
- Aos meus amigos de curso e de trabalho, pela troca de experiências e apoio incondicional;
- À Universidade de Brasília que se manteve funcionando mesmo em momento de pandemia; e
- À minha amorosa mãe, que cuidou de mim.

*Ora, suponhamos que a mensagem posta dentro da garrafa seja encontrada por um estudioso de linguística, hermenêutica ou semiótica. Esse novo destinatário acidental (um pouco mais sabido do que Robinson [Crusoé]) estará apto a elaborar um monte de hipóteses, e das mais sutis.*

Umberto Eco

## RESUMO

Este trabalho, inserido no campo disciplinar dos Estudos da Tradução, mais especificamente em seu ramo aplicado orientado às ferramentas de auxílio à tradução (*translation aid software*), tem o objetivo de apresentar o *Grinder*, uma ferramenta criada para o mapeamento sistemático de unidades de tradução. O trabalho está organizado de forma a expor o desenvolvimento da ferramenta *Grinder*, desde sua concepção à apresentação de resultados preliminares. O referencial teórico foi organizado para elucidar questões em torno de unidades de tradução por meio de postulados referentes à convencionalidade das línguas e às práticas de inteligência artificial em interface com contribuições realizadas por autores e pesquisadores em Estudos da Tradução. Os resultados indicam que se pode mapear unidades de tradução a partir de sentenças bem formadas na língua fonte e na língua alvo, uma vez que a comunicação se sustenta dos significados semânticos cristalizados (modelo superpragmático de interpretação). O trabalho faz parte do rol dos estudos empreendidos pelo grupo de pesquisa MapTrad (Mapeamentos em tradução) registrado no CNPq e certificado pela Universidade de Brasília.

### **Palavras-chaves:**

*Grinder*, unidade de tradução, modelo superpragmático de interpretação, convencionalidade, intercompreensão, inteligência artificial, algoritmo, reconhecimento de padrões.



## RESUMEN

Este trabajo se inserta en el campo disciplinar de los Estudios de Traducción, más concretamente en la rama de los Estudios Aplicados de Traducción, basados en herramientas de apoyo a la traducción (*translation aid software*), y tiene como objetivo presentar *Grinder*, una herramienta creada para el mapeo automático de unidades de traducción. El trabajo está organizado para exponer el desarrollo de la herramienta *Grinder*, desde su concepción hasta la presentación de resultados preliminares. El marco teórico se organizó para dilucidar cuestiones en torno a las unidades de traducción a través de postulados referentes a la convencionalidad de los lenguajes y las prácticas de inteligencia artificial en interfaz con las contribuciones de autores e investigadores en Estudios de Traducción. Los resultados indican que es posible mapear unidades de traducción a partir de oraciones bien formadas en los idiomas de origen y destino, ya que la comunicación se apoya en significados semánticos cristalizados (modelo superpragmático de interpretación). El trabajo es parte de la lista de estudios realizados por el grupo de investigación MapTrad (Mapeamientos en la traducción) registrados en el CNPq y certificados por la Universidad de Brasilia.

### **Palabras clave:**

*Grinder*, unidad de traducción, modelo superpragmático de interpretación, convencionalidad, intercomprensión, inteligencia artificial, algoritmo, reconocimiento de patrones.

## ABSTRACT

This work, inserted in the disciplinary field of Translation Studies, more specifically in its applied branch oriented to translation aid software, aims to present *Grinder*, a tool created for the automatic mapping of translation units. The work is organized in order to expose the development of the *Grinder* tool, from its conception to the presentation of preliminary results. The theoretical framework was organized to elucidate issues around translation units through postulates referring to the conventionality of languages and artificial intelligence practices in interface with contributions made by authors and researchers in Translation Studies. The results indicate that it is possible to map translation units from well-formed sentences in the source and target languages, since communication is supported by crystallized semantic meanings (superpragmatic model of interpretation). The work is part of the list of studies undertaken by the MapTrad research group (Mappings in translation) registered with CNPq and certified by the University of Brasília.

### **Keywords:**

*Grinder*, translation unit, unit of translation, superpragmatic model of interpretation, conventionality, intercomprehension, artificial intelligence, algorithm, pattern recognition.

## SUMÁRIO

1	INTRODUÇÃO .....	13
2	REFERENCIAL TEÓRICO .....	17
2.1	Unidades de Tradução, Equivalência e Tecnologia.....	17
2.2	Linguística de Corpus, Convencionalidade e Unidades de Tradução .....	26
2.3	Intercompreensão e Modelos de Interpretação .....	33
2.4	Reconhecimento de Padrões e Tecnologias da Inteligência.....	37
3	METODOLOGIA .....	47
4	O GRINDER.....	52
4.1	OS ALGORITMOS .....	52
4.1.1	Sentences.....	60
4.1.2	Nodes .....	61
4.1.3	Reduce.....	62
4.1.4	Normalize.....	62
4.1.5	Ordenize .....	63
4.1.6	Filtrate .....	63
4.2	Interface do Usuário.....	59
4.2.1	Configurações.....	67
4.2.2	Modelo Linguístico.....	68
4.2.3	Modelo de Tradução .....	71
5	RESULTADOS PRELIMINARES.....	75
	CONSIDERAÇÕES FINAIS .....	82
	REFERÊNCIA BIBLIOGRÁFICA.....	84
	APÊNDICE A. Unidades de tradução mapeadas na Constituição Brasileira .....	89
	APÊNDICE B. Unidades de tradução mapeadas na versão da Constituição Brasileira .....	94
	APÊNDICE C. Alinhamento entre unidades de tradução da Constituição Brasileira .....	105

## LISTA DE FIGURAS

Figura 1 - Mapa das disciplinas com interface com os Estudos da Tradução .....	23
Figura 2 - Níveis de Convencionalidade .....	29
Figura 3 - Três padrões redundantes para “A” .....	38
Figura 4 - Entrada para Programa de Reconhecimento de Letras .....	41
Figura 5 - Dois níveis de uma Árvore de Busca em amplitude .....	43
Figura 6 - Algoritmo A* .....	45
Figura 7 - Modelo de estrutura de grafos do <i>Grinder</i> .....	55
Figura 8 - Estrutura de grafos do <i>Grinder</i> .....	58
Figura 9 - Tela de Configurações do <i>Grinder</i> .....	67
Figura 10 - Tela de Configurações do <i>Grinder</i> .....	69
Figura 11 - Barra de ferramentas da tela de Modelo Linguístico .....	69
Figura 12 - Modelo Linguístico de Constituições de Língua Portuguesa .....	70
Figura 13 - Modelo Linguístico de Constituições de Língua Espanhola .....	70
Figura 14 - Tela de modelo linguístico .....	71
Figura 15 - Barra de ferramentas da tela de Modelo de Tradução .....	72
Figura 16 - Modelo de Tradução da Constituição Brasileira .....	73
Figura 17 - UTs da Constituição Brasileira .....	73
Figura 18 - Arquivo .json do modelo linguístico das Constituições hispanoamericanas .....	74
Figura 19 - Configuração de arquivos .json de modelos linguísticos .....	74
Figura 20 - Arquivo html com resultado do processamento de texto .....	74
Figura 21 - Arquivo html com resultado do processamento de texto de Rui Barbosa .....	80
Figura 22 - Arquivo html com resultado do processamento de texto de Gabo .....	81

## LISTA DE QUADROS

Quadro 1 - Mapeamento de UTs de Vinay e Darbelnet com base no conto Mateo Falcone ..	18
Quadro 2 - Modelos de Tradução identificados por Barbosa .....	20
Quadro 3 - Conceitos de unidade de tradução levantados por Alves .....	25
Quadro 4 - Convencionalidade do espanhol / português .....	30
Quadro 5 - Correspondências entre línguas românicas .....	34
Quadro 6 - Algoritmo A* .....	44
Quadro 7 - Unidades de tradução da Constituição brasileira .....	105

## 1 INTRODUÇÃO

Entre as notícias publicadas durante a Copa do Mundo de 2018, algumas deram destaque ao uso do aplicativo Google Tradutor<sup>1</sup> nos celulares dos torcedores na Rússia. Embora produzisse algumas traduções imprecisas ou inadequadas, os torcedores relataram o quão indispensável o aplicativo provou ser para contornar as barreiras linguísticas com os cidadãos locais. O Google Tradutor faz parte da nova geração de sistemas de tradução automática como Bing<sup>2</sup>, DeepL<sup>3</sup>, Linguee<sup>4</sup>, Reverso<sup>5</sup> e Babylon<sup>6</sup>, que são capazes de entregar algumas traduções com aspecto natural, inclusive em sentenças longas e complexas, graças às abordagens estatísticas da Linguística de Corpus na década de 1990. Esse tipo de abordagem requer grandes quantidades de dados, o que explica por que essa proposta não se tornou popular antes disso.

O início da sinergia entre tradução e tecnologia está registrado nos anais da história da computação<sup>7</sup>. Em meio a componentes eletrônicos e conceitos que foram o sustentáculo para o advento do computador, estão os primeiros sistemas de tradução automática, doravante TA. A Fundação Rockefeller financiou projetos de desenvolvimento em computação e Andrew Booth, celebrado por suas contribuições no desenvolvimento de dispositivos de memória, relatou seu encontro com Warren Weaver em 1947, à época presidente da Fundação, para obter apoio ao seu projeto ARC (*Automatic Relay Computer*). Cálculos matemáticos não seriam atrativos e então pensaram na tradução como se as línguas fossem códigos que pudessem ser quebrados. Booth obteve o apoio que esperava e a ideia de tratar línguas como códigos se mostrou impraticável. As pesquisas de TA, no entanto, seguiram.

---

<sup>1</sup> Disponível em: <https://translate.google.com.br/hl=pt-BR>. Último acesso em: 01 dez. 2020.

<sup>2</sup> Disponível em: <https://www.bing.com/translator/?cc=pt>. Último acesso em: 01 dez. 2020.

<sup>3</sup> Disponível em: <https://www.deepl.com/translator>. Último acesso em: 01 dez. 2020.

<sup>4</sup> Disponível em: <https://www.linguee.com/portugues-spanish/>. Último acesso em: 01 dez. 2020.

<sup>5</sup> Disponível em: [https://www.reverso.net/text\\_translation.aspx?lang=PT](https://www.reverso.net/text_translation.aspx?lang=PT). Último acesso em: 01 dez. 2020.

<sup>6</sup> Disponível em: <https://tradutor.babylon-software.com/>. Último acesso em: 01 dez. 2020.

<sup>7</sup> Disponível em: <https://history.computer/pioneers/> e <http://www.mt-archive.info/Weaver-1949.pdf>. Último acesso em: 01 dez. 2020.

Entre 1948 e 1965, setenta grupos de pesquisa em quinze países trabalharam no desenvolvimento de TA, “atenção em tempo integral de alguns milhares de trabalhadores distribuídos de leste a oeste” (BOOTH, 1965, p.7). Alguns dos projetos eram bastante promissores, como o caso do projeto do Instituto Central de Pesquisa Russo para tradução de informações sobre patentes. Tratava-se de um algoritmo baseado na análise de segmentos para delineamento de unidades de texto, tais como grupos de nomes e combinações verbais. “A tradução era realizada com a ajuda de um dicionário compilado de padrões especificados de construções sintáticas” (JOSSELSO, 1971, p.43). E aconteceu um hiato nas pesquisas após o ALPAC (*Automatic Language Processing Committee*), comitê constituído em 1964 para avaliar as perspectivas da TA, publicar seu relatório em 1966<sup>8</sup>.

A conclusão a que se chegava era que, em 10 anos, o governo estadunidense havia gasto cerca de US\$ 20 milhões em TA, sem que houvesse emergência no campo da tradução. “Existem, contudo, vários problemas cruciais de tradução. Esses são qualidade, velocidade e custo” (ALPAC, 1966, p. 16). Mais cara e menos acurada que a tradução provida por um tradutor humano, a repercussão do relatório foi negativa e as consequências foram cortes de recursos dos programas de pesquisa ligados à TA, não só nos Estados Unidos como em outros países. Contudo, os resultados promissores do algoritmo russo e de outros algoritmos persistiram na linguística quanto ao desenvolvimento de procedimentos de reconhecimento automático nos níveis sintático e semântico, “útil não apenas na análise de sentenças, mas também na construção de regras para geração de sentenças” (JOSSELSO, 1971, p.49).

Possamai (2010) demonstrou a relação fundamental entre Tecnologia e Tradução “no que tange ao levantamento de dados que possam ser usados no auxílio a tradutores como no estudo da tradução e dos textos especializados” (p. 229). A partir de coletânea de textos especializados de Medicina, empreendeu importantes estudos empíricos que comprovaram e demonstraram o perfil das associações no nível sintagmático. Possamai (2010) observou que tais associações são combinações mais ou menos fixas e frequentes de palavras, nem sempre são contínuas, estão associadas no nível sintagmático por uma atração semântica e/ou pragmática e/ou gramatical, e não necessariamente contêm um termo<sup>9</sup> em sua composição, chegando à conclusão que “essas associações se configuram como unidades de tradução

---

<sup>8</sup> Disponível em: <http://www.mt-archive.info/ALPAC-1966.pdf>. Último acesso em: 01 dez. 2020.

<sup>9</sup> Para Possamai (2010), termo é uma expressão lexical associada a um conceito. Seguiremos o mesmo entendimento.

Especializadas porque são recorrentes, fazem parte dos ‘modos de dizer’ dos gêneros e do domínio em questão e apresentam riscos e vantagens ao tradutor” (POSSAMAI, 2010, p. 23). Neste trabalho, seguiremos a trilha aberta por Possamai, pretendendo também contribuir “para o paradigma de desenvolvimento de novas maneiras de desenvolver produtos de auxílio e referência para tradutores” (POSSAMAI, 2010, p. 228).

Este trabalho, inserido no campo disciplinar dos Estudos da Tradução, mais especificamente em seu ramo aplicado orientado às ferramentas de auxílio à tradução (*translation aid software*), tem o objetivo de apresentar o *Grinder*, uma ferramenta (reunião de algoritmos) criada para o mapeamento sistemático de unidades de tradução. O trabalho faz parte do rol dos estudos empreendidos pelo grupo de pesquisa MapTrad (Mapeamentos em tradução)<sup>10</sup> registrado no CNPq e certificado pela Universidade de Brasília. Para tanto, utiliza-se de postulados referentes à convencionalidade das línguas (i.e. HALLIDAY e HASAN, 1976; SINCLAIR, 2004; TAGNIN, 2013; JENSEN, 2013) e às práticas de inteligência artificial (i.e. RICH, 1988; LÉVY, 2010; DASCAL, 2003) em interface com contribuições realizadas por ALVES (2005, 2018), BAKER (1993), BARBOSA (1990), HOUSE (2014), PAGANO (2018), PYM (2017), SCHNAIDERMAN (2011) e TOURY (1995).

Por ser a tradução uma operação linguística em sua essência e os computadores não compartilham da capacidade humana de utilizar uma língua natural, servindo-se de conhecimento léxico, sintático, semântico e pragmático, é preciso partir de um dos conceitos centrais da teoria da tradução — a equivalência na tradução — de modo que raciocínio estatístico e probabilístico seja empregado para reconhecimento de padrão. Daí esse resgate ingênuo da equivalência, necessário para o desenvolvimento de ferramentas de assistência à tradução, leva diretamente ao debate sobre a tradução e os fenômenos linguísticos, sobretudo na relação entre o original e a tradução. Não é objetivo deste trabalho ponderar sobre equivalência e qualidade de tradução, mas mapear unidades de tradução partindo do conceito de equivalência associado ao raciocínio estatístico para sair do modelo criptográfico e chegar ao modelo superpragmático de Dascal (2003).

De modo prático, a tarefa a ser cumprida pelos algoritmos é comparar originais a traduções, compor estrutura de conhecimento que promova o raciocínio estatístico suficiente para

---

<sup>10</sup> Disponível em: <http://dgp.cnpq.br/dgp/espelhogrupo/2060501981311209>. Último acesso em: 01 dez. 2020.

mapear unidades de tradução, considerando que em texto especializados há frequentes combinações de palavras mais ou menos fixas, tal como demonstrado por Possamai (2010). Esta abordagem é uma tentativa de simplificar a sistematização de unidades de tradução a partir de corpora paralelos: “é difícil, e como é árdua a tarefa de entender e sistematizar tantos dados e que nos são apresentados com tantas variáveis. Há que se ter um tanto de resignação para entender dados que escapam aos desejos iniciais”. (POSSAMAI, 2010, p.229)

O objetivo geral deste trabalho é apresentar o *Grinder* como uma ferramenta que possa contribuir com os Estudos da Tradução ao ser capaz de realizar mapeamento sistêmico de unidades de tradução. Quanto aos objetivos específicos, esperamos:

1. Apresentar propostas linguísticas que apresentem o uso da língua em escolhas sistêmicas;
2. Descrever procedimentos para mapeamento de unidades de tradução com o uso de algoritmos de processamento de texto e Linguística de Corpus; e
3. Realizar mapeamento de unidades de tradução no par linguístico português-espanhol no contexto jurídico.

Este trabalho está organizado de forma a expor o desenvolvimento da ferramenta *Grinder*, desde sua concepção com base em referencial teórico, à apresentação de resultados preliminares. O Capítulo 2 começa com uma visão geral de teorias da tradução, debates em torno dos conceitos de equivalência e unidades de tradução e os impactos de ferramentas de Memória de Tradução ao ato de traduzir e aos Estudos da Tradução. Em seguida, concentra-se em fenômenos linguísticos a serem utilizados como parâmetros para o desenvolvimento de algoritmos baseado em técnicas de Inteligência Artificial.

O Capítulo 3 apresenta o projeto dos *corpora* e a metodologia de coleta e seleção de corpora e os procedimentos executados com o *Grinder*. O Capítulo 4 apresenta os algoritmos desenvolvidos, relacionando aos conceitos vistos no Referencial Teórico e apresenta a interface do usuário e como pode ser utilizado, inclusive com “quadriangulação”. O Capítulo 5 enfoca os resultados preliminares obtidos com o *Grinder* no processamento dos corpora, em especial os Corpora selecionados referem-se a Constituições, isto é, as próprias Constituições em língua portuguesa, cujos países têm o português como língua oficial, e Constituições em língua espanhola, cujos países são hispano-americanos. Capítulo 6 é o espaço que utilizamos para reflexões sobre a experiência e desafios futuros.



## 2 REFERENCIAL TEÓRICO

### 2.1 Unidades de Tradução, Equivalência e Tecnologia

Staël, ao longo da obra *Da Alemanha*<sup>11</sup>, destacou as virtudes da tradução quando “não somente podem ser encontradas na tradução as ideias do original, mas também as mesmas sensações” (STAËL, 2017, p. 210). Para a autora, assim como uma música composta para determinado instrumento não poderia ser executada com êxito em outro instrumento, ocorreria o mesmo na tradução; isto é, o alemão acomodava-se à tradução do inglês, porque ambas as línguas teriam a mesma origem, seria difícil, contudo, obter o mesmo resultado em francês. Staël sustentou que as línguas teutônicas traduzem-se muito bem entre si, o mesmo ocorreria entre as línguas românicas, mas as línguas românicas não poderiam exprimir a literatura alemã. Depois haveria diferenças culturais, sobretudo, a liberdade criativa da qual os autores alemães usufruíram. Os franceses não teriam feito justiça à literatura alemã pela barreira da língua e claras diferenças culturais entre as duas nações, segundo a autora.

As fronteiras de traduzibilidade observados por Staël (2017) no trato da língua e da cultura sobre a tradução ainda são válidas, atravessaram o século e se instauraram especialmente nos questionamentos ao redor do conceito de equivalência, imanescentes aos Estudos de Tradução inaugurados no século XX. Vinay e Darbelnet<sup>12</sup> foram pioneiros em descrever os procedimentos de tradução e na “conceituação de *unidade de tradução*<sup>13</sup>, problema que vem a muito preocupando pesquisadores, desde que Vinay e Darbelnet emitiram o conceito” (BARBOSA & NEIVA, 1997, p. 15).

Para qualquer ciência, uma das etapas preliminares essenciais e frequentemente mais controversa é definir as unidades com as quais operar.

---

<sup>11</sup> A primeira edição *De l'Allemagne* foi concluída em 1810 e publicada em 1813.

<sup>12</sup> A primeira edição de *Stylistique comparée du français et de l'anglais: méthode de traduction* foi publicada em 1958.

<sup>13</sup> Do francês *unité de traduction* (UT).

Isso é igualmente verdadeiro para a tradução, onde até recentemente a atenção estava concentrada nas palavras, como se esses segmentos do enunciado fossem tão óbvios que não precisassem de definição. Mas basta olhar as páginas dos principais periódicos linguísticos dos últimos vinte anos para ver que nada é menos claramente definido que o conceito de palavra; alguns linguistas, notadamente Delacroix, chegaram a descrever a palavra como “nebuluse intellectuelle”, ou mesmo se recusaram a considerá-la como tendo qualquer existência concreta (VINAY; DARBELNET, 1995, p. 20).

Para linguistas franceses, os termos unidade de pensamento, unidade lexicológica e unidade de tradução transmitiriam o mesmo conceito, porém com ênfase em diferentes pontos de vista; e postularam as unidades de tradução como unidades lexicológicas dentro das quais os elementos lexicais são agrupados em relação às suas associações para formar um elemento de pensamento, não devendo ser traduzidos individualmente. Também definiram a unidade de tradução como o elemento de pensamento predominante no segmento de tradução, podendo haver superposição de ideias dentro da mesma unidade, no entanto, a tradução seria capaz de reter apenas um significado, de preferência aquele de prioridade no contexto. Subdividiram o conto *Mateo Falcone* (1829) de Prosper Mérimée para exemplificar a técnica de segmentação envolvendo o novo conceito, conforme mostrado no quadro a seguir.

Quadro 1 - Mapeamento de UTs de Vinay e Darbelnet com base no conto Mateo Falcone

1	Fortunato,	1	Fortunato
2	lorgnant	2	kept darting
3	la montre	3	longside glances at
4	du coin de l'oeil,	4	the watch,
5	ressemblait à	5	like
6	un chat	6	a cat
8	qui	7-8	who,
9	l'on	9-10	presented with
10	présente		
11	un poulet	12	a whole
12	tout entier.	11	chicken and
13	Et comme il sent qu'	13	suspecting that
14	on se moque de lui,	14	she is being made fun of,
15	il,		
16	n'		
17	ose	15-7	dares
18	y porter la griffe,	16	not
19	et	18	reach out for it,
20	de temps en temps,	19	and
21	il détourne les yeux	20	at times
22	pour	21	looks away
23	ne pas		
24	s'exposer à		
25	la tentation;	22-5	to resist temptation,

26	mais	28	all the while
27	il se lèche les babines	27	licking her chops
28	à tout moment,		
29	et	29	and
30	il a l'air de	30	wanting to
31	dire à	31	tell
32	son maître:	32	her master
33	“Que	33	how
34	votre plaisanterie	36	mean
35	est	34-5	he is.
36	cruelle!” (Mérimée)		

Fonte: (VINAY; DARBELNET, 1995, p. 47)

Os autores apresentaram a segmentação do texto em unidades de tradução como método de determinação e verificação de conteúdo. Cada segmento formaria uma unidade sintática na qual um elemento determinaria a tradução de outro de maneira que todo o texto seria traduzido, especialmente em casos de sentenças complexas. Vinay e Darbelnet defenderam que um texto subdividido em unidades poderia então ser sistematicamente analisado especialmente quanto aos seus aspectos semânticos e estilísticos que se estendem pelos elementos lexicais puramente formais criados pelos sistemas linguísticos. E, de modo mais prescritivo, definiram a unidade de tradução “como o menor segmento do enunciado cujos signos estão ligados de forma que não devam ser traduzidos individualmente” (VINAY; DARBELNET, 1995, p. 21).

Sob o ponto de vista de Vinay e Darbelnet (1995), são as unidades de tradução que permitem a segmentação de um texto ao mesmo tempo que promovem o limite de contexto necessário à tradução de palavras. Então apresentaram sete procedimentos de tradução em torno das unidades de tradução: empréstimo, decalque, tradução literal, transposição, modulação, equivalência e adaptação. Os três primeiros caracterizam soluções diretas de tradução, os quatro últimos caracterizam soluções oblíquas de tradução. O conceito de equivalência, neste momento, é introduzido como procedimento de tradução em que a unidade de tradução é onomatopeia, expressão idiomática, clichê, provérbio, isto é, haveria equivalências entre a língua fonte e a língua alvo de modo que a mensagem pudesse ser transmitida funcionalmente. Como os linguistas destacaram:

Ao contrário de outros métodos, a necessidade de equivalência, em vez de correspondência um a um, torna-se evidente assim que o significado preciso de um segmento de texto é compreendido. A necessidade de criação de equivalências surge da situação e é na situação do texto que o tradutor deve buscar uma solução. Dessa maneira, a mesma situação pode ser refletida por

meios estilísticos e sintáticos bastante diferentes nas duas línguas. (...) Todas as equivalências resultam de um mesmo processo de reconhecimento global, que, dado um conhecimento sólido de ambas as línguas, abandona a análise das unidades de tradução e se concentra na situação (VINAY; DARBELNET, 1995, pp. 255-256).

Com o termo equivalência, os autores quiseram destacar os segmentos de texto referentes a expressões já conhecidas, portanto segmentos cujos significados não estariam no texto, mas estariam presentes na cultura. Contudo, o termo equivalência foi ganhando novos contornos, descolando-se da unidade de tradução. Nida (1964) foi o primeiro a falar de outros tipos de equivalência, cujo foco deslocou-se do segmento para a mensagem (equivalência formal) ou para a relação entre o receptor e a mensagem (equivalência dinâmica). Com os conceitos de equivalência formal e dinâmica, Nida aventou práticas como comparar continuamente a mensagem na cultura fonte com a mensagem na cultura alvo para determinar padrões de precisão e correção, ou relacionar o receptor a modos de comportamento relevantes dentro do contexto de sua própria cultura para a completa naturalidade de expressão.

Novos modelos de tradução foram propostos de modo que a antiga tensão entre tradução literal e tradução livre motivava a descrição de procedimentos técnicos que orientassem a prática de tradução e as maneiras opostas de se traduzir foram sendo apresentadas como estratégias. Como Barbosa destaca, procedimentos técnicos de tradução são propostos como uma tentativa de responder à pergunta “como traduzir?” (BARBOSA, 1990, p. 63). A partir do trabalho inovador de Vinay e Darbelnet em enumerar procedimentos de tradução, muitos outros subsequentes referiram-se diretamente ao trabalho dos linguistas, ou fizeram reformulação do mesmo. Barbosa realizou revisão da literatura dedicada a esses procedimentos de tradução.

Quadro 2 - Modelos de Tradução identificados por Barbosa

Vinay e Darbelnet (1957):	Tradução direta	Tradução oblíqua
Nida (1964):	Equivalência formal	Equivalência dinâmica
Catford (1965):	Tradução plena Tradução total Tradução limitada Tradução literal	Tradução parcial Tradução restrita Tradução não limitada Tradução livre
Vásquez-Ayora (1977):	Tradução literal	Tradução oblíqua
Newmark (1981):	Tradução semântica	Tradução comunicativa

Fonte: BARBOSA (1990)

Equivalência, como um termo, deixou de designar um procedimento de tradução para tornar-se uma ideia fundamental e, em seguida, tornar-se questionável. Os limites da traduzibilidade foram expostos quando do surgimento de linha de investigação do processo tradutório como essencialmente discursivo, centrada nas dificuldades de tradução e nas diferenças entre os sistemas linguísticos. Nord (2016) acerca da traduzibilidade de textos referendou a quase impossibilidade de reprodução fiel do que é expresso no texto fonte ao texto alvo. Propôs que o conceito de tradução fosse reformulado para que se levasse em conta a transferência intercultural como uma ação comunicativa que sem a tradução não poderia acontecer pelas barreiras culturais e linguísticas existentes.

Corroborando Sanders (2019) à análise de Nord (2016). As palavras intraduzíveis ou “buracos lexicais” existem por conta de vários fatores diferentes, segundo a autora. Um deles é o fato de que algumas palavras nunca foram necessárias no vocabulário de algumas nações, ao mesmo tempo em que tais palavras são como guardiães que preservam a diversidade cultural e linguística no mundo. Sanders (2019) observa que palavras intraduzíveis funcionam como lembretes de que não é possível simplificar tudo. Sanders (2014) listou mais de cinquenta palavras que compartilham do fato de que expressam emoções e são intraduzíveis. Na lista estão inclusas as palavras *sobremesa*, *cafuné* e *saudade*.

Segundo Nord (2016), diante de casos em que a dificuldade é cultural, a tradução buscada no nível linguístico tenderia à intraduzibilidade, a tradução buscada no nível da função comunicativa resolveria a ausência de equivalência porque toda língua seria capaz de conservar os traços funcionalmente relevantes da mensagem original. “Uma tradução ‘fiel’ não se constitui em representar o TP [texto de partida] (...) com a mesma situação, função e efeito, considerando conteúdo e forma, etc., mas sim em informar sobre as características relevantes do TP para o contexto de chegada, preservando a lealdade ao emissor do TP e ao receptor do TC [texto de chegada]” (NORD, 2016, p. 22).

Para Pym (2017), embora a equivalência seja a característica marcante da tradução pelo simples fato de ser tradução, de uma perspectiva crítica, são frequentes as contraposições à sua fruição. “A tradução pode ser definida pela noção de equivalência, mas a equivalência não é um conceito suficientemente estável” (PYM, 2017, p. 13), isto é, se por um lado admite-se que a equivalência é o elo entre o original e a tradução, por outro lado não se alcança a equivalência. Schnaiderman (2011), recapitulando diálogos entre Borges e Sábato, abordou o risco de se seguir, sem maior reflexão, as equivalências indicadas por um dicionário, porque

não existiria entre uma língua e outra coincidência exata e acreditar na existência de equivalentes seria culpa dos dicionários.

Para Schnaiderman (2011), caberia ao tradutor, como criador e artista, dedicar atenção para que o original e a tradução compartilhassem dos mesmos valores semânticos. Segundo o autor, cada tradutor se aperfeiçoa pela autocrítica. Em seu caso, superou riscos a equívocos semânticos com a prática de cotejo com o original, em que a tradução era lida para ele em voz alta enquanto acompanhava a leitura no original, uma prática em que foram precisos anos para desenvolvê-la. Também foi necessário outro longo trabalho seu para que aprendesse que a verdadeira fidelidade só se obteria com arrojo, ousadia, imaginação e liberdade no trato com os textos, pois não se trataria de uma traição ou desapego à fidelidade e sim de permitir maior aproximação da tradução ao original.

Sobre a ideia de traição em oposição à fidelidade, representada pelo adágio *traduttori, traditori*, Pagano (2018) resgata a noção de tradução perfeita como uma transposição ideal e única na possibilidade de um processo tradutório que ocorre em um vácuo temporal e cultural no qual uma ideia formulada em uma língua pode ser transportada automaticamente a outra língua mediada pelas equivalências constantes em dicionários. Quando se realiza comparação entre o original e a tradução, buscando-se preencher requisitos específicos de um avaliador, o resultado será rotulado de traição, imperfeição ou inexatidão.

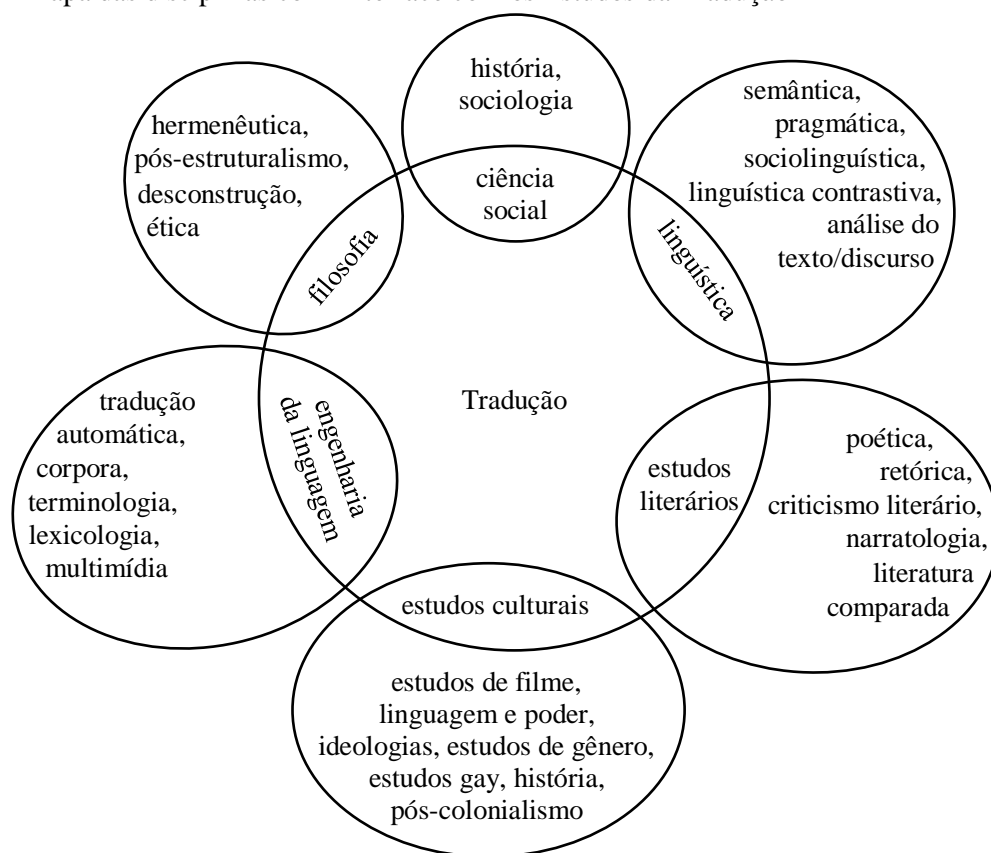
A suposta traição trata-se em verdade de necessário realinhamento à cultura alvo e objetivos para isso. Destacando ao que Pagano (2018) discorre, diferentes traduções podem ser realizadas de um mesmo original, uma vez que há aspectos relacionados à produção e recepção de textos que levam a decisões a serem tomadas na recriação de um texto numa nova língua e cultura que variam conforme objetivos pretendidos, público-alvo, função do texto traduzido e outros fatores mercadológicos. Assim como Schnaiderman (2011), Pagano (2018) atribui ao tradutor a responsabilidade da boa tradução como resultado do exercício consciente da profissão e de sua contínua qualificação.

Todo esse debate em torno da ideia de tradução, isto é, (1) a relação entre original e a tradução, (2) a fidelidade da tradução ao original e (3) a intraduzibilidade versus a traduzibilidade, deslocou a equivalência ao campo da crença, da ilusão, e implicou em expectativas do papel do tradutor no ato de traduzir. Os argumentos ressaltaram a diferença entre as línguas e as escolhas a serem feitas pelo tradutor, o que poderia se resumir na “noção

de tradução presumida” de Toury (1995, p. 33). Ainda que toda tradução seja presumida por razões linguísticas e culturais, Toury (1995) defende que os relacionamentos que se obtêm entre uma tradução presumida e seu original podem ser emparelhados de tal modo a não restar sobras, do inglês *no leftovers* (TOURY, 1995, p.80). Esses pares (parciais), que seriam então submetidos à análise comparativa, produziriam conclusões muito diferentes, incluindo busca de padrões regulares que poderiam ter governado todos esses pares, ou subgrupos deles.

Para Toury (1995) os segmentos do original e da tradução devem ser definidos simultaneamente, determinando um ao outro. Nesse sentido, as unidades de análise comparativa sempre emergiriam como pares acoplados de segmentos. O emparelhamento estaria sujeito a um princípio heurístico em que nenhuma sobra estaria presente. Assim estabelecidos, os pares acoplados seriam submetidos a análises e “as relações encontradas entre seus membros é que fundamentariam qualquer generalização sobre o tipo pertinente de equivalência de tradução” (TOURY, 1995, p. 89).

Figura 1 - Mapa das disciplinas com interface com os Estudos da Tradução



Fonte: MUNDAY & HATIM (2019, posição 444 [versão eletrônica])

A virada cultural nos Estudos da Tradução abriu novos caminhos à tradução e ao papel do tradutor. Paralelamente, e graças ao advento do computador pessoal nos idos da década de

1980, novas interfaces precisaram ser analisadas focalizando a relação tradutor-computador, ou tradutor-tecnologia. Murray e Hatim (2019) então propuseram um novo mapa para representar a interdisciplinaridade dos Estudos da Tradução, fazendo interface com toda uma série de outros campos.

Segundo os autores, há uma miríade de abordagens teóricas e metodologias a partir do cruzamento com novos campos de estudos tão variados. Aquelas pesquisas de tradução que optaram por observar tradutores experientes e/ou estudantes de tradução, levantaram evidências favoráveis ao conceito de unidade de tradução e a tendência pela busca da equivalência. Ballard (2010), por exemplo, identificou que tradutores profissionais pensam instintivamente em termos de unidades de tradução quando desejam avaliar ou descrever o seu trabalho e revela-se igualmente necessário e útil para os teóricos quando pretendem analisar o trabalho do tradutor; e os equivalentes não estavam reportados em dicionários, mas de uma descoberta do tradutor, resultante de um processo de criatividade.

Barbosa e Neiva (1997) empreenderam investigação do processo tradutório de uma tradutora aprendiz. Segundo o relato das pesquisadoras, quando a tradutora analisou seu próprio protocolo verbal em busca de unidades de tradução, identificou que havia segmentado o texto em 314 unidades das quais 28,34% do total ocorreram no nível da oração, 56,6% no nível sintagmático, 12,1% no nível da palavra e 2,86% em combinações de palavras que não correspondem a constituintes sintáticos. Quanto à equivalência, a preocupação com a fidelidade levou à produção de textos que não se ativeram a “critérios de fluência e elegância textuais” (BARBOSA; NEIVA, 1997, p. 21).

Alves (2005, 2018) fez amplo estudo das abordagens teóricas sobre os diferentes conceitos de unidades de tradução (UTs). Como nos lembra o autor, é senso comum que a tradução de um texto seja feita por partes e os Estudos da Tradução dão a essas partes o nome de unidades de tradução. Contudo, ainda há muitas questões sem resposta definitiva em torno das UTs, especialmente ao tamanho e escopo de uma UT que culminam no processo de segmentação. Segundo Alves (2005), o processo complexo, o qual não pode ser delimitado exclusivamente por características morfológicas, lexicais ou sintáticas e é dependente de uma série de fatores cognitivos e contextuais que abarcam como a memória humana, os processos inferenciais de solução de problemas e de tomada de decisão. Temos aqui uma lista sucinta dos diferentes conceitos de unidades de tradução levantados por Alves:



Quadro 3 - Conceitos de unidade de tradução levantados por Alves

Cícero (I a.C.):	Não como o que interpreta, mas como o que fala.
Vinay e Darbelnet (1957):	O menor segmento de um enunciado cuja coesão de sinais seja tal que esses não possam ser traduzidos separadamente.
Haas (1968):	Tão pequena quanto possível e tão longa quando necessária.
Reiß e Vermeer (1984):	O texto é a única unidade de tradução possível.
Newmark (1988):	Quanto mais livre a tradução maior será a unidade de tradução, quanto mais fiel a tradução menor será a unidade de tradução.
Gerloff (1987):	As unidades de tradução estão no nível do morfema, da sílaba, da palavra, da frase, da oração, do período e, até mesmo, no nível do discurso.
Alves (2000):	A unidade de tradução é um segmento do texto de partida, independente de tamanho e forma específicos, para o qual, em um dado momento, se dirige o foco de atenção do tradutor.
Dragsted (2004):	A unidade de tradução é identificada como sendo um segmento linguístico que pode ser processado pela memória de trabalho do tradutor.
Alves (2005):	A unidade de tradução pode ser definida como a compreensão simultânea/consecutiva na língua de partida e a produção na língua de chegada de um segmento textual cujo tamanho é limitado pela capacidade da memória de trabalho e cujas fronteiras podem ser identificadas através de pausas.

Fonte: ALVES (2005, 2018)

Conforme Alves (2005) destaca, os sistemas de memória de tradução configuram mudanças significativas na forma como tradutores segmentam e processam textos para fins de tradução, o que demandaria investigação para explorar a segmentação humana e eletrônica do processo de tradução, discutir o papel da memória humana e computacional e identificar as características básicas dos sistemas de memória em sua interface com o tradutor humana. Na análise de Alves (2005), a segmentação nesses sistemas é sempre feita através de marcadores ortográficos não haveria conhecimento público dos princípios de programação computacional que norteiam esta segmentação. Tais sistemas calculam a semelhança das ocorrências ortográficas identificadas e número de caracteres idênticos que ocorrem em uma ordem específica no texto-fonte e na base de memória, algo muito distinto da segmentação humana.

Segundo Pym (2017), a reutilização de texto promove a generalidade da repetição, o que impede a especificidade da situação, por conseguinte os equivalentes limitam-se cognitivamente ao nível do segmento que resulta em uma confusa mistura de frases, onde o texto-alvo teria características de vários tradutores diferentes e, provavelmente, de diversas situações discursivas, o que só se resolveria com uma série de processos de checagem. E

então, segundo Pym (2017), a revisão ganharia um posto de destaque nos processos de tradução, merecendo teorização própria. Se o conceito de equivalência ainda não tinha atingido estabilidade com décadas de investigações dos Estudos de Tradução, a tecnologia havia promovido uma equivalência padronizada e artificial, certo retorno às linhas de pensamento dos anos 1960 e 1970, “com toda sua simplicidade ilusória” (PYM, 2017, p. 302).

Alves (2005) destaca que os segmentos processados cognitivamente têm uma natureza idiossincrática, comparáveis mediante parâmetros de intersubjetividade entre os tradutores. Em contraposição à segmentação automática, um experimento acerca do conceito de unidade de tradução precisaria, por exemplo, ser operacionalizado em termos de tempo, observando-se as pausas realizadas pelos tradutores ao longo do processo de tradução, um fenômeno existente apenas na mente do tradutor. Os padrões de segmentação cognitiva entre tradutores observados por Alves (2005) seriam decorrentes dos intervalos de pausa durante o tempo de produção textual. A diferença fundamental constatada pelo pesquisador entre os processos de segmentação (pausas *versus* marcações ortográficas para delimitação das unidades de tradução) afeta o processo de tradução em si.

Teorias convencionais acerca da equivalência não contemplam os nuances que a tecnologia provoca no processo de tradução, ao que Pym (2017) reage ao declarar fascínio e objeção à tecnologia: “sou fascinado pelos efeitos da tecnologia, que nos oferecem um futuro melhor, assim como fico chocado com a forma ingênua como a equivalência retornou nesse paradigma [Localização] com toda a sua simplicidade ilusória” (PYM, 2017, p. 302).

O mapeamento sistemático de unidades de tradução consequente ao uso de tecnologia leva à ruptura do arcabouço de teorias e procedimentos de tradução. Novas investigações são requeridas, retomando-se questões latentes há muito abordadas pelos Estudos da Tradução. O armazenamento de unidades de tradução em sistemas de memória resgata, por exemplo, a suposição de equivalência entre línguas ou culturas previamente ao ato de traduzir. As seções a seguir apresentam iniciativas que buscaram alinhar a tecnologia ao uso de línguas ou ao processo tradutório em si e que são base para a construção do *Grinder*.

## **2.2 Linguística de Corpus, Convencionalidade e Unidades de Tradução**

Baker (1993) propôs que haveria benefícios em investigar a tradução partindo de métodos baseados em corpus. A Linguística de Corpus forneceria a metodologia para a realização de

investigações empíricas, enquanto a teoria da tradução identificara as áreas de investigação e elaboraria hipóteses, a fim de pesquisar questões centrais como universais de tradução e procedimentos como explicitação. Segundo Baker (1993), os dois parceiros, Linguística de Corpus e Teoria da Tradução, trabalhariam em harmonia principalmente para o benefício do avanço do ramo descritivo dos Estudos da Tradução. Kenny (2009) centrou-se no conceito-chave da unidade de tradução, destacando que, embora seja um conceito básico, a unidade de tradução recebeu pouca atenção de estudos baseado em corpus, apesar do fato de que corpora paralelos provavelmente contêm uma riqueza de dados sobre unidades de tradução.

Não se questiona que a tradução de um texto se faz por partes, mas o caso é que há várias questões que surgem da discussão sobre unidades de tradução; Baker (1993), Kenny (2009) e Alves (2018) destacaram algumas: se essas unidades são ou não unidades da língua/texto de origem, se são semânticas ou sintáticas, em que nível linguístico são realizadas, se tem qualquer base cognitiva, se são convencionalizadas em qualquer extensão significativa, se as unidades são encontradas no nível da sentença, se dependem da coerência textual, se delimitadas no âmbito do texto, se delimitadas no nível do morfema, qual o tamanho e a natureza da unidade de tradução, ou mesmo o tipo de equivalência que é alcançada pela unidade de tradução na prática.

O uso de sistemas de memórias de tradução conduz o debate acerca dos conceitos tradução, unidade de tradução e equivalência para o nível linguístico, uma vez que as unidades mapeadas, quando reutilizadas, presumem a equivalência anterior ao ato de traduzir. Isso liquefaz a equivalência enquanto estrutura de crença e promove testes reais de sua existência. Segundo House (2014), testar a existência da equivalência, desenterrando as restrições envolvidas na produção, análise ou avaliação do produto de tradução, está no cerne nos estudos empíricos, como os estudos baseados em *corpus*.

Os Estudos da Tradução Baseados em *Corpus* iniciaram com o trabalho inovador de John Laffling no início da década de 1990. Laffling (1991) se baseou em textologia contrastiva para lexicografia bilíngue, identificando equivalentes de tradução de ocorrência natural em textos paralelos e, com isso, defendeu uma modelagem semântica em TA que pudesse enfrentar uma variedade de problemas semânticos. O método consistiu em uma análise sistemática de textos paralelos para “dotar o computador de uma base de conhecimento comparável ao repertório linguístico enriquecido de que o tradutor dispõe” (LAFFLING, 1991, p.83).

John Sinclair fez importantes colaborações para análise de *corpus*, especialmente no reconhecimento das unidades lexicais, ainda na década de 1980. Segundo Sinclair (2004), padrões de colocação são muito presentes, a idiomaticidade seria tão significativa quanto a sintaxe na construção do significado. Em sua teoria das unidades de significado, o autor ilustrou como os padrões lexicais e os padrões sintáticos não poderiam ser divorciados, como as escolhas de padrões lexicais envolveriam padrões de significado e cada significado distinto de uma palavra estaria associado a uma forma distinta. Haveria uma indivisibilidade de forma e significado.

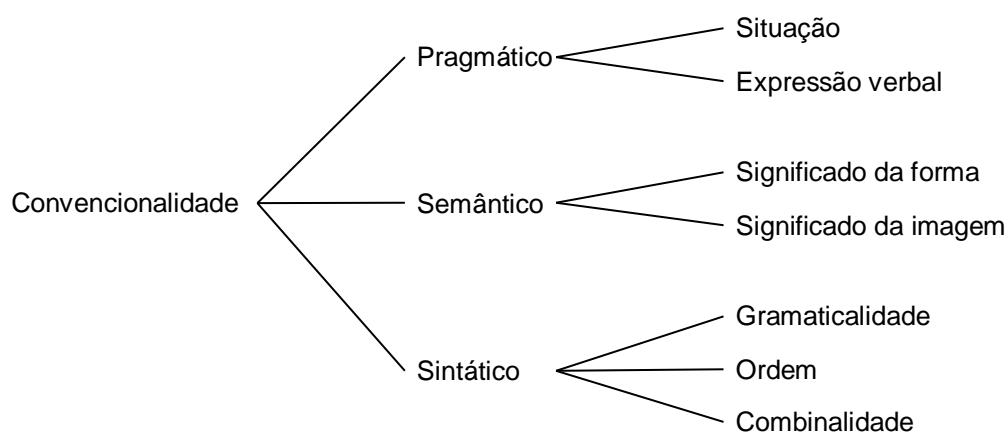
Na acepção de Sinclair (1991), uma unidade de significado é a combinação de uma palavra com outras “para entregar uma única multi-palavra” (p. 71). Essa posição descritiva de Sinclair (2004) foi absorvida de extensas coleções de textos com milhões de palavras e suporte estatístico fornecido pelas técnicas de Linguística de Corpus. Cada detalhe do texto para Sinclair (2004) deveria ser examinado, não só pela perspectiva do próprio texto, mas também deveria ser examinado como parte de uma rede de outros textos no corpus, de modo que as afirmações feitas sobre os textos fossem verificadas com referência a outros textos do mesmo tipo ou semelhantes ou de tipos diferentes, cada um com sua própria disposição de características linguísticas e cada um com sua própria proveniência no corpus.

A partir de uma série de estudos de textos do inglês moderno, em 1976, Halliday e Hasan apresentaram o conceito de coesão em diversos níveis e não puderam deixar de observar as várias relações lexicais que caracterizam relações de significado e trataram de atribuir o título de colocação ou coesão colocacional para a coesão que resulta da co-ocorrência de itens lexicais que estão de uma forma ou de outra tipicamente associados uns aos outros, porque tendem a ocorrer em ambientes semelhantes: “os tipos específicos de relações de co-ocorrência são variáveis e complexas e teriam que ser interpretados à luz de uma descrição semântica geral da língua inglesa” (HALLIDAY; HASAN, 1976, p. 287).

Tagnin (2013) afirma que as convenções linguísticas são os “jeitos” aceitos pela comunidade que fala determinada língua. Em suas investigações concluiu que há uma gama de unidades de significado convencionais. A pesquisadora chama de convencionalidade esse aspecto que caracteriza a forma peculiar de expressão numa língua ou comunidade linguística e apresenta as variadas associações: coligações, colocações, coletivos, binômios, trinômios, expressões convencionais, expressões idiomáticas, metáforas, símiles, marcadores conversacionais e fórmulas situacionais.

Essa noção de convenção pode se aplicar à língua tanto no nível social, quanto no nível linguístico (figura 2). O foco de Tagnin (2013) é o aprendizado da língua estrangeira, contudo conhecer os padrões que a pesquisadora mapeou é muito útil para o mapeamento sistemático de unidades de tradução.

Figura 2 - Níveis de Convencionalidade



Fonte: TAGNIN (2013, posição 413 [versão eletrônica])

O nível sintático compreende a faculdade que os elementos linguísticos têm de se combinar (combinabilidade), sua ordem e sua gramaticalidade. Existem palavras que se associam de forma tão natural que, para Tagnin (2013), essa associação só poderá ter sido consagrada pelo uso, tornando-se convencional. Casos brasileiros que co-ocorrem preferencialmente: “velha coroca”, “lucros e perdas”, “arroz e feijão”, “Feliz Natal”, “mundos e fundo”, “mão de obra escrava”, “roxo de ciúmes”.

No nível semântico observa-se a convencionalidade na relação não motivada entre uma expressão e seu significado como “bater as botas” que significa “morrer”. O significado de uma imagem também pode ser convencionalizado, como é o caso das metáforas. Na cultura ocidental, tudo que é “para cima” é considerado bom, enquanto o que for “para baixo” é mau. São exemplos: “ela está num alto astral”, “ela vive na fossa”, “caiu de cama”, “cabeça erguida”, “toca a levantar os ânimos”.

A ordem em que os elementos ocorrem também pode ser resultado de convenção, “cama e mesa e sempre”, “feijão com arroz”, “de vez em quando” nessa ordem. Esse fenômeno aplica-se principalmente aos binômios, que podem ser considerados um tipo especial de colocação. Um binômio é geralmente formado por duas palavras pertencentes à mesma categoria gramatical e ligado por conjunção ou preposição.

O nível pragmático abrange o uso da língua em situações de interação entre falantes. Aqui temos dois outros aspectos passíveis de convenção: a situação que exige certo comportamento social e a expressão a ser empregada nessa ocasião. Por exemplo, sempre que se recebe algo de alguém a situação exige que o receptor agradeça o que foi recebido. Agradecer é o comportamento exigido pela situação. Esse comportamento pode ser realizado verbalmente, mediante o uso de uma expressão linguística: “obrigado”, “muitíssimo obrigado” e outras similares.

O foco de Tagnin (2013) é o aprendizado de língua estrangeira, ela agraciou seu público leitor, alunos, professores e pesquisadores, com quadros comparativos (tabela 2.4), sempre que possível contrastivo, com exemplos similares nas línguas inglês, português, alemão, espanhol, francês e italiano. Como o interesse aqui é pelo par linguístico espanhol-português, citaremos alguns exemplos dos quadros comparativos deste par linguístico a fim de observar se a convencionalidade entre o espanhol e o português é possível de ser mapeado de modo sistemático que possibilite regras de mapeamento de unidades de tradução por algoritmos.

Quadro 4 - Convencionalidade do espanhol / português

Convencionalidade	Espanhol	Português
Coligações	llegar a + lugar	chegar a/em + lugar
	ir en + medio de transporte	ir de + meio de transporte
	parecerse a	parecer(-se) com
	venir a + lugar	vir para/a + lugar
	preocupación por	preocupação com
	respeto hacia	respeito para com
	olor a	cheiro de
	sabor a	gosto de
	triste por + substantivo	triste com + substantivo
	avergonzado de	envergonhado com
	igual que	igual a
	de al lado	ao lado
	hacia atrás	para trás
	desde lejos	de longe
	a causa de	por causa de
	a costa de	à custa de
	en cuanto a	quanto a
	al azar	ao acaso
	por casualidad	por acaso
	al gusto	a gosto
a oscuras	no escuro	
a menudo	com frequência / amiúde	
Colocações	leche entera	leite integral
	azúcar moreno	açúcar mascavo
	papel pintado	papel de parede
	leche descremado/desnatado	leite desnatado

Convencionalidade	Espanhol	Português
	armario empotrado/placad	armário embutido
	pan de molde	pão de forma
	palabra de honor	palavra de honra
	traje de gala	traje a rigor
	carné de conducir	carta de motorista
	libreta de ahorro	caderneta de poupança
	tener éxito	fazer/ter sucesso
	formar parte	fazer parte
	tener en cuenta	levar em consideração
	quedarse dormido	cair/pegar no sono
	volver loco / volverse loco	deixar louco / ficar louco
	ponerse enfermo	ficar doente
	quedarse helado	ficar paralisado (de susto, medo...)
	repetir obstinadamente	repetir insistentemente
altamente ofendido	profundamente ofendido	
Expressões especificadoras de unidade	una barra/tableta de chocolate	uma barra de chocolate
	una barra de jabón	uma pedra de sabão
	un bloque de hielo	um bloco de gelo
	una hoja de papel	uma folha de papel
	una franja de tierra	uma faixa de terra
Coletivos	una nidada de pollitos	uma ninhada de pintinhos
	una colonia de hormigas	uma colônia de formigas
	una bandada de pájaros	um bando de pássaros
	una camada de perritos	uma ninhada de cachorrinhos
	un enjambre de abejas	um enxame de abelhas
Binômios	cara o cruz	cara ou coroa
	blanco y negro	branco e preto
	sí o no	sim ou não
	de acá para allá	daqui para lá
	de allá para acá	de lá para cá
	café con leche	café com leite
	pérdidas y ganancias	perdas e ganhos
	[llevarse] como el gato y el perro	[brigar/ser] como cão e gato
	poco a poco	aos poucos / pouco a pouco
	cara a cara	cara a cara / frente a frente
	más y más	cada vez mais / mais e mais
	día tras día	dia após dia
	oferta y demanda	oferta e procura
	a trancas y barrancas	aos trancos e barrancos
	a capa y espada	com unhas e dentes
	de rompe y rasga	firme e decidido / de pulso
	sin pies ni cabeza	sem pé nem cabeça
a tontas y a locas	de qualquer jeito	
Trinômios	cama, mesa y baño	cama, mesa e banho
	casa, comida y buen sueldo	casa, comida e roupa lavada
	comida, postre y café	almoço, café e sobremesa
	ciego, sordo y mudo	cego, surdo e mudo
	padre, madre e hijos	pai, mãe e filhos
Expressões convencionais	estar fuera de servicio	estar de folga
	tener una rueda pinchada	estar com um pneu furado
	estar de moda	estar na moda
	aprender de memoria	aprender de cor

Convencionalidade	Espanhol	Português
	ser todo un caballero	ser um autêntico cavalheiro
Expressões idiomáticas	estirar la pata	bater as botas
	tomar el pelo a	tirar sarro (com a cara) de
	dar en el clavo	acertar na mosca
	meter la pata	dar bola fora / dar um fora
	no tener pelas en la lengua	não ter papas na língua
Metáfora	altos y bajos (altibajos)	altos e baixos
	levantarle el ánimo a	levantar o ânimo a
	por todo lo alto	com toda a pompa, com luxo
	estar en auge	estar em alta
	estar en su apogeo	estar no auge
	caer en un pozo sin fondo	estar num beco sem saída
	ser un bajón	ser deprimente
	criticar por lo bajo	criticar pelas costas
el tiempo es dinero	tempo é dinheiro	
Símbolos	como una cuba	bêbado como um gambá
	ser pan comido	se sopa
	más feo que pegarle a la madre	mais feio que o diabo
	como anillo al dedo	a calhar; como uma luva
	como unas castañuelas	feliz da vida
Marcadores conversacionais	Desde mi punto de vista...	Do meu ponto de vista...
	Bueno, pero...	Tudo bem, mas...
	Lo que pasa es que...	Acontece que...
	Por cierto...	Aliás... / por falar nisso...
	Yo, que tú, ...	Eu, se fosse você, ...
	Perdona (que te corte)... / perdone (que le interrumpa)...	Desculpe (interromper)...
	Me gustaría decir algo.	Eu queria dizer uma coisa.
	Esto es todo	É isso
	¿Y qué?	E daí?
	Si quieres saber lo que pienso...	Se você quer saber o que eu acho...
	No es asunto tuyo	Não é da sua conta
	Me quitaste las palabras de la boca	Você tirou as palavras da minha boca
	No me consta que...	Não me consta que...
	¿Queda claro?	Ficou claro?
	Perdona, ¿qué has dicho?	Desculpa, o que você disse?
En ningún momento quise decir que...	Em nenhum momento eu quis dizer que...	
Fórmulas situacionais	¿Se puede fumar aquí?	Pode fumar aqui?
	Todo indica que...	Tudo indica que...
	No soy muy amigo de...	Não sou muito fã de...
	No fue de todo mal	Até que não foi tão ruim...
	No saques conclusiones apresuradas	Não tire conclusões precipitadas
	Que te mejores	Melhoras
	¿No te pongas así?	Não fique assim!
	¡A otro perro con ese hueso!	Vai contar para outro!
	Hola, ¿qué tal?	Oi, tudo bem?
	Gracia, muy amable	Obrigado/a, muito gentil de sua parte
	Te / le acompaño en el sentimiento	Meus pêsames
Invito yo	É por minha conta	



Todo esse mapeamento linguístico de Tagnin (2013) e seus colaboradores<sup>14</sup> apresenta diversos aspectos convencionais próprios das culturas espanhola e brasileira tal com as contribuições de Halliday e Hasan (1976) e Sinclair (2004) propuseram. Podemos observar que há a recursividade linguística se revela em diversos níveis: palavras e combinação de palavras (léxico) são empregadas em estruturas conhecidas (sintaxe) e devem fazer sentido (semântica) conforme contexto de uso (pragmática). Em sociedades ocidentais, as palavras são formadas por letras enumeradas no alfabeto latino, letras ou combinação de letras podem ser analisadas pelo som (fonologia) ou pela composição mínima que mantém significado (morfologia).

A próxima seção, seção 2.3, corroborará com os aspectos de convencionalidade linguística, mas sob o ponto de vista de intercompreensão e modelos de interpretação.

### **2.3 Intercompreensão e Modelos de Interpretação**

Segundo Bogantes (2005), Jørgen Jensen foi um gênio das línguas, dominava as línguas românicas, podia passar de uma língua à outra com uma pronúncia impecável, sem esforço e sem misturá-las. Seu segredo seria conhecer muito bem o nível sintático de todas elas, incluindo uma série de dialetos e variantes dessas línguas. Também falava inglês, alemão, russo, latim e grego antigo. A ideia de Jensen (2013) para um projeto de intercompreensão em línguas românicas originou-se da situação linguística de seu país, “em que um dinamarquês, um norueguês e um sueco podem compreender-se entre si falando cada um a própria língua” (JENSEN, 2013, p. 15).

Havia semelhanças evidentes e diferenças sistemáticas entre as três línguas escandinavas, conforme relatado por Jensen (2013). Os programas de ensino das escolas apresentavam aos estudantes as bases da gramática de outras línguas o suficiente para compreendê-las na forma escrita e oral. Então a questão para o linguista dinamarquês foi saber se este tipo de intercompreensão poderia também funcionar entre as línguas românicas e, para começar, entre o português, o espanhol, o italiano e o francês. De antemão, já considerava que as diferenças que separam as quatro línguas românicas talvez fossem maiores do que as que foram constatadas entre as línguas escandinavas.

---

<sup>14</sup> Heloísa Cintrão foi a colaboradora de Tagnin (2013) para o quadro comparativo do par linguístico espanhol-português.

O experimento, conforme relatado por Jensen (2013), foi realizado com estudantes universitários e percebeu-se rapidamente que funcionava. Um estudante formado numa dessas quatro línguas românicas conseguia, através de uma aprendizagem apropriada, compreender relativamente depressa uma das outras três. Um trimestre bastava, em geral, para que os universitários participantes do experimento conseguissem ler enunciados escritos simples em uma das outras três línguas.

Jensen (2013) e sua equipe começaram a construir uma nova pedagogia que se constitui em expor as correspondências em sincronia, a partir do uso atual e sem exigir dos estudantes um estudo prévio do latim. Existem inúmeras correspondências entre as quatro línguas românicas dizendo respeito não só à fonética, mas também à morfologia, à sintaxe e à organização do léxico. A análise destes domínios constitui uma gramática comparativa resultante da participação de representantes das instituições Universidade de Aarhus, Universidade de Copenhague, Universidade de Valladolid, Universidade de Rouen, Universidade Complutense de Madrid, Universidade Franche-Comté, Universidade de Florença, Universidade de Paris IV-Sorbonne, Universidade do Porto, Universidade Federal do Rio de Janeiro e Universidade do Estado do Rio de Janeiro.

Quadro 5 - Correspondências entre línguas românicas

Português	Espanhol	Italiano	Francês
O João é simpático, o que não se pode dizer da Maria.	Juan es simpático, lo cual no se puede decir de María.	Giovanni è simpático, il che non si può dire di Maria.	Jean est sympathique, ce qu'on ne peut pas dire de Marie.
Apreciei-o por suas qualidades.	Le he apreciado por sus cualidades.	L'ho apprezzato per le sue qualità.	Je l'ai apprécié pour ses qualités.
Conheço o seu amigo.	Conozco a su amigo.	Conosco il suo amico.	Je connais son ami.
É possível que João venha.	Es posible que venga Juan.	È possibile che venga Giovanni.	Il est possible que Jean vienne.
Prefiro não encontrá-lo.	Prefiero no encontrarlo	Preferisco non incontrarlo.	Je préfère ne pas le rencontrer.
O melhor de todos.	El mejor de todos.	Il migliore de tutti.	Le meilleur de tous.
Chamar alguém de ladrão	Tachar a uno de ladrón.	Dare del ladro a qualcuno.	Traiter quelqu'un de voleur.

Fonte: BRITO et al. (2013, pp. 141, 206, 229, 231)

Tagnin (2013) apresentou a convencionalidade, inclusive entre o par linguístico espanhol-português. Jensen(2013) e sua equipe apresentaram a intercompreensão entre línguas românicas. À concepção de língua como estrutura, segundo Koch (2018), corresponde a de

sujeito determinado. A autora ressalta o princípio explicativo de que todo e qualquer fenômeno e todo e qualquer comportamento individual baseia-se sobre o sistema linguístico.

O próprio conceito de texto, conforme Koch (2018) analisa, dependeria das concepções que se tenha de língua e de sujeito. Na concepção de língua como representação do pensamento e de sujeito como senhor absoluto de suas ações de seu dizer, o texto é visto como um produto do autor, cabendo ao leitor/ouvinte senão captar as intenções do produtor, bastando o conhecimento da língua. Juntos, língua, sujeito e texto, constituem uma atividade interativa altamente complexa de produção de sentidos que se realiza com base no sistemas linguístico.

O *sentido* de um texto é, portanto, *construído* na interação texto-sujeitos (...). Também a *coerência* deixa de ser vista como mera propriedade ou qualidade do texto, passando a dizer respeito ao modo como os elementos presentes na superfície textual, aliados a todos os elementos do contexto sociocognitivo mobilizados na interlocução, vêm a constituir, em virtude de uma construção dos interlocutores, uma configuração veiculadora de sentidos (KOCH, 2018, posição 148 [recurso eletrônico]).

Dascal (2003) apresentou seu modelo de idiossincrático de interpretação e compreensão enfocando o que alguém está fazendo quando está buscando significado. Segundo o autor, o significado é um produto de uma interação de forças que sustentam e determinam a atividade humana. Para o autor, o computador foi inicialmente concebido como uma máquina que realiza cálculos. Na fala cotidiana, calcular é realizar certas manipulações numéricas. Porém, não é um problema trivial generalizar intuições e dizer exatamente quais operações (quais funções) em números são cálculos. À solução para esse problema, Dascal (2003) chamou de classe de funções recursivas — um conjunto de procedimentos claramente definidos e que correspondem tanto às intuições quanto ao que é um cálculo. Um cálculo pode ser definido como um procedimento que pode ser realizado por tal conjunto de regras. Essa maneira de chegar à computação foi utilizada por Chomsky (1957).

Para Dascal (2003), o computador foi desenvolvido com base em uma teoria que tentava capturar uma capacidade cognitiva humana, a capacidade de computar. Segundo o autor, as habilidades cognitivas superiores estão intimamente relacionadas à capacidade humana de se comunicar, fazendo induções e deduções em termos linguísticos. Portanto, para que o computador fosse capaz de realizar essas tarefas, ele precisaria primeiro de ser ensinado a lidar com as línguas ou com um sistema simbólico tão expressivo e poderoso quanto um

sistema linguístico. Para ensinar linguística ao computador, conforme Dascal concluiu, seria preciso destacar três componentes: sintático, semântica e pragmático.

Em sua explicação, Dascal (2003) destacou que o primeiro componente consistiria da habilidade de reconhecer expressões bem formadas na língua, de segmentar tais expressões de acordo com sua estrutura gramatical e de produzir expressões gramaticalmente corretas. O segundo e o terceiro componentes, teriam a ver com o significado das expressões linguísticas, não apenas com sua forma. O conhecimento semântico seria o conhecimento do que as expressões da língua significam, pura e simplesmente: o que é significado por várias palavras e qual é o significado das expressões complexas que podem ser feitas a partir dessas palavras. Na formação de expressões linguísticas complexas a partir de outras mais simples, a sintaxe entraria em jogo, portanto, o conhecimento semântico dependeria do conhecimento sintático.

O conhecimento pragmático, para Dascal (2003) também teria a ver com significado, consistindo na habilidade de usar expressões linguísticas significativas nas muitas maneiras possíveis, conforme intenção de comunicação, se informar, assustar, prometer ou alertar. Todas essas maneiras requerem algum conhecimento para serem realizadas — conhecimento que depende de se saber do significado das expressões linguísticas. Seguindo nessa tese, Dascal (2003) sugere que a reprodução da habilidade pragmática no computador seria bem-sucedida. Em sua conclusão, o autor afirma que um processo meramente computacional executado por um computador não poderia ter qualquer aspecto semântico. Apenas haveria soluções de contorno, de modo que o processamento semântico se baseie em expressões interpretadas como meras entidades sintáticas. Portanto, poder-se-ia dizer que um processo puramente sintático no computador simularia a capacidade semântica humana. Dessa forma, poder-se-ia deixar de lado as questões sobre significado e pensamento e lidar com a semântica indiretamente.

Segundo Dascal (2003), na tradução a questão é semântica, mas também tem uma contrapartida puramente sintática, na qual um programa de computador poderia ser escrito para imitar a tradução correta de acordo com as intuições semânticas humanas. A semântica estaria confinada ao significado conforme ocorre no cotidiano. Dascal (2003) então propõe três modelos de interpretação. No modelo “criptográfico”, o significado está sob a superfície do signo, e o intérprete tenta descobrir o significado subjacente. No modelo “hermenêutico”, o significado não está “lá”, mas “aqui”; é o intérprete que dá sentido ao signo relacionando-se

com ele enquanto intérprete. Porém, por razões diferentes, ambos os modelos negligenciam ou minimizam o papel do produtor do signo.

Então Dascal (2003) propõe mais um modelo, o modelo pragmático. Nesse modelo, o significado não é um dado independente nem um construto de intérprete, mas advém de uma ação comunicativa. O modelo pragmático compartilha com o modelo criptográfico a suposição de que existem significados objetivos, associados a seus signos em virtude de regras semânticas, porém dentro do contexto. No entanto, do reconhecimento do papel essencial do contexto na interpretação, Dascal (2003) sugere o modelo superpragmático que consiste em negar à semântica qualquer papel significativo na interpretação. Nesse modelo, o intérprete de alguma forma apreende o significado do falante diretamente com base nas informações contextuais, sem ter que levar em consideração o significado semântico da expressão do falante. Desse modo, recursos de Inteligência Artificial poderiam lidar com o processamento de sistemas linguísticos.

Na próxima seção, serão apresentadas técnicas de Inteligência Artificial.

## **2.4 Reconhecimento de Padrões e Tecnologias da Inteligência**

Knuth (1980), ao propor a discussão sobre o papel real da noção de algoritmo em ciências matemáticas, entregou uma definição de algoritmo adotada neste trabalho.

Nem todos os meus colegas concordam comigo, mas acontece que a fonte de nossa discordância é simplesmente que minha definição de algoritmos é muito mais ampla que a deles: costumo pensar em algoritmos como abrangendo a variedade de conceitos que lidam com processos bem definidos, incluindo a estrutura dos dados que estão sendo acionados, bem como a estrutura da sequência de operações que estão sendo executadas; outras pessoas pensam nos algoritmos meramente como métodos diversos para a solução de problemas particulares, análogos aos teoremas individuais da matemática (Knuth, 1980, p. 1).

Fazendo uma trajetória de suas próprias experiências como estudante, Knuth (1980) abordou um marco importante para a linguística. Ele relembrou que começou a estudar matemática no mesmo ano em que começou a trabalhar com computadores (1957). Nunca tinha misturado seu pensamento matemático com seu pensamento de ciência da computação até 1961, quando ficou empolgado com a ideia de que a matemática e a ciência da computação poderiam ter algum ponto em comum, porque a notação BNF (Backus-Naur Form) parecia matemática, então comprou uma cópia de *Syntactic Structures* de Chomsky e "partiu para encontrar um

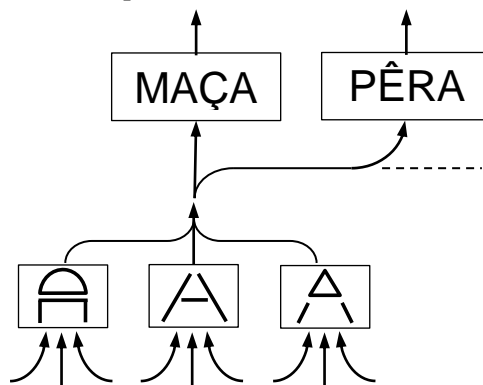
algoritmo para decidir o problema de ambiguidade das gramáticas livres de contexto" (KNUTH, 1980, p. 7).

Chomsky (1957) reformulou os objetivos da teoria linguística e elaborou um modelo que resolvesse o problema de desenvolver e esclarecer os critérios para a seleção correta da gramática para cada língua, ou seja, a teoria correta de cada língua. Tal algoritmo supõe uma gramática da forma normal  $G [\Sigma, F]$ , onde  $\Sigma$  é o corpus finito de enunciados e  $F$  é o conjunto regras gramaticais na forma  $X \rightarrow Y$ , interpretada: “reescreva  $X$  como  $Y$ ” (CHOMSKY, 1957, p. 26, 27 e 51).

Segundo Kurzweil (2015), os seres humanos têm uma capacidade reduzida para processar lógica, mas uma capacidade muito profunda e central para reconhecer padrões e pensar de maneira lógica basicamente é reconhecer padrões. O autor informa que a estimativa do número de padrões com que a mente humana é capaz de lidar chega a trezentos milhões de processadores de padrões, o que permitiu ao Homo Sapiens o desenvolvimento do processo de comunicação.

A teoria da mente baseada em reconhecimento de padrões apresentada por Kurzweil (2015) leva em consideração que há padrões de nível inferior (o padrão principal) e padrões de nível superior. As descrições para cada um desses padrões de nível inferior não precisam ser repetidas para cada padrão de nível superior que se referencia. Por exemplo, no mundo da linguagem, os padrões de nível superior são as palavras, os padrões de nível inferior são as letras. A palavra “maça” inclui duas vezes a letra “A”. A descrição da letra “A” não precisa ser repetida, mas usará a mesma descrição.

Figura 3 - Três padrões redundantes para “A”



Fonte: Kurzweil, 2015, posição 688 [versão eletrônica]

Esse território da recursividade, que busca imitar a mente humana, faz parte da Inteligência Artificial (I.A.). Parafraseando Rich (1988), a I.A. “é o estudo de como fazer os computadores realizarem tarefas em que as pessoas são melhores” (RICH, 1988, p. 1). Nos primeiros vinte anos de pesquisa de I.A., um dos poucos resultados concretos obtidos é que “inteligência requer conhecimento” (RICH, 1988, p.6). Rich ressalta que o conhecimento possui propriedades: é volumoso, é difícil de caracterizar com precisão e está em constante mutação. Tais propriedades são imanentes à competência linguística humana e “uma técnica de I.A. é um método de explorar o conhecimento” (RICH, 1988, p. 6). Na definição de Rich (1988), fica estabelecido que o conhecimento deve ser representado de tal forma que possa:

- capturar generalizações, ou seja, não é necessário que ele represente separadamente cada situação individual. Ao contrário, situações que partilhem propriedades importantes deverão ser agrupadas;
- ser compreendido por pessoas que necessitam supri-lo;
- ser facilmente modificado para corrigir erros e para refletir mudanças no mundo e em nossa visão de mundo; e
- ser utilizado para superar seu próprio volume.

Rich (1988) também ressaltou que, embora as técnicas de I.A. sejam projetadas tendo em vista as restrições impostas pelos problemas de I.A., há certo grau de independência entre os problemas e as suas técnicas de resolução. Tanto é possível resolver problemas de I.A. sem utilizar as técnicas de I.A., como é possível aplicar técnicas de I.A. à solução de problemas que não sejam de I.A. e quase tudo que se denomina I.A. parte do reconhecimento de padrões. Como um princípio, a autora observou que “à medida que a pesquisa de I.A. progrediu e técnicas para tratar de quantidades maiores de conhecimento do mundo foram desenvolvidas, conseguiu-se certo progresso nas tarefas” (RICH, 1988, p. 2).

Entre as tarefas citadas por Rich (1988), estão “a percepção (visão e fala), compreensão de línguas e resolução de problemas em domínios especializados, como a diagnose médica e a análise química” (RICH, 1988, p. 2). Especialmente a partir de 2016, podemos observar a dinâmica de evolução da I.A. que Rich apresentou em 1983, ano de publicação da primeira versão de Inteligência Artificial. A edição 364 da revista Superinteressante, publicada em agosto de 2016, anunciou em manchete de capa “a nova aposta do Google”.

O artigo tratava de narrar cenário futurista em que um enorme sistema de inteligência artificial, alimentando com todas as informações do mundo, seria capaz de nos assistir no

controle e execução de ações cotidianas graças a tecnologias de I.A. já existentes e em funcionamento, 2016 foi o ano em que I.A. se consolidou como produto, inclusive em uso por ferramentas de assistência à tradução. Naquele ano o Google anunciou uma rede neural para tradução automática: a *Google Neural Machine Translation* (GNMT), resultado da estratégia conhecida como “aprendizado de máquina” (*machine learning*).

Em lugar de se tentar ensinar o computador, o novo método consiste em fazê-lo aprender sozinho, usando as chamadas redes neurais que mapeiam o conhecimento de mundo em estrutura de dados que facilite não só a busca de informação, como também a inferência. A GNMT foi iniciada com o mapeamento de sentenças em uma língua para que as mesmas sentenças fossem produzidas em outra língua, considerando que uma sentença é uma unidade de tradução. Para isso, um sistema de tradução baseado em frases fosse construído, previamente.

Lévy (2010) propôs o uso do termo *hipertexto* como uma metáfora válida para todas as esferas da realidade em que significações estejam em jogo e caracterizou o modelo através de seis princípios abstratos em árvores e grafos, estruturas de dados populares para representar o conhecimento:

- Princípio de metamorfose - a rede hipertextual está em constante construção e renegociação.
- Princípio de heterogeneidade - os nós e as conexões de uma rede hipertextual são heterogêneos.
- Princípio de multiplicidade e de encaixe das escalas - o hipertexto se organiza como um fractal, isto é, um nó ou conexão, quando analisado, pode revelar-se uma rede e assim por diante, indefinidamente.
- Princípio de exterioridade - a rede não possui unidade orgânica, nem motor interno. Seu crescimento e sua diminuição, sua composição e sua recomposição permanente dependem da adição de novos elementos.
- Princípio de topologia - nos hipertextos, tudo funciona por proximidade, por vizinhança e o curso dos acontecimentos é uma questão de topologia,
- Princípio de mobilidade dos centros - a rede não tem centro, mas diversos centros, saltando de um nó a outro.

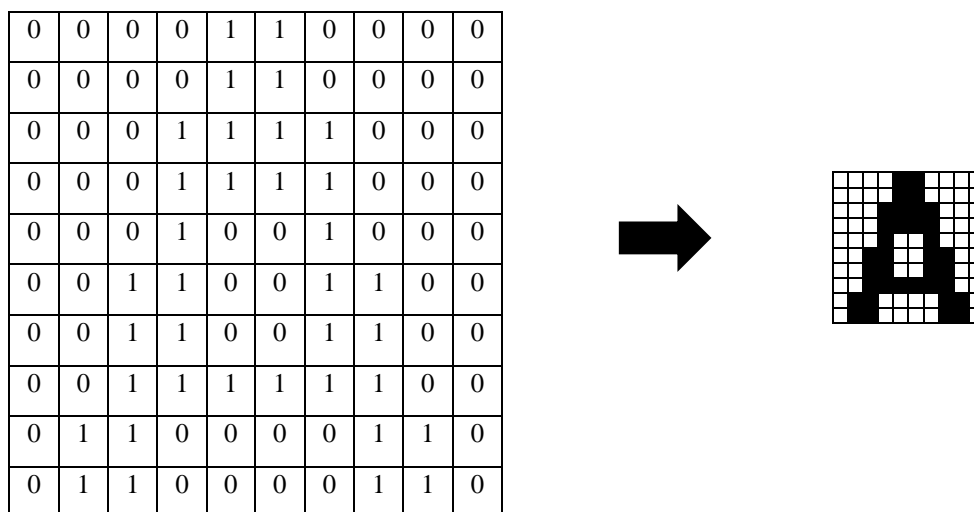
Tais princípios se inspiram na mente humana que funciona através de associações, há um processo que embasa o exercício da inteligência. “Ela pula de uma representação para outra



ao longo de uma rede intrincada, desenha trilhas que se bifurcam, tece uma trama” (LÉVY, 2010, p.28). O reconhecimento de padrões seria uma classificação de dados com base no conhecimento já adquirido e, de um ponto de vista computacional, de estatísticas extraídas de padrões e/ou suas representações. A frequência de ocorrência de palavras-chave no texto, por exemplo, faria parte da representação. No reconhecimento estatístico de padrões, seria usado espaços vetoriais para representar padrões e classes. As abstrações normalmente lidariam com densidade/distribuição de probabilidade de pontos em espaços multidimensionais.

Rich (1988) apresenta, entre exemplos de algoritmo de reconhecimento de padrões, o problema de identificar corretamente padrões de letras do alfabeto, de modo que o sinal analógico produzido pela caneta seja digitalizado. O tamanho das unidades a serem utilizados na digitalização está condicionado ao tamanho das próprias letras; mas, em qualquer caso, a entrada para o programa será uma matriz de 0 e 1 com 1 representando tinta e 0 representando falta de tinta. A figura 3 apresenta um exemplo dessa entrada para visualização do padrão. Os programas de reconhecimento de caracteres que descreveremos darão saída, para matrizes de entradas que receberem, indicando a letra que mais se assemelha ao padrão de entrada.

Figura 4 - Entrada para Programa de Reconhecimento de Letras



Fonte: RICH, 1988, p. 14

Estrutura de Dados:

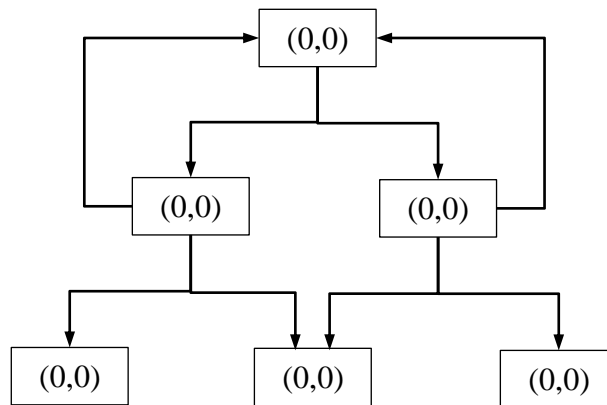
Padrão de entrada      Uma matriz de 0 e 1, conforme descrito acima, representando o padrão a ser classificado.



- seja utilizado em muitas situações, mesmo quando não totalmente preciso e completo;
- seja utilizado para superar seu próprio volume, ao ajudar a estreitar a faixa de possibilidades que deverão normalmente ser consideradas.

Conforme Rich (1988) destacou, o conhecimento precisa estar estruturado de tal forma que decidir qual caminho seguir, ou qual regra aplicar, durante o processo de busca por uma solução, produza um estado-meta. A estrutura deve causar movimento e ser sistemática como *busca em amplitude*. Cada nó da árvore é expandido pelas regras de produção para um conjunto de nós sucessores, cada um dos quais pode, por sua vez, ser expandido, continuando até que um nó representando uma solução seja encontrado. Um procedimento de busca de árvores pode ser convertido em um procedimento de busca de grafo modificando-se a ação realizada cada vez que um nó for gerado.

Figura 5 - Dois níveis de uma Árvore de Busca em amplitude



Fonte: RICH, 1988, p. 70

Segundo Lévy (2010), um procedimento de busca em árvores pode ser convertido em um procedimento de busca de grafo modificando-se a ação realizada cada vez que um nó for gerado. Em vez de simplesmente acrescentar o nó no grafo, é interessante:

- Examinar o conjunto de nós que foram criados até o momento e verificar se o novo nó já existe.
- Se não existir, simplesmente acrescente-o ao grafo da mesma forma que para a árvore.
- Se já existir, fixar o nó que estiver sendo expandido para apontar para o nó já existente correspondente ao seu sucessor.

Conforme conclui Lévy (2010), com um conjunto de neurônios hipotéticos, estruturados como grafos ou árvores, poder-se-ia ordená-los de tal modo a formar conceitos de modo que

máquina de estados finitos probabilística seja construída por meio de um algoritmo da busca pela melhor escolha (algoritmo A\*, pronuncia-se “A estrela”), apresentado abaixo. Uma máquina de estados finitos probabilística é uma máquina de estados finitos onde a função do próximo estado é uma distribuição de probabilidade sobre o conjunto completo de estados da máquina.

Quadro 6 - Algoritmo A\*

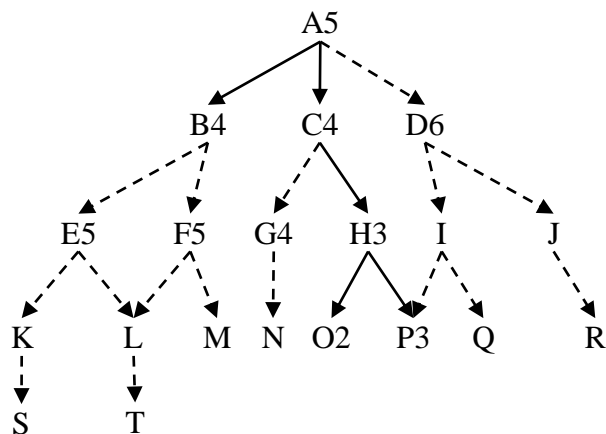
**Algoritmo A\***  
 função busca-melhor-escolha  
 início  
   abertos = [inicial]  
   fechados = [ ]  
   enquanto abertos ≠ [ ] faça  
     início  
       retire o estado mais à esquerda de abertos e chame-o de X  
       se X = objetivo então retorna o caminho de inicial até X  
       senão início  
         gera filhos de X.  
         para cada filho de X faça  
           caso  
             o filho não está em abertos ou em fechados:  
               início  
                 atribua ao filho um valor heurístico  
                 acrescente o filho a abertos  
               fim  
             o filho já está em abertos:  
               se o filho foi alcançado, por um caminho mais curto  
               então dê ao estado em abertos o caminho mais curto  
             o filho está em fechados:  
               se o filho foi alcançado por um caminho mais curto, então  
               início  
                 retire o estado de fechados  
                 acrescente o filho em abertos  
               fim  
             fim  
             coloque X em fechados  
             reordene estados em aberto pelo mérito heurístico (melhor mais à esquerda)  
           fim  
     fim  
   retorna FALHA  
 fim

1. abertos = [A5]; fechados = [ ]
2. avaliar A5; abertos = [B4, C4, D6]; fechados = [A5]
3. avaliar B4; abertos = [C4, E5, F5, D6]; fechados = [B4, A5]
4. avaliar C4; abertos = [H3, G4, E5, F5, D6]; fechados = [H3, C4, B4, A5]
5. avaliar H3; abertos = [O2, P3, G4, E5, F5, D6]; fechados = [O2, H3, C4, B4, A5]
6. avaliar P3; a solução foi encontrada!

Considere a função de avaliação  $f(n) = g(n) + h(n)$ , onde:

- $n$  é um estado qualquer encontrado na busca.
- $g(n)$  é o custo de  $n$  a partir do estado inicial.
- $h(n)$  é a estimativa heurística do custo de ir de  $n$  até um objetivo.
- Se essa função de avaliação for usada como algoritmo busca-melhor-escolha, o resultado é o chamado algoritmo A.
- Um algoritmo de busca é admirável se, para qualquer grafo, ele sempre terminar no caminho de solução ótima.
- Se o algoritmo A for usado com uma função de avaliação, na qual  $h(n)$  é menor que o custo do caminho mínimo de  $n$  para um objetivo, ou igual a ele, o algoritmo de busca resultante será chamado de algoritmo A\*.

Figura 6 - Algoritmo A\*



Fonte: LÉVY, 2010, p. 123

Riesen (2015) argumenta que a capacidade de reconhecer padrões tem sido essencial à sobrevivência e essa capacidade existe graças à sofisticação neural do sistema cognitivo humano. Segundo o autor, a Ciência da Computação adotou o reconhecimento de padrão como um método para imitar a capacidade humana de percepção e inteligência. Em outras palavras, segundo Riesen (2015), reconhecimento de padrão visa definir algoritmos que automatizam, ou ao menos apoiam, o processo de reconhecimento de padrões provenientes do mundo real.

Nesses algoritmos de reconhecimento de padrões, os grafos são comumente usados porque consistem em conjuntos finitos de nós que permitem representar diversas relações aos pares.

Consequentemente, os grafos são capazes não apenas de descrever as propriedades de um padrão, mas também em relacionamentos (binários) entre as diferentes partes do padrão subjacente, ou arranjo de padrões. Além disso, os grafos não são limitados a um tamanho fixo, ou seja, o número de nós e arestas não é limitado a priori, mas pode ser adaptado ao tamanho e à complexidade de cada padrão individual em consideração (RIESEN, 2015, posição 390 [recurso eletrônico]).

Nos próximos capítulos o *Grinder* será apresentado.

### 3 METODOLOGIA

Assim como modelos matemáticos, a construção de algoritmos parte de princípios ideais e abstração para a criação de modelos. Todo o referencial teórico é aplicado aqui, à medida que se intenciona:

1. Simular a mente humana que tem o potencial de processar, analisar e sistematizar coisas em sistemas linguísticos, imitando sua recursividade. A estrutura de dados para representar essa capacidade humana será grafo.
2. Criar modelos linguísticos (do inglês *language model*) a partir de recursos da Linguística de Corpus aplicados por Sinclair (2004) e Tagnin (2013), isto é, a partir de grandes volumes de sentenças extraídos de um conjunto de textos de mesmo gênero textual, encontrar associações convencionais entre as palavras e agrupá-las. É o mapeamento de unidades de significado na acepção de Sinclair (1991)
3. Criar modelos de tradução (do inglês *translation model*) fazendo comparação entre modelos linguísticos, textos originais e suas traduções. Podemos dizer que ocorre aqui uma “quadriangulação”, pois deve haver um modelo linguístico da língua-fonte para comparação com o original (texto-fonte) e um modelo linguístico da língua-alvo para comparação com a tradução (texto-alvo).
4. Criar alinhamento com sistemática de cores de modo que possa ser observado o padrão identificado no modelo linguístico, o padrão identificado no modelo linguístico (em verde), o padrão identificado no próprio texto (em azul) e palavras que não constituíram padrão são identificadas em preto.

Esse processamento textual se embasa no modelo superpragmático de Dascal (2003) e na intercompreensão de Jensen (2013), uma vez que a inteligência aqui construída não realizará qualquer ação de interpretação de significado, nem alinhamento de unidades de tradução entre pares linguísticos. O desenvolvimento do *Grinder* segue a hipótese de que haverá alinhamento natural entre as unidades de tradução se houver um padrão de composição de unidades, algo próximo da proposta de Toury (1995): o emparelhamento de unidades de tradução de tal modo a não restar sobras, *no leftovers*.

Os Estudos da Tradução têm se debruçado sobre o ato de traduzir e como a experiência do tradutor conduz para o aperfeiçoamento de sua memória de trabalho, como Alves (2005) constatou: com a experiência o tradutor armazena unidades de tradução maiores. O *Grinder* não deve se preocupar com o tamanho de unidades de tradução, mas delimita as unidades pela repetição de uso, o foco é no reconhecimento de padrão. Ademais, é considerado neste trabalho que as decisões realizadas pelo tradutor, brevemente contempladas no referencial teórico, sobretudo a equivalência, constituem fonte de conhecimento da qual os algoritmos *Grinder* devem ser capaz de reconhecer padrões no momento em que ocorre o processamento textual na tradução e, assim, mapear as unidades de tradução sistematicamente, ocorrendo o emparelhamento *no leftovers* proposto por Toury (1995).

Em resumo, grafos são capazes de armazenar *corpus*, mantendo as funcionalidades básicas de ferramentas de análise de corpus do tipo listador: calcular o número total de palavras no *corpus* (*tokens*) e representar a ocorrência de cada palavra (*type*) (cf. Laviosa, 2010). Para o processamento textual, são aplicados seis algoritmos: (1) *sentences*, (2) *nodes*, (3) *reduce*, (4) *normalize*, (5) *ordenize* e (6) *filtrate* a serem apresentados no capítulo 4. Juntos, esses algoritmos realizam as tarefas de segmentar o texto, organizá-lo em grafo, reduzi-lo conforme associações identificadas conforme frequência de repetição e, por fim, normalizar o grafo em unidades de significado, na acepção de Sinclair (2004). Tais unidades comparadas em textos paralelos tornam-se unidades de tradução, em consonância ao legado de Laffling (1991) e Toury (1995).

E como os algoritmos pretendidos aplicarão teoria estatística da probabilidade se encaixa bem a metodologia de Linguística de Corpus. Este trabalho emprega corpora monolíngues comparável de textos não traduzidos das Constituições de países de língua oficial portuguesa e Constituições de países hispano-americanos. Para exame das características dos corpora bilíngues paralelos, empregamos a Constituição Brasileira e sua tradução. Para verificar benefícios da prática de textologia contrativa, empregamos textos de Rui Barbosa e Gabriel Garcia Márquez, ambos referente a análises de artigos de Constituições, respectivamente do Brasil de da Colômbia. As Constituições foram copiadas a partir do sítio oficial de cada Governo. A soma do corpus em espanhol é 16.436 *types* e 555.935 *tokens* e a soma do corpus em português é 9.307 *types* e 211.034 *tokens*. Para tal contagem de frequência, foi utilizado o software AntConc 3.4.4w ([http://www.antlab.sci.waseda.ac.jp/antconc\\_index.html](http://www.antlab.sci.waseda.ac.jp/antconc_index.html)).



Tais corpora foram divididos em seis grupos:

- O corpus da Constituição Brasileira, que apresenta 5.796 *types* e 53.500 *tokens*.
- O corpus da tradução ao espanhol da Constituição Brasileira, que apresenta 5.070 *types* e 63.076 *tokens*.
- O corpus do espanhol hispano-americano, que apresenta 15.608 *types* e 491.968 *tokens*, composto pelas Constituições dos países Argentina, Bolívia, Chile, Colômbia, Costa Rica, Cuba, Equador, El Salvador, Guatemala, Honduras, México, Nicarágua, Panamá, Peru, Porto Rico, República Dominicana, Uruguai e Venezuela.
- O corpus de países de língua oficial português, exceto o Brasil, que apresenta 6.450 *types* e 157.534 *tokens*, composto pelas Constituições dos países Angola, Cabo Verde, Guiné Bissau, Macau, Moçambique, Portugal, São Tomé Príncipe e Timor Leste. Cabe ressaltar que as diferenças ortográficas entre os países de língua oficial portuguesa são poucas, graças ao acordo ortográfico de 1990.
- O corpus do português é o texto *O Artigo 6º da Constituição e A intervenção de 1920 na Bahia*<sup>16</sup> de Rui Barbosa sobre sua análise acerca do art. 6.º da Constituição Brasileira.
- O corpus do espanhol é o texto *Hay que manosear la constitución*<sup>17</sup>, resultante da entrevista concedida por Gabriel García Márquez a Pizano para *El Tiempo*. Durante a entrevista, o escritor emitiu sua análise a vários artigos da Constituição da Colômbia.

A principal motivação para a escolha das Constituições é que existe uma relação em construção entre Textos Jurídicos e Inteligência Artificial. Em rápida pesquisa realizada no Google, com contexto brasileiro, diversas notícias serão apresentadas sobre agilidade com uso de inteligência artificial no sistema judiciário e quantidade de projetos nos tribunais brasileiros. Do ponto de vista linguístico, podemos compreender o fenômeno jurídico graças à convencionalidade no conjunto de textos que compõem gêneros textuais no âmbito jurídico. São contratos, leis, regulamentos, estatutos, certidão de batismo, certidão de casamento,

---

<sup>16</sup> BARBOSA, Rui. *O Artigo 6º da Constituição e A intervenção de 1920 na Bahia* in Obras Completas de Rui Barbosa. Vol. XLVII. 1920 Tomo III. Rio de Janeiro: Fundação Casa de Rui Barbosa, 1975. Disponível em: [http://www.stf.jus.br/bibliotecadigital/RuiBarbosa/20144\\_V47\\_T3.pdf](http://www.stf.jus.br/bibliotecadigital/RuiBarbosa/20144_V47_T3.pdf). Acessado em: 5 out. 2020.

<sup>17</sup> PIZANO, Daniel S. *Hay que manosear la constitución*, El Tiempo, 16 de septiembre de 1990. CROCE, Juan E. C. La nostalgia de las almendras amargas [livro eletrônico]. Bogotá: Intermedio Editores, 2014. Posição 1637.

certidão de óbito, documentos pessoais, autorização e funcionamento, sentença de condenação, medida provisória, editais dentre outros.

Estão embarcados nestes projetos de I.A. do sistema judiciário brasileiro algoritmos de Aprendizado de Máquina (*Machine Learning*), Aprendizado Profundo (*Deep Learning*), Redes Neurais (*Neural Networks*) e Internet das Coisas (*Internet of Things*). Neste trabalho, estamos simplificando os conceitos que estão por trás de todas essas soluções tecnológicas para que eles possam estar ao alcance de quem interessar-se pelo assunto e quiser construir seus próprios algoritmos de inteligência artificial para suas investigações e atividades de tradução.

As etapas seguidas neste trabalho foram:

1. Obtenção de dados – cópia das Constituições Federais a partir dos sítios oficiais dos países supracitados. Todas as Constituições são públicas e de interesse geral, não sendo necessárias autorizações para sua cópia.
2. Conversão dos dados em formato único – as cópias foram salvas com o formato texto plano utf-8 (\*.txt). Não foi utilizado software especial para isso, a não ser ferramentas comuns de escritório como Microsoft Office, leitores de arquivos PDF e bloco de notas.
3. Transformação de dados pré-processamento – retirada dos cabeçalhos, rodapés, numeração de páginas e quebras de linha provocadas pela formatação do documento original em pdf. Para esta etapa é muito útil utilização do software Notepad ++ v7.6.2.
4. Transformação de dados – realizado no *Grinder* (ver capítulo 4):
  - a. verificação e substituição de números e marcas ortográficas que possam provocar quebra de sentença indevidamente, no caso, ponto-final em siglas e marcadores ordinais (.º ou .ª).
  - b. quebra dos textos em parágrafos, sentenças e palavras; e
  - c. vetorização das palavras em grafos.
5. Seleção de características:
  - a. Identificação de binômios, trinômios e expressões fixas.
  - b. Eliminação de palavras que formam única uma única vez no texto e, quando verificado nos corpora, foram utilizados indicadores baseados na sequência de Fibonacci.
  - c. Armazenamento de n-grams.
6. Extração de informação – neste momento ocorre o reconhecimento de padrões.

7. Junção dos textos – também conhecido como *merging*, uma estrutura de dados é salva com as unidades de tradução mapeadas.

As etapas “transformação de dados”, “seleção de características”, “extração de informação” e “junção dos textos” serão apresentadas em esboços de métodos de código<sup>18</sup>, porque o *Grinder* ainda está em construção. E, dividindo o processamento em fases, a primeira fase é obter dados de texto, optamos aqui pelas Constituições, mas textos podem ser coletados de diferentes fontes, como mídia social, páginas da web, memorandos corporativos, cartas, relatórios de diagnóstico, prescrições, artigos de pesquisa científica, entre outros. Para o *Grinder*, no momento, é necessário que os textos estejam no formato utf-8 texto plano (do inglês *plain text*), isto é, formato identificado pela extensão txt.

A partir daí, o *Grinder* está projetado para processar textos de tal modo que o resultado seja unidades de significado na acepção de Sinclair (2004).

---

<sup>18</sup> Em Ciência da Computação um esboço de método é conhecido como *stub*, isto é, um código construído para simular o comportamento de um código existente.

## 4 O GRINDER

### 4.1 OS ALGORITMOS

A tecnologia, ao longo do tempo, tem contribuído para produzir conhecimento aplicado em forma de ferramentas a práticas profissionais e acadêmicas. Já dissemos que a introdução do microcomputador na década de 1980 marcou transição para os Estudos da Tradução. Nesses últimos quarenta anos, além do aumento exponencial de conteúdo a ser traduzido, bem como a ser investigado, novas perspectivas foram incorporadas ao ato de traduzir, graças aos novos recursos de produção e pesquisa de conteúdo. Nesse ínterim, o papel do computador nos Estudos da Tradução mudou totalmente; ninguém pensaria agora em realizar pesquisa sem um computador.

Também já dissemos que, a começar pela década de 1990, Linguística de Corpus em Estudos da Tradução e sistemas de memórias de tradução. O efeito de maior alcance foi nos estudos de corpus, onde o computador tornou possível gravar e processar indefinidamente grandes quantidades de texto, de modo que agora há dados adequados para métodos estatísticos de análise. Isso abre o caminho para a descrição do uso de línguas em termos probabilísticos e reanimou o debate em torno dos conceitos de equivalência e unidades de tradução e do próprio ato de traduzir com as novas possibilidades de investigação e a reutilização de unidades de tradução. O tradutor há muito trabalha em um ambiente mediado por ferramentas que o auxiliam no desempenho de suas tarefas. Os pesquisadores de Estudos da Tradução dispõem de várias ferramentas para disponibilidade e análise de corpora.

Se queremos mapear unidades de tradução sistematicamente, devemos avaliar estratégias para encontrar correspondências. O que ficou claro é que quando a convencionalidade se faz no nível sintático ou pragmático, a identificação de que as unidades de significado estão contidas em uma mesma faixa semântica é mais espontânea. E pensando que estamos buscando estratégias para a construção de algoritmos que mapeiem UTs, é preciso encontrar regras que

possam identificar unidades, cuja convencionalidade se faz no nível semântico de unidades de que há de como contornar problemas de interpretação e compreensão.

Este projeto, o *Grinder*, é um tipo de colaboração que ainda é incomum, onde se une linguística, teorias da tradução e ciência da computação em sinergia. Podemos aprender muito com corpora computadorizados em uma escala que agora temos. Os corpora são um recurso primário para nos permitir teorizar sobre o uso das línguas, como uma série de características que foram investigadas nos estudos de corpus. Aqui associamos tais características a modelos de interpretação: criptográfico, hermenêutico e pragmático e superpragmático. Recapitulando:

- Modelo criptográfico – o sentido está objetivamente no texto;
- Modelo hermenêutico – o sentido é criado pelo intérprete; e
- Modelo pragmático – o sentido está na intenção do produtor do texto; e
- Modelo superpragmático – o sentido está na intenção do produtor do texto e o intérprete capta o sentido diretamente do contexto.

Uma alta proporção do uso da língua é rotineira, convencional e idiomática. É claro que a rotina está conciliada com a criatividade, mas é precisamente no uso de recursos rotineiros, convencionais e idiomáticos que a criatividade é exibida. Não pode haver sentido em se afastar de uma norma da língua, a menos que já haja uma norma em vigor para se afastar. O produtor de um texto estabelece algum tipo de padrão no domínio interpessoal, talvez envolvendo coligações, colocações, binômios, trinômios, estruturas cristalizadas, expressões convencionais, expressões idiomáticas, marcadores conversacionais, fórmulas situacionais, e constrói sobre o padrão conteúdo novo para comunicar a mensagem que deseja.

Há uma longa tradição de estigmatizar padrões que não se conformam aos cânones da língua culta, e que o que precisamos compreender aqui é que ainda assim, existem padrões. O fato é que os níveis linguísticos léxico, sintático e semântico são altamente padronizados no domínio interpessoal e no texto, a depender do gênero textual aplicado, há uma estrutura bem definida. Neste trabalho, estamos avaliando Constituições Federais de diversos países, cujo uso da língua culta é altamente padronizado, tendo uma estrutura comum: preâmbulo, corpo normativo subdividido em títulos, capítulos, seção e subseção, e ato das disposições constitucionais. As seções e subseções estão subdivididas em artigos, caput (caso o artigo tenha subdivisões), parágrafos, incisos, alíneas e itens.

Presumivelmente, algoritmos seriam capazes de usos mecânicos da língua como um truque, realizando o processamento textual em busca de unidades de significado (Sinclair, 1991), reconhecendo padrões de convencionalidade e intercompreensão, em um modelo de interpretação superpragmático. E é claro que tais afirmações não poderiam ser observadas sem o benefício de um corpus. A produção de sentido está no texto, seja o original, seja a tradução, e as unidades de significado são extraídas daí, porque no texto está a mensagem e quem produziu a mensagem fez utilizando normas, quem lê esta mensagem sabe decodificá-la e, no transporte de uma língua a outra, sobretudo de uma mesma família, adiciona-se o fato de que há uma intercompreensão entre línguas próximas.

Na tradução, o problema básico para o tradutor é o problema da escolha, assim como a decisão de um escritor de preferir esta forma de expressão a outra. Mas essas são apenas as ocasiões em que a escolha está, ou pode ser colocada, sob o foco da atenção, como podemos ver nos registros protocolares de reflexão em voz alta dos tradutores são escolhas feitas conscientemente. Na verdade, todo uso da língua é um exercício de escolha e muitas das aplicações de uma teoria da tradução dependem de trazer à tona as escolhas específicas que foram feitas, ou que precisam ser feitas, em determinados contextos situacionais e textuais, em outras palavras, em localizá-los em sua função no geral sistema da língua. Uma forma de atividade na qual o processo, ou processos, de escolha estará mais claramente em primeiro plano é a tradução de uma língua a outra. A função de uma teoria da tradução especificar e explicar os aspectos em que podemos estabelecer equivalência

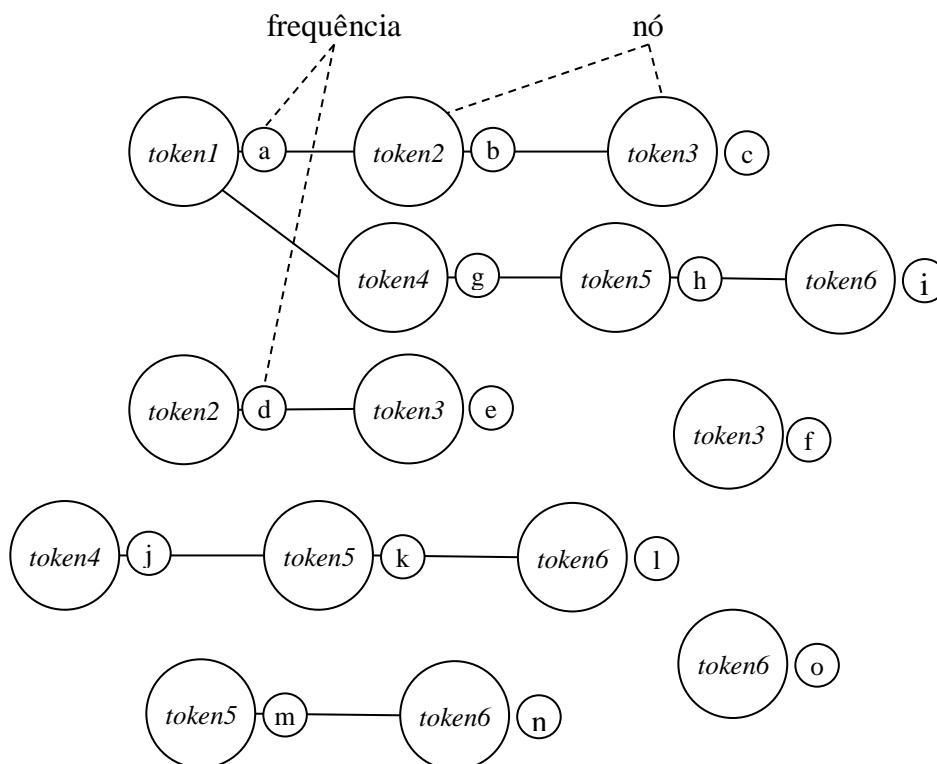
Mas então, o mesmo acontece quando uma teoria que oferece explicações funcionais tanto para o sistema de linguístico quanto para cada instância de uso da língua, e localizando qualquer característica da língua em seu contexto paradigmático, como seleção de um sistema multidimensional recurso de criação de significado. A tradução é uma área em que é valioso poder refletir sobre o que estamos fazendo e explorar o potencial de significado de uma língua, de duas ou mais línguas à medida que entram em contato, de forma explícita e recuperável. Criamos algoritmos de análise, pouco sofisticados, que apenas se apropriam das características linguísticas para reconhecer padrões e identificar possíveis unidades de significado. Se estamos trabalhando com corpora paralelo, então estaremos identificando unidades de tradução. O reconhecimento de características se faz pelo método estatístico.

O *Grinder*, como chamamos nossa ferramenta de mapeamento sistemático de unidades de tradução, se propõe a ser uma ferramenta que facilita a identificação das unidades

convencionais em textos paralelos, não necessariamente original e tradução. Antes da apresentação da ferramenta em si e os resultados preliminares (Capítulo 5), apresentaremos os algoritmos, desenvolvidos com práticas de Inteligência Artificial, que compõem o *Grinder*, inclusive no intuito de que outras pessoas colaborem com o seu desenvolvimento, porque o que mostraremos aqui é só o início.

Considerando que uma sentença não é simplesmente uma sequência de palavras, pois as palavras se associam de maneira convencional para produzir sentido, organizamos todas as sentenças de um corpus em grafos. Cada sentença passa por um processo conhecido no processamento textual como tokenização, no qual as palavras são separadas. Até aí, sem muita novidade, já que ferramentas aplicadas à Linguística de Corpus fazem isso normalmente, levantando o número de *types* e *tokens*. Os *tokens* são organizados na estrutura de grafos, que cada *token* se torna um nó e sua frequência se torna um subnó. Essa frequência é calculada com base na associação entre *tokens* em uma mesma sentença. Novas associações são criadas, sempre mantendo leitura da esquerda para direita.

Figura 7 - Modelo de estrutura de grafos do *Grinder*



Fonte: Elaboração própria.

No exemplo hipotético da figura 7, temos duas sentenças:

[1] Token1 token2 token3.

[2] Token1 token4 token5 token6.

Os valores de frequência para as associações possíveis são: a=2, b=1, c=1, d=1, e=1, f=1, g=1, h=1, i=1, j=1, k=1, l=1, m=1, n=1 e o=1.

Novas associações são criadas para verificação de binômios e trinômios, ou outras fórmulas fixas que em processamento textual são os n-grams. Do exemplo:

token1(2) token2(1) token3(1)  
 token2(1) token3(1)  
 token4(1) token5(1) token(6)  
 token3(1)  
 token4(1)  
 token5(1)  
 token6(1)

Perini (2020) propõe dividir a sentença:

[3] O carro de Luísa é azul.

E obtém as partes:

o carro de Luísa  
 Luísa  
 Luísa é azul  
 de Luísa é

Tal exercício é para que se observe verificar que nem todas as partes parecem satisfatórias e como significado das partes contribui para o significado da sentença:

[4] É azul, o carro de Luísa.

[5] \*Luísa é azul, o carro de.

[6] [O carro de Luísa] [é azul]

[7] \*[O carro de] [Luísa é azul]

Realizando o mesmo exercício proposto por Perini (2020) com o *Grinder*, teremos:

O carro de Luísa é azul  
 carro de Luísa é azul  
 de Luísa é azul  
 Luísa é azul  
 é azul  
 azul

E normalização das partes gera vazio ( $\emptyset$ ), porque o exercício foi feito com a apenas uma sentença. O ideal é que se use o *Grinder* com um corpus composto por coleção de textos



agrupados de acordo com critérios específicos. Continuando o exemplo, incluiremos mais uma sentença do teste de Perini (2020):

[8] Ela disse que Luísa é azul.

Faremos novo processamento com *Grinder* incluindo as sentenças [7] e [8] e então obtemos o resultado:

O carro de Luísa é azul  
 carro de Luísa é azul  
 de Luísa é azul  
 Luísa é azul  
 é azul  
 azul  
 Ela disse que Luísa é azul  
 disse que Luísa é azul  
 que Luísa é azul

Com o mapeamento das unidades de significado *Luísa é azul*, *é azul* e *azul*. Estas unidades foram identificadas graças à frequência:

O(1) carro(1) de(1) Luísa(1) é(1) azul(1)  
 Carro(1) de(1) Luísa(1) é(1) azul(1)  
 De(1) Luísa(1) é(1) azul(1)  
 Luísa(2) é(2) azul(2)  
 é(2) azul(2)  
 azul(2)  
 Ela(1) disse(1) que(1) Luísa(1) é(1) azul(1)  
 Disse(1) que(1) Luísa(1) é(1) azul(1)  
 Que(1) Luísa(1) é(1) azul(1)

Incluiremos mais uma sentença do teste de Perini (2020):

[9] O carro de Luís é branco.

Faremos novo processamento com *Grinder* incluindo as sentenças [7], [8] e [9] e então obtemos o resultado:

O(2) carro(2) de(2) Luísa(1) é(1) azul(1)  
 Luís(1) é(1) branco(1)  
 Carro(2) de(2) Luísa(1) é(1) azul(1)  
 Luís(1) é(1) branco(1)  
 De(2) Luísa(1) é(1) azul(1)  
 Luís(1) é(1) branco(1)  
 Luísa(2) é(2) azul(2)  
 é(3) azul(2)  
 branco(1)  
 azul(2)  
 Ela(1) disse(1) que(1) Luísa(1) é(1) azul(1)

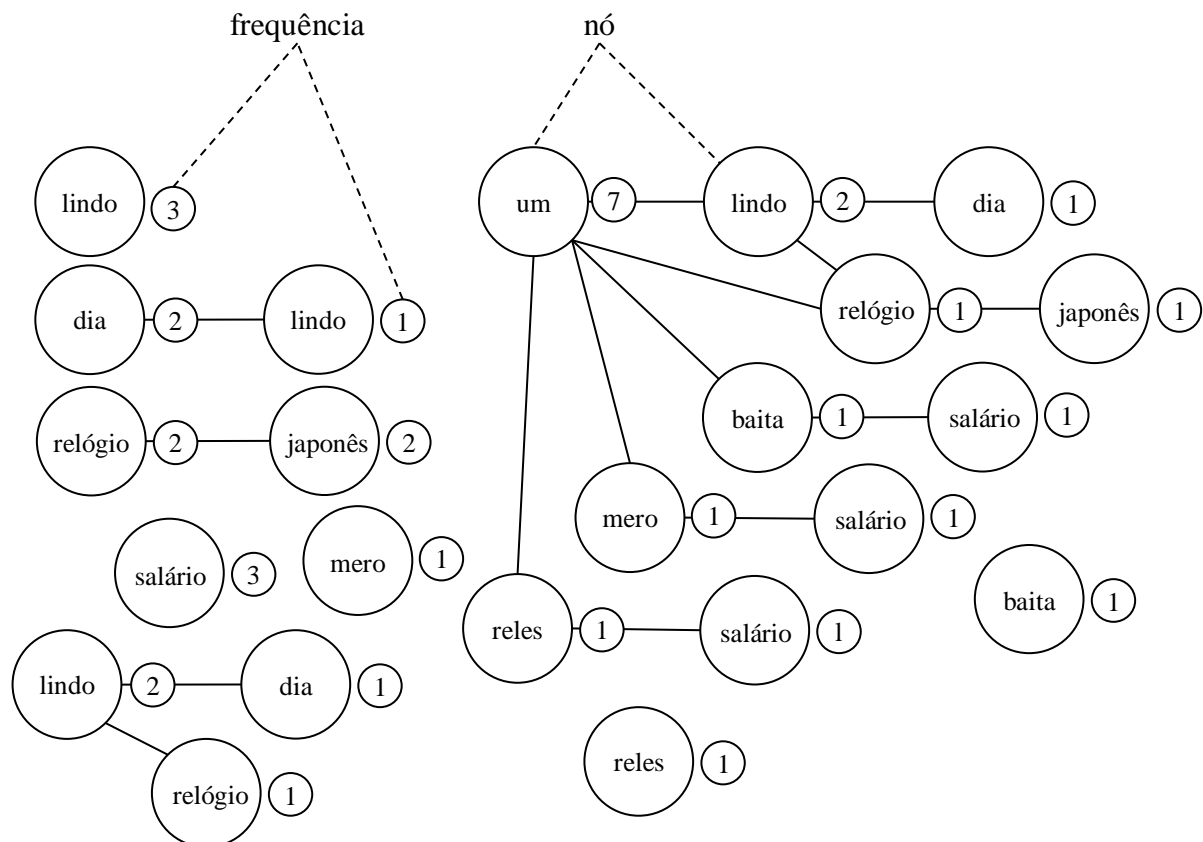
Disse(1) que(1) Luísa(1) é(1) azul(1)  
 Que(1) Luísa(1) é(1) azul(1)

E novo mapeamento de unidades é: *o carro de, carro de, de, Luísa é azul, é azul e azul.*

A preposição *de* está apresentada como unidade, isto ocorre haja vista o baixo número de sentenças. E com mais alguns casos de Perini (2020), ilustramos em grafo (Figura 7).

- [10] Um lindo dia
- [11] Um dia lindo
- [12] Um relógio japonês
- [13] Um lindo relógio japonês
- [14] Um baita salário
- [15] Um mero salário
- [16] Um reles salário

Figura 8 - Estrutura de grafos do *Grinder*



Fonte: Elaboração própria.

Foram unidades mapeadas *um lindo, lindo, dia, relógio japonês, japonês e salário.*

E agora podemos observar uma poderosa ferramenta natural em ação: a simetria. Os matemáticos da Grécia Antiga foram apaixonados por beleza, simetria e geometria. Platão descreveu cinco objetos 3D multifacetados na obra *Timeu* por volta de 350 a.C. Os sólidos platônicos são poliedros cujas faces são polígonos idênticos, com lados de comprimento igual e ângulos de graus iguais, e unidos segundo uma proporção. Platão, Euclides, Leonardo de Pisa, também conhecido como Fibonacci, Lucas Pacioli, Fermat, Descartes ou Kepler, entre tantos outros matemáticos<sup>19</sup>, descreveram, cada um à sua maneira, a proporção que aparece com incrível frequência na natureza, trata-se do número de ouro 1,618 que se manifesta pela harmonia. O conceito de simetria aqui aplicado é o de similaridade ou igualdade para realizar comparações mesmo em cenários de oscilações<sup>20</sup>. Outros métodos estatísticos podem ser testados futuramente, mas o uso dos índices de Fibonacci, habitualmente utilizados em análises de mercado, sobretudo no eixo de tempo do mercado, foi bem sucedido aplicado nas sentenças (nível sintagmático).

Voltando às sentenças [10], [11], [12], [13], [14], [15] e [16] e ao grafo do *Grinder* para estas sentenças (Figura 8), observemos alguns números:

- O corpus é composto por 9 *types* e 22 *tokens*;
- $[(1,618)^3 \times 9] - [(0,618)^3 \times 9] = 36$ , equivalente à soma de todas as frequências;
- $9 - [(0,618)^3 \times 9] \approx 6$ , equivalente ao número de *types* das unidades mapeadas;
- $(0,618)^3 \times 9 = 9^{(0,236)} \approx 2$ , excluindo-se todos os *tokens* cuja frequência é menor que 2, encontraremos as unidades mapeadas;
- o mapeamento de unidades de significado realizado pelo *Grinder* é com base na razão 1,618 e os resultados se mostraram promissores; e
- chegamos a um modelo superpragmático onde tudo o que acontece é o reconhecimento da convencionalidade por meio do método estatístico distribuição de frequência baseada na proporção áurea.

Uma ferramenta de processamento textual pode ser desenvolvida com diferentes linguagens de programação, ou mesmo com base em outras ferramentas. Optamos por escrever todo o código do *Grinder*, do zero, nem mesmo usamos bibliotecas, para que tivéssemos controle

---

<sup>19</sup> 250 marcos cronológicos na história da matemática estão disponíveis em PICKOVER, Clifford A. *The math book*. New York: Sterling Publishing, 2009.

<sup>20</sup> BORODEN, Carolyn. *Fibonacci Trading: how to master time and price advantage*. New York: McGraw-Hill, 2008

total do processamento, de modo a permitir que fossem feitos ajustes finos conforme resultados. Isso nos proporcionou, por exemplo, chegar à potência de três para cálculo das razões em que a simetria é observada. A linguagem de programação que tem sido utilizada para a construção do *Grinder* é o *Python*, mas, neste trabalho, estamos apresentando apenas o esboço dos métodos, justamente porque o *Grinder* ainda é uma ferramenta em desenvolvimento.

#### 4.1.1 Sentences

O algoritmo *Sentences* realiza a segmentação dos textos que compõem os *corpora* que serão comparados para mapeamento das unidades de tradução. Basicamente, este algoritmo divide cada texto em parágrafos e, por conseguinte, os parágrafos em sentenças, considerando, ponto, ponto e vírgula, interrogação, exclamação, travessão, parêntese, colchete e vírgula marcam segmentação do texto em blocos de sentido. A lógica desenvolvida promove distinção entre apostro, vocativo e separação de itens. Por exemplo, a sentença “como se, de repente, quisesse transformar a rua em passarela” é transformada em dois segmentos: (1) “de repente” e (2) “como se quisesse transformar a rua em passarela”.

```
# sentences.py

import json
import re

def pos(mark,line,start):
    return span

def trim(text):
    return string

def sentence(line,dicmarks,lines,temp):
    return lines

thread = []
files = []
rows = []
file = ''

# ler arquivo json de sinais ortograficos
with open(r'lib\marks.json', 'r', encoding='utf-8-sig') as f:
    dicmarks = json.load(f)

# ler arquivos txt do corpus
nfiles = input('Numero de arquivos: ')
for i in range(int(nfiles)):
    file = input(str(i) + ': ')
    files.append(file)

for file in files:
    with open(file,'r',newline='',encoding='utf-8-sig') as f:
```



### 4.1.3 Reduce

O algoritmo *Reduce* este algoritmo verifica as frequências e exclui *tokens* que estão abaixo da estatística calculada. O número de *types* é multiplicado por 0,618 elevados a três, que equivale a 0,236. Em seguida, o valor encontrado é utilizado para realizar a redução das sequências de palavras para reconhecimento do padrão de convencionalidade no texto ou no *corpus* processado.

```
# reduce.py

import json
import math
import re

def removeChild(object,freq,limit):
    return object

# ler resultado do processamento de nodes.py
with open(r'results\nodes.json', 'r', encoding='utf-8') as f:
    corpus = json.load(f)

# outras razões áureas podem ser testadas (0.09, 0.146, 0.236, 0.382, 0.618)
limit = math.pow(len(corpus),0.236)
corpus = removeChild(corpus,[],round(limit))

with open(r'results\reduce.json','w',newline='',encoding='utf-8') as f:
    f.write(json.dumps(corpus,ensure_ascii = False,sort_keys = False,indent=4))
```

### 4.1.4 Normalize

A partir do grafo reduzido, o algoritmo *Normalize* recupera as sentenças que agora constituem unidades de significado.

```
# normalize.py

import json
import re

def getRootNode(node,parent,grandpa,root,textContent):
    return textContent

# ler resultado do processamento de reduce.py
with open(r'results\reduce.json','r',encoding='utf-8') as f:
    nodes = json.load(f)

textContent = getRootNode(nodes','',[],'')

with open(r'results\units.txt','w',newline='',encoding='utf-8') as f:
    f.write(textContent)
```

#### 4.1.5 Ordenize

A partir do resultado dos seis algoritmos-base (*sentences*, *nodes*, *reduce* e *normalize*) e organiza as unidades para localização. Os arquivos gerados por este algoritmo é que são utilizados para o mapeamento de unidades de significado nos textos.

```
# ordenize.py

import json
import re

def trim(text):
    return string

def appendUnit(lexia,lexis):
    return lexis

lexis = {}

# ler resultado do processamento de normalize.py
with open(r'results\units.txt','r',newline='',encoding='utf-8-sig') as f:
    rows = f.read()

units = re.split('\r\n',rows)

for unit in units:
    appendUnit(unit,lexis)

keys = [*lexis]

for key in keys:
    lexis[key].sort(reverse=True)

with open(r'results\units.json','w',newline='',encoding='utf-8') as f:
    f.write(json.dumps(lexis,ensure_ascii = False,sort_keys = False,indent=4))
```

#### 4.1.6 Filtrate

Realiza comparação entre texto e corpus de referência, isto é, modelo linguístico processado previamente e mantido como memória nas configurações do *Grinder*. As unidades de significado são mapeadas em três cores: (1) preto para unidades que foram localizadas apenas no texto/corpus e não foram usadas dentro da frequência mínima ( $0,236 \times \text{types}$ ) no texto/corpus, (2) verde para unidades que foram localizadas no corpus de referência, e azul para unidades localizadas no corpus e acima da frequência mínima. Estas unidades de significado tornam-se unidades de tradução conforme emparelhamento *no leftovers*.

```
# filtrate.py

import json
import re
```

```

def utrim(text):
    return txt

def filtrate(rows,units,units1,unreplaces):
    return lines

# ler resultado do processamento de ordenize.py (modelo linguístico)
with open(r'results\units.json','r',encoding='utf-8') as f:
    units = json.load(f)

# ler resultado do processamento de ordenize.py (modelo de tradução)
with open(r'results\units1.json','r',encoding='utf-8') as f:
    units1 = json.load(f)

# ler arquivo de substituicoes para evitar quebra de linha na segmentacao
with open(r'lib\unreplaces.json','r',encoding='utf-8-sig') as f:
    unreplaces = json.load(f)

# ler resultado do processamento de sentences.py
with open(r'results\master.txt','r',newline='',encoding='utf-8-sig') as f:
    rows = f.readlines()

lines = filtrate(rows,units,units1,unreplaces)
html = '<!DOCTYPE html>\r\n<html>\r\n<head>\r\n<title>Grinder - Translation
Technology</title>\r\n<style rel="stylesheet"
type="text/css">\r\n.tag1{color:
#009961;}\r\n.tag2{color:#2251dd;}\r\n.tag3{color:#ff5233;}\r\nbody{font-
size:medium;}\r\nh1{font-size:large;}\r\nh2{font-
size:medium;}\r\n</style>\r\n</head>\r\n<body>\r\n<p>'
row = ''

for line in lines:
    if line[1] == 0:
        html += line[0].replace('\r\n','<br>')
    elif line[1] == 1:
        html += '<span class="tag1">' + line[0].replace('\r\n','<br>') + '</span>'
    else:
        html += '<span class="tag2">' + line[0].replace('\r\n','<br>') + '</span>'

    if line[2] != '' and line[2] != '\r\n':
        row += line[2] + '<br>'

html += '</p>\r\n<h2 class="tag3">Translation Units</h2>\r\n'
html += '<p>' + row + '</p>\r\n'
html += '</body>\r\n</html>\r\n'

with open(r'results\grinder.html','w',newline='',encoding='utf-8') as f:
    f.writelines(html)

```

O arquivo *unreplaces.json* pode ser substituído no algoritmo pelo dicionário de expressões regulares (*regex*), utilizadas aqui para substituição de números, datas e abreviações como parâmetros (*\_num\_*, *\_ano\_*, *\_mês\_*, *art\_*, *inc\_*, *a\_l*). Também é possível incluir novas expressões regulares, conforme necessidade de substituições para tratamento do texto.



```

unreplaces = {
  "_num_": "(\\d+\\.\\d+|\\d+\\.\\d+|\\d+|a\\. |a\\) |b\\) |c\\. |c\\) |d\\.
|d\\) |e\\. |e\\) |f\\. |f\\) |g\\. |g\\) |h\\. |h\\) |i\\. |i\\)
|j\\. |j\\) |k\\. |k\\) |l\\. |l\\) |m\\. |m\\) |n\\. |n\\) |ñ\\.
|ñ\\) |o\\. |o\\) |p\\. |p\\) |q\\. |q\\) |r\\. |r\\) |s\\. |s\\)
|t\\. |t\\) |u\\. |u\\) |v\\. |v\\) |w\\. |w\\) |x\\. |x\\) |y\\.
|y\\) |z\\. |z\\)
|i|ii|iii|iv|ix|l|li|lii|liii|liv|lix|lv|lvi|lvii|lviii|lx|lxi|lxii|l
xiii|lxiv|lxix|lxv|lxvi|lxvii|lxviii|lxx|lxxi|lxxii|lxxiii|lxxiv|lxxv
|lxxvi|v|vi|vii|viii|x|xi|xii|xiii|xiv|xix|xl|xli|xlii|xliii|xliv|xli
x|xlvi|xlvii|xlviii|xv|xvi|xvii|xviii|xx|xxi|xxii|xxiii|xxiv|xxix
|xxv|xxvi|xxvii|xxviii|xxx|xxxi|xxxii|xxxiii|xxxiv|xxxix|xxv|xxxvi|x
xxvii|xxxviii|cem|cento|cinco|cinqüenta|dezenove|dezesesseis|dezesete|
dezoito|dez|dois|doze|duzentos|duzentas|novecentos|novecentas|noventa
|nove|oitenta|oito|oitocentos|oitocentas|onze|quarenta|quatorze|quatr
o|quatrocentos|quatrocentas|quinhentos|quinhentas|quinze|seis|seiscen
tos|seiscentas|sessenta| sete
|setecentos|setecentas|setenta|três|treze|trezentos|trezentas|trinta|
vinte|primeira|primera|segunda|tercera|cuarta|quinta|sexta|septima|oc
tava|novena|decima|decimoprimeria|decimosegunda|decimotercera|decimocu
arta|decimoquinta|decimosexta|decimoseptima|decimooitava|decimonovena|
vigésima|vigésima
primera|vigésimoprimeria|vigésimosegunda|vigésimotercera|vigésimocuart
a|vigésimoquinta|vigésimosexta|vigésimoséptima|vigésimoitava|vigésimo
octava|vigésimo|vigésimonovena|vigésimo novena|trigésima|trigésima
primera|trigésima segunda|trigésima terceira|trigésima
cuarta|trigésima quinta|trigésima sexta|trigésima séptima|trigésima
octava|trigésima novena|cuadragésima|cuadragésima
primera|cuadragésima segunda|cien)",
  "_mês_": "(janeiro|fevereiro|março|abril|maio|junho|julho|agosto|setembro|outu
bro|novembro|dezembro|enero|febrero|marzo|abril|mayo|junio|julio|sept
iembre|octubre|noviembre|diciembre|jan|fev|mar|abr|mai|may|jun|jul|ag
o|set|sep|out|oct|nov|dez|dec)",
  "_ano_": "(\\d+\\.\\d+|\\d+\\.\\d+|\\d+|\\d+)",
  "art_": "art\\.\"",
  "inc_": "inc\\.\"",
  "a_1": "a\\.1\\.\"",
  "n_": "n\\.\""
}

```

## 4.2 Interface do Usuário

Como já mencionamos, Estudos de Tradução Baseados em Corpus surgiram do casamento entre Linguística de Corpus e Estudos Descritivos de Tradução e objetivam fazer análises estatísticas de textos traduzidos em relação a textos não traduzidos ou seus respectivos textos fonte com base no uso de corpora. O foco na descrição poderá ser nas características da tradução nos níveis léxico, sintático, semântico e pragmático, de modo a desvendar as características dos textos traduzidos, o estilo do tradutor e as normas de tradução.

Geralmente, o que se espera de ferramentas de processamento e análise de corpus são duas funcionalidades: (1) apresentem o número total de palavras corridas, ou *tokens*, e (2) apresentem as ocorrências de cada palavra, ou *type*. O *Grinder* utiliza essas informações, mas não as apresenta ao usuário. Seu objetivo é listar as unidades de significado como unidades de tradução quando o processamento textual é realizado em corpora paralelos. O *Grinder* foi pensando com foco em procedimentos analíticos acerca das unidades de significado voltado a Estudos da Tradução, o que não deixa de permitir aos pesquisadores usarem o *Grinder* em um processo cíclico de observação, descoberta, formação de hipóteses e teste a partir do processamento textual.

Como uma ferramenta de auxilia à tradução (CAT), o *Grinder* permite métodos de investigação baseados em corpus ou corpora paralelo à busca por unidades de tradução ou, ainda, de características da tradução. Como uma ferramenta de Linguística de Corpus, pretende contribuir com Estudos da Tradução, propondo um quadro metodológico em que a convencionalidade é investigada no corpus em modelo superpragmático, são combinados de princípios linguísticos relacionados à tradução (por exemplo, a relação entre textos paralelos de gêneros correspondentes em pares linguísticos — Laffling (1991) usou manifestos de sindicatos da Inglaterra e da Alemanha).

Em particular, a interface do *Grinder* foi construída para realizar a “quadriangulação”. Essa “quadrangulação” é composta de dois tipos principais de triangulação: (1) uma envolvendo a combinação de diferentes corpora, e (2) outra que consiste na combinação de métodos de análise quantitativos e qualitativos. A ênfase é colocada aqui na triangulação de corpora, seguindo os passos da textologia contrastiva. Detalhes dessa triangulação são apresentados na seção 4.1.6 em que as unidades de tradução mapeadas são apresentadas em diferentes cores para demonstrar contraste entre o texto analisa do e corpora de referência.

Há vantagens consideráveis de combinar dados de corpora paralelos, comparáveis e sincrônicos a métodos estatísticos diferentes gêneros, contextos e pares linguísticos, especialmente pares cujas línguas são de mesma família, como é o caso do par linguístico português-espanhol e esperamos demonstrá-los aqui ao apresentar a interface que aciona os algoritmos.

### 4.2.1 Configurações

Embora as configurações necessárias ao *Grinder* estejam na última aba (*Settings*), deve ser o primeiro passo revisar os parâmetros.

Figura 9 - Tela de Configurações do *Grinder*.



Fonte: Elaboração própria.

As opções apresentadas em *Fibonacci Ratios* são para calibragem da frequência mínima. Apresentamos na seção 4.1 o uso da razão 0.236. Já as opções apresentadas em *Other Options* se referem a

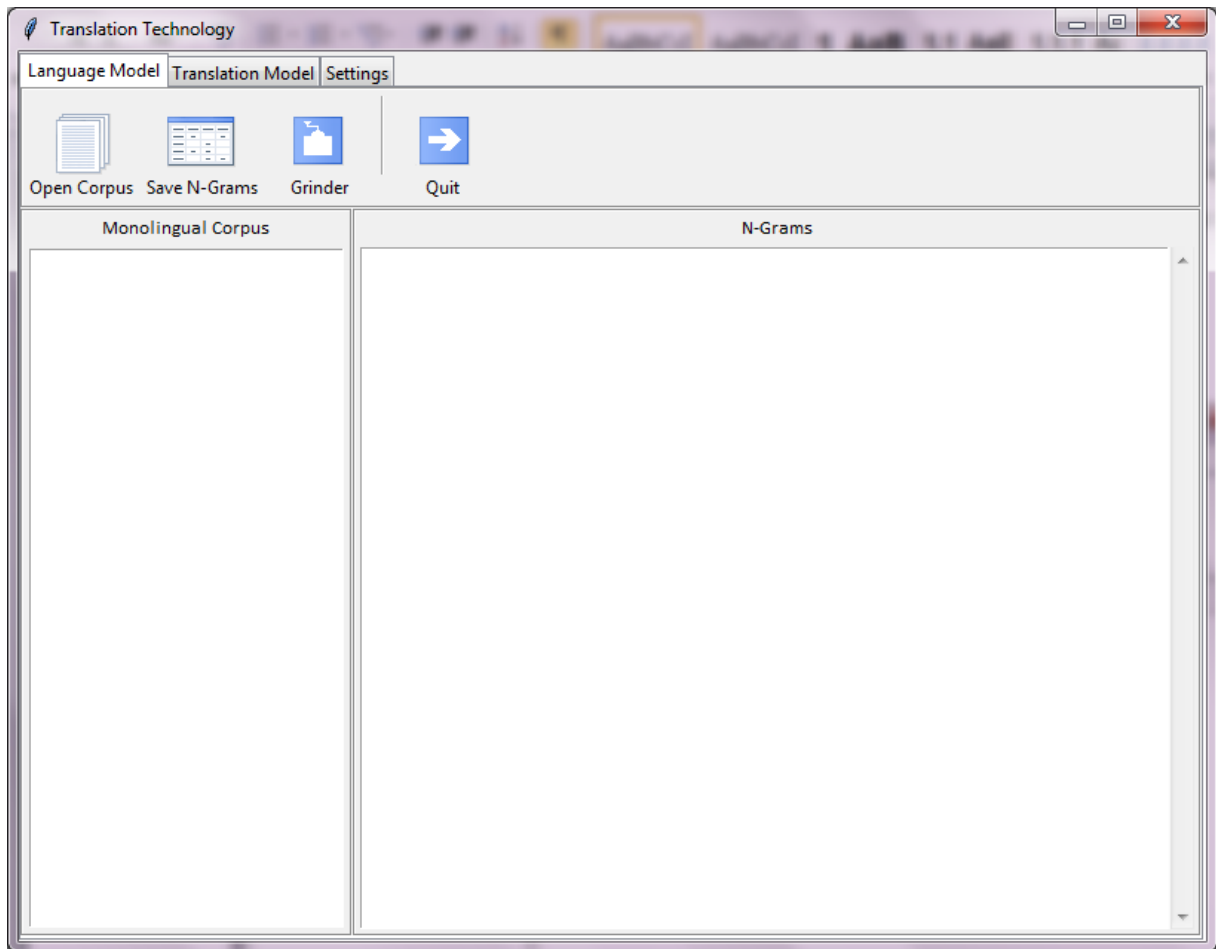
- *Treat all data as lowercase*: opção assinalada por padrão, utilizada no algoritmo *Filtrate* (ver seção 4.1.6). Determina se haverá distinção entre caracteres maiúsculos e minúsculos no mapeamento de unidades de tradução.
- *Treat all numbers*: opção assinalada por padrão, utilizada ao longo dos algoritmos para substituir todos os números por um parâmetro genérico, a saber `_num_`.
- *Treat all dates*: opção assinalada por padrão, utilizada ao longo dos algoritmos para substituir todos os números por um parâmetro genérico, a saber `_mês_` e `_ano_`.
- *Use spelling file*: se os corporas envolvidos necessitam de alguma substituição ortográfica, pode-se configurar um arquivos json para as substituições.
- *Replace all personal data*: se os corporas envolvidos contêm dados pessoais, os mesmo podem ser anonimizados.
- *Replace all URLs*: se os corporas envolvidos contêm, os mesmos podem ser substituídos.

- *Language Models*: Estes parâmetros implicam diretamente na triangulação ora comentada. Apresentaremos como criar os modelos de linguagem na seção a seguir.

#### **4.2.2 Modelo Linguístico**

Na aba *Language Model* os modelos de linguagem são criados. Os modelos constituem-se de unidades de linguísticas mapeadas e que serão usadas como referência para comparação entre o texto-fonte ao texto-alvo, isto é, a triangulação. São três os elementos de interface: a barra de ferramentas, a janela do corpus monolingual e a janela de N-Grams.

Figura 10 - Tela de Configurações do *Grinder*



Fonte: Elaboração própria.

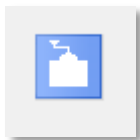
Figura 11 - Barra de ferramentas da tela de Modelo Linguístico



Abre um ou mais textos que compõe o corpora monolingual.



Salva N-Grams em um arquivo csv.



Aciona os algoritmos *sentences*, *nodes*, *redude*, *normalize* e *ordenize*.



Fechar o *Grinder*.

Fonte: Elaboração própria.

Figura 12 - Modelo Linguístico de Constituições de Língua Portuguesa

Monolingual Corpus	N-Grams
constituicao_Angola.txt constituicao_Cabo_Verde.txt constituicao_Guine_Bissau.txt constituicao_Macau.txt constituicao_Mocambique.txt constituicao_Portugal.txt constituicao_Sao_Tome_Principe.txt constituicao_Timor_Leste.txt	constituição são constituição que constituição ou constituição não constituição e pela lei constituição e na lei constituição e das leis constituição e da lei constituição e a constituição de constituição da república de angola constituição da república constituição da assembleia legislativa constituição assembleia popular nacional assembleia nacional é assembleia nacional popular assembleia nacional pode assembleia nacional ou assembleia nacional e assembleia nacional assembleia legislativa e

Fonte: Elaboração própria.

A compilação de corpora dependerá de que o usuário pense em questões como “Qual deve ser o tamanho de um corpus?”, “Que tipo de estratégia de amostragem deve ser usada para garantir a representatividade de um corpus?”, “Os textos traduzidos são comparáveis aos textos não traduzidos?”. Ao fornecer respostas a essas questões, pode-se ter certeza de que os modelos de linguagem serão úteis, ofereceram um número adequado de unidades de linguísticas para evidências convincentes para Estudos de Tradução.

Figura 13 - Modelo Linguístico de Constituições de Língua Espanhola

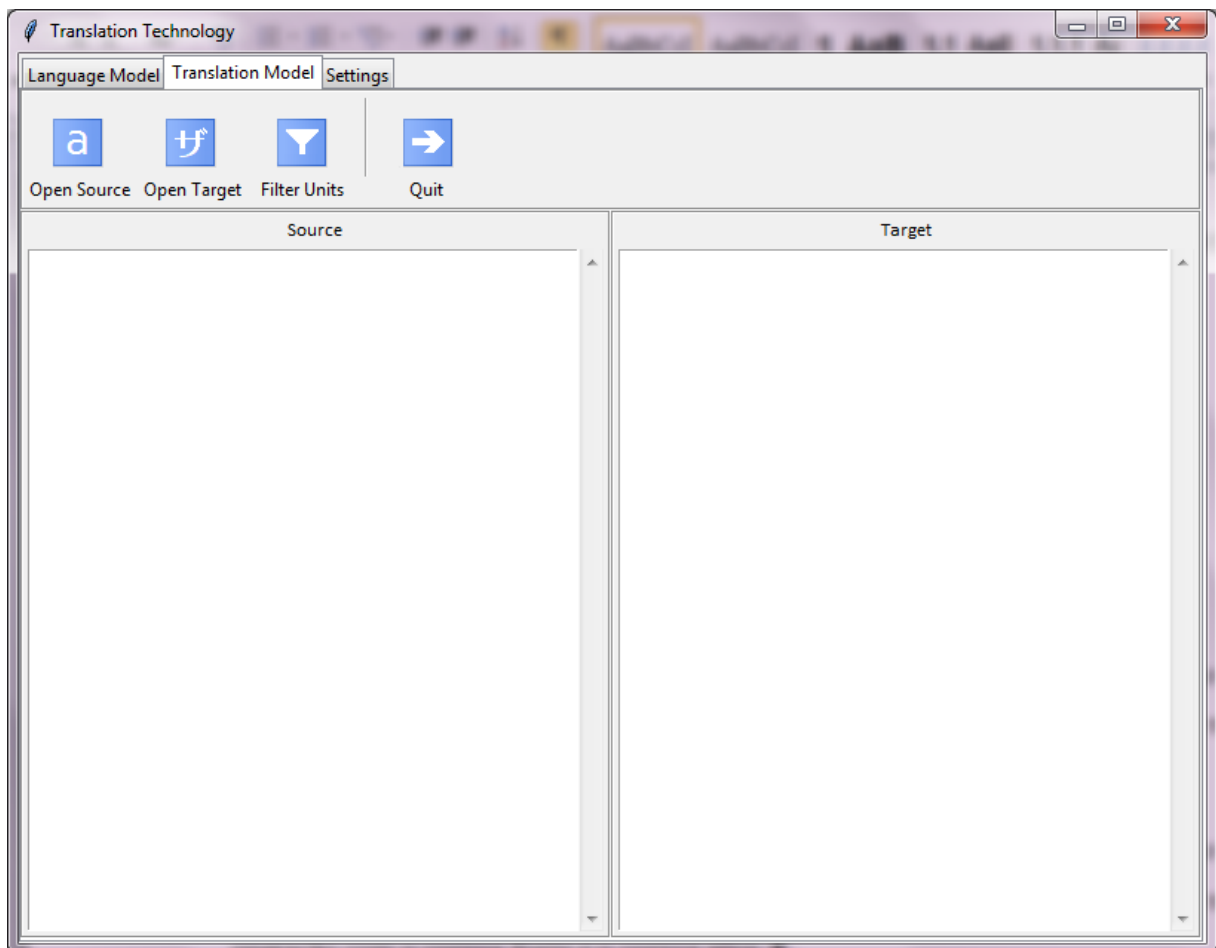
Monolingual Corpus	N-Grams
contituicion_Argentina.txt contituicion_Bolivia.txt contituicion_Chile.txt contituicion_Colombia.txt contituicion_Costa_Rica.txt contituicion_Cuba.txt contituicion_Ecuador.txt contituicion_El_Salvador.txt contituicion_Guatemala.txt contituicion_Honduras.txt contituicion_Mexico.txt contituicion_Nicaragua.txt contituicion_Panama.txt contituicion_Paraguay.txt contituicion_Peru.txt contituicion_Puerto_Rico.txt contituicion_Republica_Dominicana.txt contituicion_Uruguay.txt contituicion_Venezuela.txt	constitución y los constitución y las leyes constitución y la ley constitución y en los constitución y en la ley constitución y de las constitución y de constitución y a las leyes constitución y a la ley constitución y constitución se constitución política quedará así constitución política de los estados unidos mexicanos constitución política de la república de constitución para constitución o las leyes constitución o la ley constitución no constitución establece constitución en constitución de la república constitución

Fonte: Elaboração própria.

### 4.2.3 Modelo de Tradução

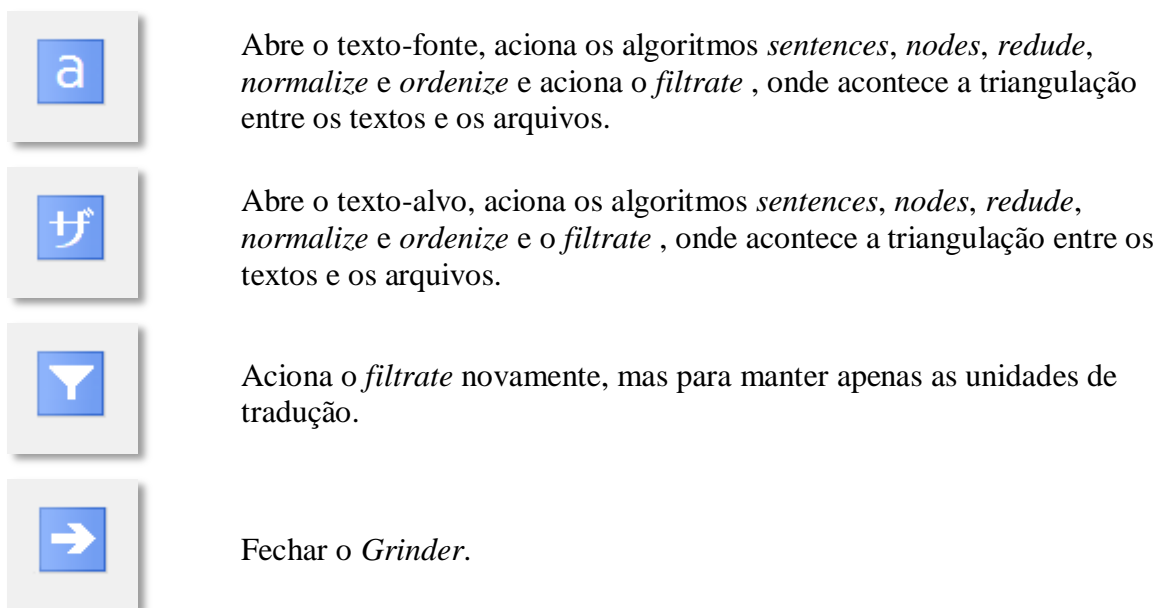
Quando se trata de Estudos de Tradução baseados em corpus, são aplicadas abordagens para observação e coleta de evidências da natureza da tradução, do processo de tradução ou das estratégias de tradução. As investigações serão acerca das características da tradução em relação ao par linguístico, compreendendo a natureza da tradução e descobrir padrões regulares de textos traduzidos, buscando descrever as características dos textos traduzidos em vários níveis (léxico, sintático ou discurso), as idiossincrasias do tradutor no uso de léxicos, estruturas sintáticas, pontuações e estruturas do discurso, bem como as motivações para essas idiossincrasias na perspectiva linguística e no contexto sociocultural; ou serão acerca da prática de tradução, explorando metodologia de corpus para identificar de equivalentes entre língua-fonte e língua-alvo, identificar regularidades da transferência de linguagem, e análise estratégias de tradução e métodos de tradução. São três os elementos de interface: a barra de ferramentas, a janela do texto-fonte e a janela do texto alvo.

Figura 14 - Tela de modelo linguístico



Fonte: Elaboração própria.

Figura 15 - Barra de ferramentas da tela de Modelo de Tradução



Fonte: Elaboração própria.

As Constituições são textos de mesmo contexto situacional, seja qual for o país a qual esteja vinculada: são fontes do Direito. Relembrando a sistemática de cores para identificação de unidades, preto é usado para unidades que foram localizadas apenas no texto, verde para unidades que foram localizadas no corpus de referência e azul para unidades localizadas no corpus acima da frequência mínima.

Essas unidades de significado tornam-se unidades de tradução conforme alinhamento. É de impressionar quão próxima do léxico está a tradução da Constituição Brasileira das demais Constituições Hispanoamercianas. Isso só é possível analisar com a triangulação é possível analisar e extrair ocorrências relevantes a comparar com outros documentos de mesmo contexto, observando tendências, padrões, fenômenos comuns e incomuns que emergem de uma amostra representativa do uso da língua, um recurso que consideramos valioso.

Podemos observe como as palavras ocorrem, como se associam, as palavras que se repetem. As unidades de tradução são salvas na sequência em são identificadas no texto. Caberá ao pesquisador formular hipóteses para os fenômenos linguísticos que forem sendo identificados.



Figura 16 - Modelo de Tradução da Constituição Brasileira

Source	Target
<p>Constituição da República Federativa do Brasil.  <b>Preâmbulo</b>  Nós, representantes do povo brasileiro, reunidos em Assembleia Nacional Constituinte para instituir um Estado Democrático, destinado a assegurar o exercício dos direitos sociais e individuais, a liberdade, a segurança, o bem-estar, o desenvolvimento, a igualdade e a justiça como valores supremos de uma sociedade fraterna, pluralista e sem preconceitos, fundada na harmonia social e comprometida, na ordem interna e internacional, com a solução pacífica das controvérsias, promulgamos, sob a proteção de Deus, a seguinte CONSTITUIÇÃO DA REPÚBLICA FEDERATIVA DO BRASIL.  <b>Título I</b>  <b>Dos Princípios Fundamentais</b>  Art. 1º A República Federativa do Brasil, formada pela união indissolúvel dos Estados e Municípios e do Distrito Federal, constitui-se em Estado democrático de direito e tem como fundamentos  I - a soberania;  II - a cidadania;  III - a dignidade da pessoa humana;  IV - os valores sociais do trabalho e da livre iniciativa;  V - o pluralismo político.  <b>Parágrafo único.</b> Todo o poder emana do povo, que o exerce por meio de representantes eleitos ou diretamente, nos termos desta Constituição.  Art. 2º São Poderes da União, independentes e harmônicos entre si, o Legislativo, o Executivo e o Judiciário.  <b>Translation Units</b>  constituição da república</p>	<p>Constitución de la República Federativa de Brasil  <b>PREÁMBULO</b>  Nosotros, los representantes del pueblo brasileño, reunidos en Asamblea Nacional Constituyente para constituir un Estado democrático, que asegure el ejercicio de los derechos sociales e individuales, la libertad, la seguridad, el bienestar, el desarrollo, la igualdad y la justicia como valores supremos de una sociedad fraterna, pluralista y sin prejuicios, fundada en la armonía social y comprometida, en el orden interno e internacional, con la solución pacífica de las controversias, promulgamos, bajo la protección de Dios, la siguiente Constitución de la República Federativa de Brasil.  <b>TÍTULO I</b>  <b>DE LOS PRINCIPIOS FUNDAMENTALES</b>  Artículo 1. La República Federativa de Brasil, formada por la unión indisoluble de los estados y municipios y del Distrito Federal, se constituye en Estado Democrático de Derecho y se fundamenta en:  I – la soberanía;  II – la ciudadanía;  III – la dignidad de la persona humana;  IV – los valores sociales del trabajo y de la libre iniciativa;  V – el pluralismo político.  <b>Párrafo único.</b> Todo el poder emana del pueblo, el cual lo ejerce por medio de representantes elegidos o directamente, en los términos establecidos por esta Constitución.  Artículo 2. Los Poderes de la Unión son, independientes y armónicos entre sí, el Legislativo, el Ejecutivo y el Judicial.  <b>Translation Units</b>  constitución de la república  federativa de brasil</p>

Fonte: Elaboração própria.

Quanto às demais funcionalidades do *Grinder*, podemos filtrar as unidades de tradução.

Figura 17 - UTs da Constituição Brasileira

Source	Target
<p><b>Translation Units</b>  constituição da república  federativa do brasil.  preâmbulo  nós  representantes do  povo  brasileiro,  reunidos  em  assembleia nacional  constituente  para  instituir  um estado  democrático,  destinado  a assegurar  o exercício dos direitos  sociais  e  individuais,  a  liberdade,  a segurança,  o bem-estar,  o desenvolvimento,  a  igualdade e  a justiça</p>	<p><b>Translation Units</b>  constitución de la república  federativa de brasil  preámbulo  nosotros,  los representantes del  pueblo  brasileño,  reunidos en  asamblea nacional constituyente  para  constituir  un estado  democrático,  que asegure  el  ejercicio de los  derechos sociales  e  individuales,  la libertad,  la  seguridad,  el bienestar,  el desarrollo,  la igualdad  y la  justicia  como  valores</p>

Fonte: Elaboração própria.

A unidades de tradução são salvas em arquivos \*.json.

Figura 18 - Arquivo .json do modelo linguístico das Constituições hispanoamericanas

```

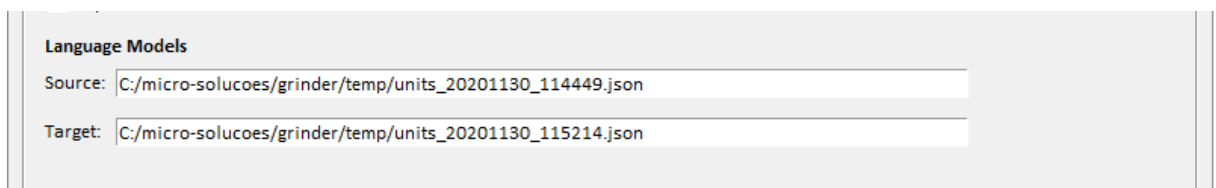
1  {
2    "constitución": [
3      "constitución y los",
4      "constitución y las leyes",
5      "constitución y la ley",
6      "constitución y en los",
7      "constitución y en la ley",
8      "constitución y de las",
9      "constitución y de",
10     "constitución y a las leyes",
11     "constitución y a la ley",
12     "constitución y",
13     "constitución se",
14     "constitución política quedará así",
15     "constitución política de los estados unidos mexicanos",
16     "constitución política de la república de",
17     "constitución para",
18     "constitución o las leyes",
19     "constitución o la ley",
20     "constitución no",
21     "constitución establece",

```

Fonte: Elaboração própria.

Os arquivos .json referentes a modelos de linguagem devem ser indicados na aba de configurações.

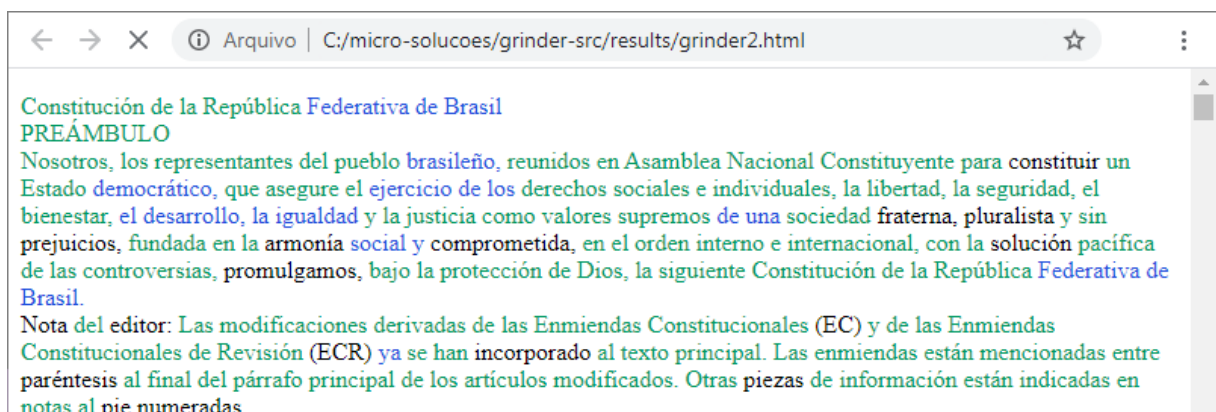
Figura 19 - Configuração de arquivos .json de modelos linguísticos.



Fonte: Elaboração própria.

Por fim, o resultado do processamento do Grinder é salvo em arquivos html, para que o pesquisador possa copiar e colar para outros formatos, matendo a sistemática de cores se assim desejar.

Figura 20 - Arquivo html com resultado do processamento de texto



Fonte: Elaboração própria.

## 5 RESULTADOS PRELIMINARES

O *Grinder* se apropria do básico da linguagem enquanto processo de comunicação: cada língua é constituída por um grande número de associações convencionais (coligações, colocações, binômios, expressões, marcadores conversacionais e fórmulas situacionais). Se o par linguístico analisado faz parte da mesma família linguística, haverá uma proximidade entre as línguas de tal modo a manter as mesmas estruturas sintáticas (intercompreensão). Há um padrão linguístico que pode ser reconhecido e contornar o fato de que computadores não são capazes de captar sentido. Como um estratagema, presumivelmente algoritmos são capazes de usos mecânicos da linguagem, realizando o processamento de modo a reconhecer padrões, ao que Dascal (2003) chama de modelo de interpretação superpragmático. Parece, portanto, natural, para esta concepção, reduzir a semântica à sintaxe:

- Segmentar e reconhecer símbolos.
- Construir e analisar sequências de símbolos bem formadas.
- Reconhecer referência a símbolos individuais.
- Reconhecer referência a sequências de símbolos.
- Usar sequências de símbolos de forma significativa, em um contexto de uso.
- Apresentar sequências de símbolos assim utilizadas.

Para estes primeiros resultados usaram-se corpora paralelos de Constituições não traduzidas em português e espanhol, com estes corpora criou-se o modelo linguístico para a análise com triangulação e corpus paralelo formado pelo texto da Constituição Brasileira e sua versão ao espanhol. A triangulação proporcionou o alinhamento de unidades de tradução sem que houvesse um algoritmo para isso. Analisando o trecho inicial da Constituição brasileira, temos os segmentos [1] e [2].

[1] [Constituição da República] [Federativa do Brasil].

[2] [Constitución de la República] [Federativa de Brasil].

O *Grinder* assinalou em verde as unidades de tradução “Constituição da República” e “Constitución de la República” porque são sequências reconhecidas nas Constituições dos diversos países. “Federativa do Brasil” é uma sequência de uso apenas no Brasil.

[3] [os recursos naturais] [da] [plataforma continental] [e da] [zona] [econômica] [exclusiva];

[4] [los recursos] [naturales] [de la] [plataforma continental] [y de la] [zona económica exclusiva]

As unidades de tradução reconhecidas nos segmentos [3] e [4], extraídos do artigo 20, não estão perfeitamente alinhadas. As unidades de tradução “recursos naturais” e “plataforma continental” são sequências encontradas nas Constituições de língua portuguesa. Nas Constituições hispano-americanas, a sequência frequente foi “los recursos naturales y”, com a conjunção ao final. Desse modo, não ocorreu a correspondência perfeita (do inglês *match* em processamento textual). Quanto às unidades “zona econômica exclusiva” e “zona económica exclusiva”, a correspondência não aconteceu pela dupla grafia para palavras proparoxítonas cuja vogal tônica possui variantes nas pronúncias (ê, é, ô, ó) permitida no Acordo Ortográfico de 1990<sup>21</sup>. Foi mapeada a unidade “zona económica exclusiva e” das Constituições de língua portuguesa oficial. Em todo caso, também ocorre a conjunção no mapeamento, a correspondência não ocorreria.

Nos segmentos a seguir, extraídos do artigo 22, novamente o uso de conjunção frequente em sequências distintas nos corpora utilizados provoca desalinhamento no mapeamento de unidades de tradução: “emprego” / “empleo y”. Há casos de frequências distintas de preposição “e condições” / “las condiciones para”, “para o exercício” / “ejercicio de”.

[5] [organização do] [sistema nacional de] [emprego] [e condições] [para o exercício] [de] [profissões]

[6] [la organización del] [sistema nacional de] [empleo y] [las condiciones para] [el] [ejercicio de] [las profesiones]

Sobre cultura refletida dos países, as organizações judiciárias são próximas, então o mapeamento das unidades de tradução em alinhamento quase perfeito, exceto pelo uso do artigo que é mais frequente em espanhol.

---

<sup>21</sup> HOUAISS. *Escrevendo pela nova ortografia: como usar as regras do novo acordo ortográfico da língua portuguesa*. 3 ed. São Paulo: Publifolha, 2009.

[7] [organização][judiciária],[do Ministério Público] [do Distrito Federal e dos Territórios] [e da] [Defensoria Pública] [dos Territórios], [bem como] [organização] [administrativa] [destes]

[8] [la] [organización] [judicial], [del Ministerio Público] [del Distrito Federal y de los Territorios] [y de la] [Defensoría Pública] [de los Territorios], [así como su] [organización] [administrativa]

A redação do artigo 125 segue estrutura sintática diferente na versão al espanhol. Apenas há correspondência das unidades “princípios” e “princípios”

[9] [Os Estados] [organizarão] [sua] [Justiça], [observados os] [princípios] [estabelecidos] [nesta Constituição]

[10][La organización de] [la] [Justicia em] [los estados] [corresponderá a] [cada] [ente federativo], [observando los] [princípios] [establecidos en esta Constitución]

No parágrafo 3º do artigo 199, há um caso de equivalência das expressões “é vedada a” e “se proíbe la” formada por sequências distintas. E mais uma vez, a frequência de uso de conjunções e preposições distintas no par linguístico não permitiu o alinhamento que seria perfeito entre as sequências “direta” “ou” “indireta” / “directa o indirecta” e “de” “empresas” / “de empresas”.

[11][É vedada a] [participação] [direta] [ou] [indireta] [de] [empresas] [ou] [capitais] [estrangeiros] [na] [assistência] [à saúde] [no País], [salvo nos casos] [previstos em lei].

[12][Se prohíbe la] [participación] [directa o indirecta] [de empresas] [o] [capitales] [extranjeros] [en la] [asistencia a] [la salud] [en el País], [salvo en los casos] [previstos por la ley].

No parágrafo único do artigo 243, observamos os casos de equivalência “qualquer” / “cada uno de los”, “apreendido” / “incautados”, “drogas afins” / “drogas relacionadas”, “destinação” / “destino” e “na forma da lei” / “de conformidad con lo establecido por la ley”.

[13][Todo] [e] [qualquer] [bem] [de] [valor] [econômico] [apreendido] [em] [decorrência] [do] [tráfico ilícito] [de] [entorpecentes] [e] [drogas afins] [e da] [exploração] [de trabalho] [escravo] [será] [confiscado] [e] [reverterá] [a] [fundo] [especial] [com] [destinação] [específica], [na forma da lei].

[14][Todos] [y] [cada uno de los] [bienes de] [valor] [económico] [incautados] [como] [resultado] [del] [tráfico] [ilícito] [de] [estupefacientes] [y] [drogas relacionadas] [y de la] [explotación] [del trabajo] [esclavo] [serán] [confiscados] [y] [revertidos] [a um] [fondo] [especial] [con um] [destino] [específico], [de conformidad con lo establecido por la ley].

E no último artigo da Constituição brasileira, artigo 250, encontramos as unidades “regime geral de previdência social” / “Régimen General de Previsión Social” que ocorrem apenas no caso brasileiro. Com este caso observamos que a triangulação formada pelo processamento de corpora monolíngues para criar o modelo linguístico versus o processamento dos corpora bilíngue para mapeamento das unidades de tradução funciona bem, porque aquelas unidades que não fazem parte do contexto comum cultural ou do gênero textual, será mapeado no próprio texto se o uso de frequência passar pela razão 0.236 (razão de *Fibonacci*).

[15][Com o] [objetivo] [de] [assegurar] [recursos] [para] [o pagamento] [dos benefícios] [concedidos] [pelo] [regime geral de previdência social], [em] [adição] [aos] [recursos de] [sua] [arrecadação], [a União] [poderá] [constituir] [fundo] [integrado] [por] [bens], [direitos e] [ativos] [de qualquer natureza], [mediante] [lei que] [disporá sobre a] [natureza e] [administração] [desse] [fundo]

[16][Con el objetivo de] [garantizar los] [recursos para] [el pago de] [los beneficios] [outorgados] [por el] [Régimen General de Previsión Social], [además de los] [recursos de] [su] [recaudación], [la Unión] [podrá] [constituir] [um] [fondo] [integrado por] [bienes], [derechos y] [activos] [de cualquier naturaliza], [mediante una ley] [que] [defina] [la naturaliza] [y la administración] [de esse] [fondo].

E revisitando os conceitos de UTs, não é o tamanho que define uma unidade linguística e, por conseguinte, uma unidade de tradução. É a associação convencional que a forma, bem como sua frequência de uso. Com o modelo linguístico de fato foi possível observar a convencionalidade, suficiente para o mapeamento de unidades de tradução. Podemos até nos arriscar a formular mais uma definição para o conceito de unidade de tradução: “unidades de tradução são unidades de significado formadas pelas associações convencionais na língua”.

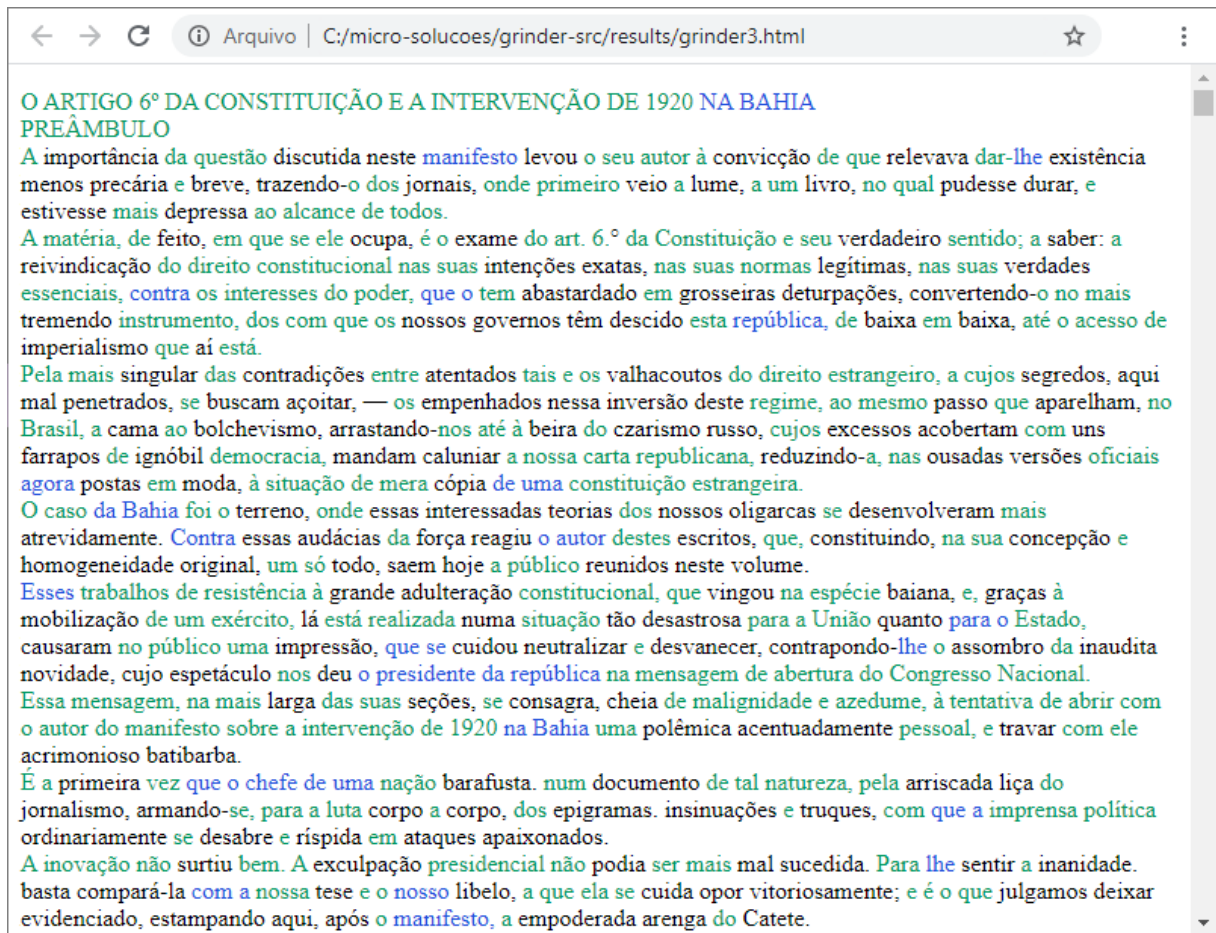
Outra análise realizada foi observar textos que realizam análise sobre artigos de uma Constituição comparados ao modelo linguístico criado para análise da Constituição Brasileira, desta vez, no modelo linguístico do português o corpus da Constituição brasileira foi incluído. O texto selecionado para este experimento foi o modelo linguístico em português foi *O artigo*

6º da constituição e a intervenção de 1920 na Bahia de Rui Barbosa (1920). O texto de Rui Barbosa comparada às Constituições de língua oficial portuguesa, apresentou 604 unidades de significado em comum e 523 unidades de significado próprias. Também foi possível observar o estilo do autor, que realiza algumas repetições como fórmula de persuasão.

a essa desordem	e tranquilidade
a intervenção	governo da bahia
a perturbação da ordem	governo do estado
a responsabilidade	governo federal
artigo _num_º da constituição	manifesto
consolidação da	necessários ao
constitucional	normas
da administração	nos termos do artigo _num_º
da constituição	o governo federal
da ordem e tranquilidade	o presidente da república
das populações	perturbação da ordem
de acordo com as	preâmbulo
declaração de guerra	presidente da república
direito	republicana
do congresso nacional	seu desenvolvimento
do poder executivo	seus interesses
e _num_ horas	

As compilações com as com os textos de Rui Barbosa e Gabo demonstraram que é possível observar o estilo do autor, quiçá tradutor se traduções fossem os textos analisados. A triangulação exposta na sistemática de cores auxilia na identificação das escolhas do autor. Se todas as UTs mapeadas pelo *Grinder* alimentarem sistemas de memória de tradução, ou mesmo sistemas de tradução automática, proveremos aos tradutores facilidades no ato de traduzir, resgatando dos pares linguísticos equivalências extraídas do uso, o que pode gerar um aprendizado de máquina e aprendizado para o tradutor.

Figura 21 - Arquivo html com resultado do processamento de texto de Rui Barbosa



Fonte: Elaboração própria

Gabriel García Márquez fez uma revisão completa da Constituição colombiana em 1990. *Hay que manosear la Constitución* é a entrevista conzida por Daniel Samper Pizano para El Tiempo. comparada às Constituições de países hispanoamericanos, apresentou 861 unidades de significado em comum e 437 unidades de significado próprias. Gabo aprofundou-se em diversos artigos, por isso a maior presença de expressões contidas nas Constituições. Outras UTs estão listadas no Apêndice deste trabalho.

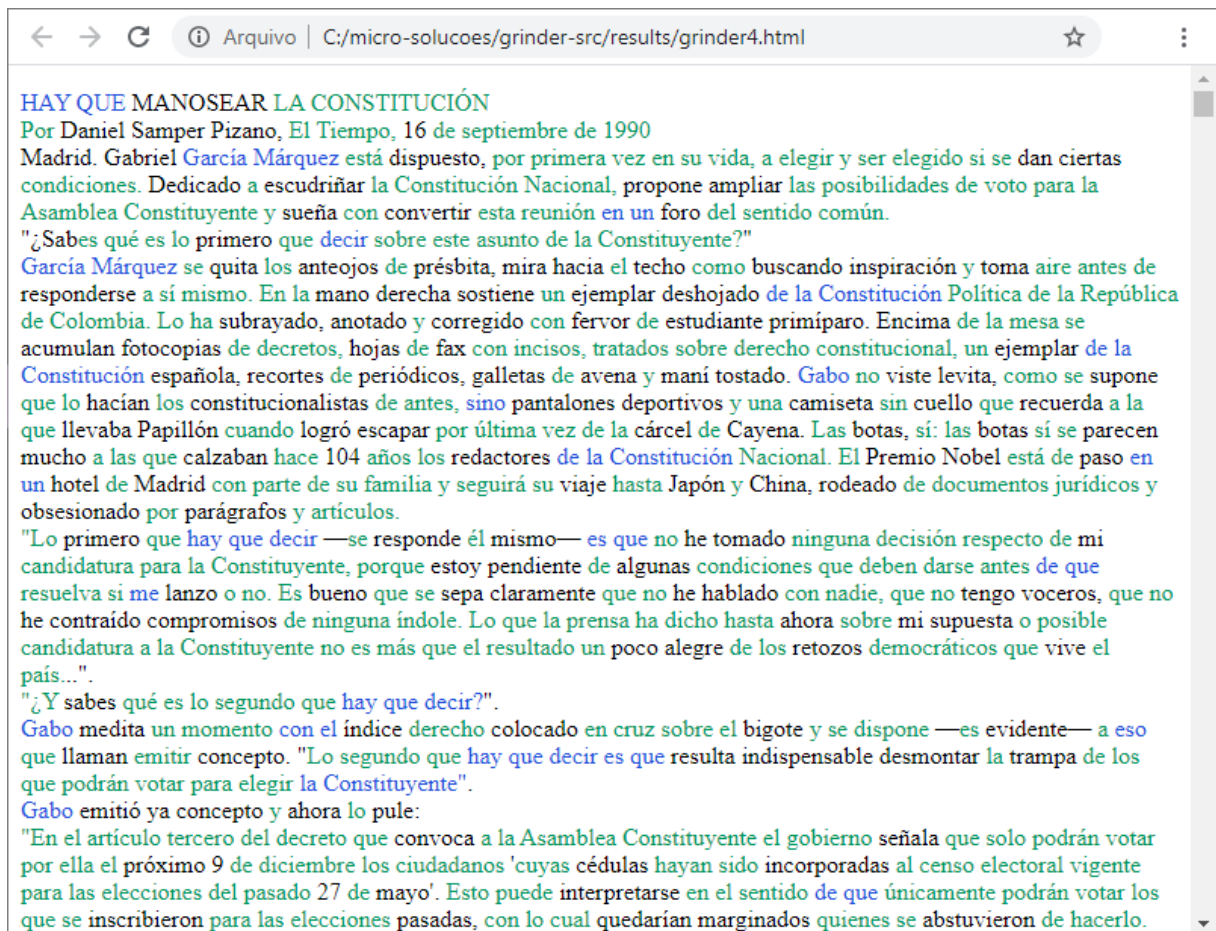
a los partidos  
 al mismo tiempo  
 asamblea constituyente  
 beneficio de la  
 cargos públicos  
 censo electoral  
 con el presidente de la república  
 condiciones que  
 de candidaturas  
 de identificación

de la constitución  
 de la educación  
 de la libertad de  
 de responsabilidad  
 del medio ambiente  
 derechos de la  
 educación pública  
 el artículo \_num\_  
 el consejo nacional electoral  
 el desarrollo de la



el ministro de	las circunstancias
el territorio de la república	los militares
elección presidencial	los procesos electorales
en caso de que	los votos de los
en consecuencia	materia electoral
en el exterior.	necesario para
en esta materia	necesidades de la
en la asamblea	para cualquier
en la elección	para las elecciones
esta constitución	para todos los
establecer las	política de la república de
están obligados a	por medio de la
extradición de nacionales	posibilidad de
información de	que la asamblea
inscripción de	recursos naturales
la asamblea constituyente	una circunscripción
la constitución	una comisión de
la constituyente	una institución
la facultad de	una obligación
la participación ciudadana	

Figura 22 - Arquivo html com resultado do processamento de texto de Gabo



Fonte: Elaboração própria

## CONSIDERAÇÕES FINAIS

Embora este seja um trabalho de graduação, portanto, com limite de tempo menor, tivemos êxito em conduzir o desenvolvimento do *Grinder*. Ainda há um caminho de amadurecimento, mas os resultados iniciais já são promissores. Ainda que não tenhamos desenvolvido algoritmo para alinhamento, as triangulações realizadas entre o modelo linguístico e o modelo de tradução constituíram um modelo superpragmático. Analisar corpus buscando identificar as associações convencionais dos pares linguísticos via algoritmos é factível. Nesta primeira experiência, os textos não foram alinhados, no entanto, através da sistemática de cores, estão em verde as unidades de significado mapeadas em corpus de referência e azul são unidades de significado não identificadas no corpus de referência, mas identificadas no próprio texto analisado, foi possível observar a equivalência naturalmente, no caso das constituições. O *Grinder* nos permitiu reconhecer padrões linguísticos e comparar corpora de tamanho significativo.

A sistemática de cores como solução para o alinhamento de texto permitiu uma visão diferenciada do que é uma marca de contexto situacional ou marca do autor do texto. Acreditamos que esta será uma fonte de conhecimento útil, para que o tradutor possa verificar terminologias, fraseologias, o estilo de traduções etc. Esta sistemática foi bem sucedida porque o alinhamento advém do reconhecimento de padrões de convencionalidade. Em suma, é o modelo de interpretação superpragmático entrando em ação. Simplesmente o que se faz é coletar estatística e tomar decisão conforme estatísticas coletadas. A sintaxe torna possível o modelo superpragmático, que funciona perfeitamente entre línguas de mesma família, ou de proximidade cultural.

Como próximas oportunidades de evolução do *Grinder*, haver tratamento de conjunções e preposições para garantir maior número de correspondências perfeitas (*matches*) ainda trará mais UTs mapeadas. Outro ponto de evolução é explorar outras formas alinhamento. Alguns softwares fazem alinhamento das unidades de tradução presumindo que a tradução geralmente segue a estrutura do texto original e que as sentenças são geralmente encadeadas da mesma forma no original e na tradução, podendo-se definir uma proporção de comprimento entre

dois pares de línguas (por exemplo, em termos de número de palavras, um texto em francês é geralmente 1,2 vezes maior do que o texto em inglês correspondente).

Esses *softwares* observam primeiro que as unidades de tradução variam em comprimento e, segundo, que geralmente há uma boa correlação entre o comprimento da sentença do texto fonte e o comprimento da sentença de o texto alvo. Pode-se tentar alinhar as unidades com base nisso; isto é, observando a diferença relativa de comprimento no original e procurar padrões semelhantes na tradução.

Outra estratégia a ser explorada é alinhar as unidades com base em correspondências lexicais, isto é, dada uma palavra ou mais palavras em sequência comum, tentar realizar a correspondência. Esta estratégia é menos genérica pode ser muito eficiente especialmente em línguas em que o princípio da intercompreensão acontece. Se considerarmos um determinado par linguístico tem frequentemente sequências semelhantes ou quase semelhantes devido às correspondências lexicais, geralmente chamadas de cognatos, o alinhamento poderá acontecer de modo eficiente. Todos esses elementos podem ser usados para identificar pontos de correspondência entre os textos de origem e de destino.

Por fim, para colocar o modelo superpragmático embasado em estatísticas, usamos razão de Fibonacci, mas outros métodos estatísticos podem ser avaliados no futuro. Podemos afirmar que o tipo de representação obtido a partir de uma análise estatística é apropriado e cognitivamente plausível. Noções como ambiguidade e polissemia estão intimamente ligadas ao uso e não são noções absolutas. Nesse sentido, é compreensível que a análise estatística possa ajudar a definir os vários significados de uma palavra, os diferentes contextos em que ela aparece, e assim por diante. Se podemos saber significado de um palavra pelo seu contexto de uso, um modelo superpragmático será capaz de mapear unidades de significado. É isso que o *Grinder* pode fazer. É preciso realizar mais mapeamentos e observar as diferença e ir estruturando melhores estratégias para o mapeamento automático de tradução.

Quanto à pesquisa que apoiou o desenvolvimento do *Grinder*, focada em unidades de tradução, equivalência, tradutor e ato de traduzir, poderá ser ampliada para contemplar Estudos Descritivistas da Tradução e abordagens prescritivistas, sobretudo para considerar as normas, as leis e os universais de tradução, ao mesmo tempo em que o *Grinder* possa ser utilizado como ferramenta que apoie uma pesquisa em tradução com objetivo de analisar traduções, diferenças entre traduções e seus originais e descrições comparativas.

## REFERÊNCIA BIBLIOGRÁFICA

ALPAC. *Language and machines: computers in translation and linguistics*. A report by the Automatic Language Processing Advisory Committee. Washington, DC: National Academy of Sciences. 1966. Disponível em: <<http://www.mt-archive.info/ALPAC-1966.pdf>>. Acessado em: 18 jul. 2020.

ALVES, Fábio. *Tradução, Cognição e Tecnologia: Investigando a interface entre o desempenho do tradutor e a tradução assistida por computador*. In Cadernos de Tradução v.2 n.14 (2005). U. F. Catarina (Ed.). Disponível em: <<https://periodicos.ufsc.br/index.php/traducao/article/view/6481/5975>>. Acessado em 10 jul. 2020.

\_\_\_\_\_. *Unidades de Tradução* in PAGANO, Adriana; et al. Traduzir com autonomia: estratégias para o tradutor em formação [recurso eletrônico]. 4. ed. São Paulo: Contexto, 2018.

BAKER, Mona. *Corpus Linguistics and Translation Studies: Implications and Applications*. In BAKER, Mona Baker; FRANCIS, Gil; TOGNINI-BONELLI, Elena (eds.). *Text and technology: in honour of John Sinclair*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 1993, pp. 233-252.

BALLARD, Michel. *Unit of translation*. In GAMBIER, Yves; DOORSLAER, Luc van (eds.). *Handbook of Translation Studies. Volume 1*. Amsterdam/Philadelphia: John Benjamins, 2010, pp. 437-440.

BARBOSA, Heloisa G. *Procedimentos técnicos da tradução: uma nova proposta*. Campinas: Pontes, 1990.

BARBOSA, Heloisa G.; NEIVA, Aurora M. S. *Investigando o processo tradutório*. Cadernos de Tradução v. 1 n. 2 (1997). U. F. Catarina (Ed.). Disponível em: <<https://periodicos.ufsc.br/index.php/traducao/article/view/5181/4607>>. Acessado em 10 jul. 2020.

BOGANTES, Claudio. *Depoimento Prof. Dr. Jørgen Schmitt Jesen in memoriam* in Revista Filologia e Linguística Portuguesa. N.7 (2005). Disponível em: <<https://www.revistas.usp.br/flp/article/view/59730/62830>>. Acessado em: 20 ago. 2020.

BOOTH, Andrew D. *Digital computers in action*. Oxford: Pergamon Press, 1965.

\_\_\_\_\_. *Andrew Donald Booth*. in *International biographical dictionary of computer pioneers*, John A N Lee (ed), 115-126. Chicago:Fitzroy Dearborn, 1995. Disponível em: <<https://history.computer.org/pioneers/booth-ad.html>>. Acessado em: 17 jul. 2020.

BRITO, Ana Maria; et al. *Gramática Comparativa: Houaiss: quatro línguas românicas*. São Paulo: Publifolha, 2013

CHOMSKY, Noam. *Syntactic Structures*. Berlin: Mouton de Gruyter, 1957.

DASCAL, Marcelo. *Interpretation and understanding*. Amsterdam/Philadelphia: John Benjamins, 2003.

HALLIDAY, Michael A. K.; HASAN, Ruqaiya. *Cohesion in English*. London: Longman, 1976.

HOUSE, Juliane. *Translation: A Multidisciplinary Approach* [recurso eletrônico]. Hampshire: Palgrave Macmillan, 2014.

JENSEN, Jørgen. *Nota Prévia* in BRITO et al. *Gramática Comparativa: Houaiss: quatro línguas românicas*. São Paulo: Publifolha, 2013

JOSSELSON, Harry H. *Automatic Translation of Languages Since 1960: A Linguist's View*. in *Advances in Computer, Volume 11*, Franz L. Alt, Morris Rubinoff, Marshall C. Yovits (eds), 1-58. New York: Academic Press, 1971. Disponível em: <<http://mt-archive.info/Josselson-1971.pdf>>. Acessado em: 19 jul. 2020.

KENNY, Dorothy. *Unit of translation*. In BAKER, Mona; SALDANHA, Gabriela (eds.). *The Routledge Encyclopedia of Translation Studies*. 2. ed. London/New York: Routledge, 2009. pp. 304-306.

KOCH, Ingedore G. V. *Desvendando os segredos do texto* [recurso eletrônico]. São Paulo: Cortez, 2018.

KNUTH, Donald. *Algorithms in Modern Mathematics and Computer Science*. Stanford Department of Computer Science Report No. STAN-CS-80-786, 1980. Disponível em: <http://i.stanford.edu/pub/cstr/reports/cs/tr/80/786/CS-TR-80-786.pdf>>. Acessado em: 15 jul. 2020.

KURZWEIL, Ray. *Como criar uma mente: os segredos do pensamento humano* [recurso eletrônico]. São Paulo: Aleph, 2015.

LAFFLING, John. *Towards High Precision Machine Translation: Based on Contrastive Textology*. Berlin and New York: Foris Publications, 1991.

LAVIOSA, Sara. *Corpora*. In GAMBIER, Yves; DOORSLAER, Luc van (eds.). *Handbook of Translation Studies. Volume 1*. Amsterdam/Philadelphia: John Benjamins, 2010, pp. 80-86.

LÉVY, Pierre. *As tecnologias da inteligência: o futuro do pensamento na era da informática*. 2 ed. Rio de Janeiro: Ed. 34, 2010.

MUNDAY, Jeremy. *Introducing Translation Studies: Theories and Applications*. London/New York: Routledge, 2001.

MUNDAY, Jeremy; HATIM, Basil. *Translation: na advanced resource book for students*. [recurso eletrônico]. 2 ed. New York: Routledge, 2019.

NORD, Christiane. *Lealdade em vez de fidelidade: proposta de uma tipologia funcional da tradução*. In *Cadernos de Tradução*, Porto Alegre, Número Especial, 2016, pp. 8-25.

PAGANO, Adriana; et al. *Traduzir com autonomia: estratégias para o tradutor em formação*. 4. ed. São Paulo: Contexto, 2018.

PERINI, Mário A. *Sintaxe* [recurso eletrônico]. São Paulo: Parábola 2020.

POSSAMAI, Viviane. *Associações sintagmáticas como unidades de tradução especializadas e proposta de ferramenta de apoio ao tradutor*. 2010. 263 p. Tese (Doutorado em Teorias do Texto e do Discurso) - Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul.

PYM, Antonie. *Explorando as teorias da tradução*; tradução de Rodrigo Borges de Faveri, Claudia Borges de Faveri, Juliana Steil. São Paulo: Perspectiva, 2017.

RICH, Elaine. *Inteligência Artificial*; tradução de Newton Vasconcellos. São Paulo: McGraw-Hill, 1988.

RIESEN, Kaspar. *Structural pattern recognition with graph edit distance: approximation algorithms and applications* [recurso eletrônico]. Switzerland: Springer, 2015

SANDERS, Ella F. *Lost in translation: na illustrated compendium of untranslatable words forma round the world* [recurso eletrônico]. Berkeley: Ten Speed Press, 2014.

\_\_\_\_\_. *De cafuné a ubuntu: 10 palavras 'sem tradução' famosas no mundo*. In BBC News Brasil, Londres, 15 nov. 2019. Disponível em: <<https://www.bbc.com/portuguese/brasil-50438128>>. Acessado em: 10 ago. 2020.

SCHNAIDERMAN, Boris. *Tradução, ato desmedido*. São Paulo: Perspectiva, 2011.

SINCLAIR, John. *Corpus, concordance, collocation*. Oxford: Oxford University Press, 1991.

\_\_\_\_\_. *Trust the text: language, corpus and discourse* [recurso eletrônico]. London/New York: Routledge, 2004.

STAËL, Madame de. *Da Alemanha*; traduzido por Edmir Míssio. São Paulo: Unespe Digital, 2017.

TAGNIN, Stella E. O. *O jeito que a gente diz* [livro eletrônico]: combinações consagradas em inglês e português. Barueri: DISAL, 2013.

TOURY, Gideon. *Descriptive Translation Studies and Beyond*. Amsterdam/Philadelphia: John Benjamins, 1995.

VINAY, Jean-Paul; DARBELNET, Jean. *Comparative Stylistics of French and English. A methodology for translation* (translated and edited by Juan C. Sager & M.J. Hamel). Amsterdam/Philadelphia: John Benjamins, 1995.

WEAVER, Warren. *Machine Translation of Languages*, MIT Press, Cambridge, MA, Reprinted from a memorandum written by Weaver in 1949.(1949/1955) Disponível em: <<http://www.mt-archive.info/Weaver-1949.pdf>. Acessado em: 18 jul. 2020.





## APÊNDICE A. Unidades de tradução mapeadas na Constituição Brasileira

A Constituição brasileira comparada, a outras Constituições de língua oficial portuguesa, apresentou 10.478 unidades de significado em comum com as demais Constituições e 10.798 unidades de significado presentes apenas na Constituição brasileira, este número contém repetições.

Foram identificadas fórmulas fixas (469 expressões):

_num_ por cento	art. _num_ é
§ _num_ a	art. _num_ o
§ _num_ °	art. _num_ os
§ _num_ ° a lei	art. _num_ são
§ _num_ ° as	art. _num_ °
§ _num_ ° compete	até _num_ habitantes
§ _num_ ° deste artigo	câmara dos deputados
§ _num_ ° do art. _num_	câmara dos deputados e
§ _num_ ° é vedada a	câmara municipal
§ _num_ ° lei	capítulo _num_
§ _num_ ° na	caput deste artigo
§ _num_ ° não	casas do congresso nacional
§ _num_ ° o estado	casos previstos nesta constituição
§ _num_ ° o presidente	com mais de _mês_ anos de
§ _num_ ° os	com mais de _num_
§ _num_ ° para	comissões parlamentares
§ _num_ ° são	competência e o funcionamento
§ _num_ ° se	competência exclusiva
§§ _num_ e _num_	competência legislativa
§§ _num_ ° e	competência para
a _num_ mil	comunicação social
à lei complementar	concessão ou permissão
a que se refere o art. _num_	concurso público de provas e títulos
a que se refere o inciso _num_	condições para a
art. _num_	congresso nacional
art. _num_ a lei	conselho da república
art. _num_ a união	conselho de defesa nacional
art. _num_ as	conselho nacional de justiça
art. _num_ compete	constitucionais e legais
art. _num_ e	constituição da república

constituição federal  
 constituição não  
 contribuições sociais  
 convenção internacional  
 convenções internacionais  
 corrente líquida  
 da administração  
 da administração direta  
 da administração pública  
 da arrecadação do imposto  
 da câmara dos deputados  
 da competência do  
 da comunicação social  
 da correspondência  
 da maioria absoluta  
 da nacionalidade  
 da ordem dos advogados do brasil  
 da ordem pública  
 da organização e  
 da pessoa humana  
 da política agrícola  
 da publicação da  
 da república para  
 da responsabilidade  
 da seguridade social  
 da sociedade e do estado  
 das decisões dos tribunais  
 das forças armadas  
 das instituições democráticas  
 das necessidades  
 de \_mês\_ anos de  
 de \_num\_  
 de \_num\_ anos  
 de \_num\_ de  
 de \_num\_ de \_num\_  
 de \_num\_ dias  
 de \_num\_ e \_num\_  
 de \_num\_ e \_num\_ anos  
 de \_num\_ habitantes e de até \_num\_  
 habitantes  
 de acordo com as  
 de acordo com os  
 de administração  
 de até \_num\_ habitantes  
 de autorização legislativa  
 de comunicação social  
 de contas da união  
 de defesa nacional  
 de desenvolvimento  
 de diretrizes orçamentárias  
 de economia mista  
 de funcionamento  
 de habeas corpus  
 de inconstitucionalidade  
 de interesse público  
 de mais de \_num\_ habitantes e de até  
 \_num\_ habitantes  
 de nacionalidade  
 de notável saber jurídico e  
 de oportunidades  
 de organização e  
 de partidos políticos  
 de previdência complementar  
 de previdência social  
 de qualquer natureza  
 de quarenta e cinco  
 de quatorze anos  
 de que trata este artigo  
 de que trata o §  
 de reforma agrária  
 de responsabilidade  
 de serviços públicos  
 de sessenta e cinco anos  
 de sua competência  
 de uma sociedade  
 décimos por \_num\_  
 decisão judicial  
 declarar a guerra e  
 decreto legislativo  
 defensoria pública  
 definidos em lei  
 dentre brasileiros  
 dentro de \_num\_  
 deputados à assembleia legislativa  
 deputados estaduais  
 desenvolvimento da  
 desenvolvimento do  
 desenvolvimento e  
 desta constituição  
 direitos e deveres  
 direitos políticos  
 direta e indireta  
 direta ou indiretamente  
 diretrizes orçamentárias  
 disposto na alínea  
 disposto no § \_num\_°  
 disposto no art. \_num\_  
 distrito federal  
 distrito federal e territórios  
 do § \_num\_°

do art. \_num\_  
do caput deste artigo  
do congresso nacional  
do desenvolvimento  
do distrito federal  
do distrito federal e dos municípios  
do distrito federal e dos territórios  
do distrito federal e territórios  
do inciso \_num\_  
do meio ambiente  
do ministério público  
do ministério público da  
do ministério público e  
do poder executivo  
do poder judiciário  
do poder público  
do presidente da república  
do presidente da república ou  
do presidente do  
do produto da arrecadação  
do respectivo ente federativo  
do senado federal  
do serviço militar  
do superior tribunal de justiça  
do supremo tribunal federal  
do território nacional  
do tribunal de contas da união  
dos deputados estaduais  
dos direitos e liberdades fundamentais  
dos estados e do distrito federal  
dos partidos políticos  
dos poderes públicos  
dos recursos naturais  
dos seus membros  
dos trabalhadores  
dos tribunais regionais  
e \_num\_ por cento  
e \_num\_  
e \_num\_ por cento  
e \_num\_ anos  
e \_num\_ e  
e \_num\_ mil  
e \_num\_ por \_num\_  
e \_num\_°  
e aproveitamento  
e da administração  
e de até \_num\_ habitantes  
e do distrito federal  
e do ministério público  
e dos municípios

é garantido o direito de  
e nos termos da lei  
e o funcionamento  
e o funcionamento dos órgãos  
e responsabilidade  
e serviços públicos  
e seus derivados  
em \_num\_  
em flagrante delito  
em lei complementar  
em qualquer caso  
em razão da matéria  
emprego ou função  
empresas públicas  
ensino fundamental  
entidades privadas  
entidades públicas  
estabelecidas na  
estabelecido por lei  
estabelecidos em lei  
estabelecidos na lei  
estabelecidos por lei  
estado de defesa  
estado democrático  
estados e do distrito federal  
exercer as funções  
exercício do direito  
exercício financeiro  
federativa do brasil  
fiscalização das  
formação profissional  
função jurisdicional  
funcionamento do  
funções públicas  
garantias fundamentais  
geral da república  
geral de previdência social  
habitantes e de até \_num\_ habitantes  
imposto previsto no inciso \_num\_  
imposto previsto no inciso iii  
incapacidade permanente  
inciso \_num\_  
inciso \_num\_ do  
inciso \_num\_ do § \_nm\_°  
inciso \_num\_ do § \_num\_°  
inciso \_num\_ do caput  
inciso \_num\_ do caput deste artigo  
incisos \_num\_  
incompatíveis com  
incompatível com

inconstitucional ou  
 independência nacional  
 independentemente da  
 infrações penais  
 instituições democráticas  
 instituições públicas  
 interesse nacional  
 interesse público  
 juízes dos tribunais  
 justiça do trabalho  
 justiça eleitoral  
 lei complementar  
 lei de diretrizes orçamentárias  
 liberdade de associação  
 liberdade de imprensa  
 língua portuguesa  
 maiores de \_num\_  
 maiores de \_num\_ anos  
 maioria absoluta  
 mais de \_num\_  
 mais de \_num\_ e  
 mais de \_num\_ habitantes e de \_mês\_ anos  
 de  
 mais de \_num\_ habitantes e de até \_num\_  
 habitantes  
 mandato de \_num\_ anos  
 máximo de \_num\_ dias  
 mediante autorização  
 medida provisória  
 medidas provisórias  
 meios de comunicação social  
 membros do ministério público  
 menos de \_num\_  
 menos de \_num\_  
 ministério público  
 ministério público e  
 ministério público é  
 ministro de estado  
 ministros de estado  
 ministros de estado e  
 ministros do supremo tribunal federal  
 municípios de mais de \_num\_ habitantes e  
 de até \_num\_ habitantes  
 na lei de diretrizes orçamentárias  
 na mesma sessão legislativa  
 nacional de justiça  
 nesta constituição  
 no § \_num\_°  
 no art. \_num\_  
 no distrito federal

no exercício dos  
 no inciso \_num\_ do  
 no prazo de \_num\_ dias  
 no território nacional  
 no todo ou em parte  
 nomeado pelo presidente da república  
 nomeados pelo presidente da república  
 nomear e exonerar os  
 nos §§ \_num\_ e \_num\_  
 nos casos previstos  
 nos demais casos  
 nos incisos \_num\_  
 nos municípios de mais de \_num\_  
 habitantes e de até \_num\_ habitantes  
 nos termos da lei  
 nos termos do art. \_num\_  
 o § \_num\_°  
 o art. \_num\_  
 o disposto no art. \_num\_  
 o imposto previsto no inciso \_num\_  
 o inciso \_num\_ do caput  
 observado o disposto no art. \_num\_  
 órgãos do poder judiciário  
 órgãos e entidades  
 para a realização  
 para assegurar o  
 para o desenvolvimento  
 para o exercício  
 para promulgação  
 parágrafo anterior  
 parlamentares de  
 participação dos  
 partidos políticos  
 partidos políticos e  
 partidos políticos que  
 pela maioria absoluta  
 pelo congresso nacional  
 pelo poder público  
 pelo presidente da república  
 pelo presidente da república e  
 pelo presidente do  
 pelo senado federal  
 pelo supremo tribunal federal  
 pelos respectivos  
 perante o tribunal  
 perda do mandato  
 plano plurianual  
 planos nacionais  
 plataforma continental  
 poder judiciário

poder legislativo  
 poder regulamentar  
 poderá estabelecer  
 poderes públicos  
 política agrícola  
 por \_num\_  
 por \_num\_ do subsídio  
 por cento do subsídio  
 por crime doloso  
 por decisão judicial  
 por intermédio de  
 por lei complementar  
 portadores de deficiência  
 possibilidade de  
 prazo de \_num\_ dias  
 presidente da república  
 presidente da república a  
 presidente da república e  
 presidente do conselho  
 presidente do senado  
 previdência complementar  
 previdência social  
 previstas em lei  
 previsto no inciso \_num\_  
 previsto no inciso i  
 previstos em lei  
 previstos nesta constituição  
 prisão ou detenção  
 processar e julgar  
 processo eleitoral  
 produto da arrecadação  
 promover o desenvolvimento  
 proposta orçamentária  
 propriedade privada  
 próprio de previdência social  
 provas e títulos  
 pública ou privada  
 quaisquer outras  
 quaisquer outros  
 qualidade de vida  
 qualquer natureza  
 que a lei determinar  
 que se refere o art. \_num\_  
 que se refere o inciso \_num\_  
 que trata este artigo  
 receita corrente líquida  
 recursos públicos  
 refere o art. \_num\_  
 refere o inciso \_num\_  
 regime geral de previdência social  
 regime próprio de previdência social  
 regionais federais  
 regulado por lei  
 relacionadas com  
 relações internacionais  
 relativamente ao  
 relevante interesse nacional  
 representantes do  
 representantes dos  
 república federativa do brasil  
 reputação ilibada  
 responsabilidade civil  
 ressalvado o disposto  
 salvo em caso de flagrante delito  
 são iguais perante a lei  
 são reconhecidos  
 se refere o art. \_num\_  
 se refere o inciso \_num\_  
 seção \_num\_  
 secretários de estado  
 seguintes princípios  
 seguridade social  
 sem prejuízo das  
 serviços auxiliares  
 serviços públicos  
 servidor público  
 servidores públicos  
 sessão legislativa  
 seu desenvolvimento  
 seus direitos e interesses  
 sistema financeiro  
 sistema nacional de  
 soberania nacional  
 social e cultural  
 sociedades de economia mista  
 sua organização e  
 sufrágio universal  
 superior a \_num\_  
 superior a \_num\_ dias  
 superior do trabalho  
 superior tribunal de justiça  
 supremo das forças armadas  
 supremo tribunal federal  
 supremo tribunal federal e  
 tempo de contribuição  
 tempo de serviço  
 termos do art. \_num\_  
 território nacional  
 título \_num\_  
 transitada em julgado

trânsito em julgado  
 tratados internacionais  
 tribunais de contas  
 tribunais e juízes  
 tribunais regionais  
 tribunais regionais federais  
 tribunais superiores  
 tribunal de contas da união

tribunal de justiça  
 tribunal militar  
 tribunal regional  
 tribunal superior do trabalho  
 última instância  
 utilidade pública  
 vigência do estado de

## **APÊNDICE B. Unidades de tradução mapeadas na versão da Constituição Brasileira**

A versão da Constituição brasileira ao espanhol, comparada a outras Constituições de países 94rasil94-americanos, apresentou 15.811 unidades de significado em comum com as demais Constituições e 7.776 unidades de significado presentes apenas na Constituição brasileira, estes números contêm repetições. Podemos aqui elaborar a hipótese de que a versão da Constituição brasileira aproxima-se mais das Constituições hispanoamericanas que a Constituição brasileira aproxima-se das Constituições de língua espanhola. Em análise inicial, foi possível verificar que as Constituições dos outros países de língua oficial portuguesa seguem a Constituição de Portugal como modelo.

Foram identificadas fórmulas fixas (1056 expressões) :

a \_num\_ de  
 a \_num\_ habitantes  
 a excepción de los  
 a la formación de  
 a la información  
 a la jurisdicción  
 a la organización  
 a la participación  
 a la presidencia de la república  
 a la seguridad social  
 a las comunidades  
 a las condiciones  
 a las necesidades  
 a las que se refiere  
 a los ciudadanos  
 a los efectos de  
 a los extranjeros  
 a los funcionarios  
 a los miembros del  
 a los ministros de estado  
 a los municipios

a los trabajadores  
 a los tribunales  
 a partir de la fecha de  
 a que se refiere el  
 a que se refiere el párrafo \_num\_  
 a que se refiere este artículo  
 a que se refieren los  
 acciones de amparo  
 acciones y servicios  
 actividad económica  
 actividades económicas  
 actividades productivas  
 actos administrativos  
 actos u omisiones  
 acuerdo con esta constitución  
 además de las que  
 administración de justicia  
 administración pública  
 administración pública directa  
 administrativa o  
 administrativa y

administrativas y  
 administrativo y  
 administrativos y  
 al \_num\_  
 al 22 de diciembre  
 al congreso nacional  
 al desarrollo de  
 al distrito federal y a los municipios  
 al financiamiento  
 al medio ambiente  
 al ministerio público  
 al poder legislativo  
 al poder público  
 al presidente de la república  
 al presidente de la república para  
 al presidente del  
 al servicio de la  
 al servicio público  
 al vicepresidente  
 años consecutivos  
 ante el congreso  
 ante el tribunal  
 antes de la fecha  
 apartado \_num\_ del  
 apartado \_num\_ del enunciado de este  
 artículo  
 apartado \_num\_ del párrafo \_num\_  
 apartados \_num\_  
 apartados \_num\_ y  
 aplicación de la  
 aplicación de las  
 aprobación del congreso  
 aprobados por el  
 aprovechamiento de  
 artículo \_num\_  
 artículo \_num\_ a  
 artículo \_num\_ corresponde a  
 artículo \_num\_ corresponde al  
 artículo \_num\_ de  
 artículo \_num\_ de esta constitución  
 artículo \_num\_ el  
 artículo \_num\_ en  
 artículo \_num\_ la ley  
 artículo \_num\_ la unión  
 artículo \_num\_ las  
 artículo \_num\_ los  
 artículo \_num\_ se  
 artículo \_num\_ son  
 artículo \_num\_ y  
 artículo anterior

artículos \_num\_  
 artículos \_num\_ y \_num\_  
 asamblea legislativa  
 asamblea nacional constituyente  
 así como también  
 asistencia social  
 asistencia técnica  
 asuntos de interés  
 atención integral  
 atribuciones del  
 atribuciones que le  
 autonomía funcional  
 autoridades administrativas  
 autorización del  
 autorización del congreso  
 autorización legislativa  
 autorización para  
 autorización previa  
 beneficios de la  
 bienes inmuebles  
 bienes y servicios  
 cada cuatro años  
 cada legislatura  
 cada una de las cámaras  
 calamidad pública  
 cámara de diputados  
 cámara respectiva  
 cámaras del congreso nacional  
 capítulo \_num\_  
 carácter nacional  
 carácter público  
 caso de flagrante delito  
 casos previstos en esta constitución  
 casos previstos por  
 causa justificada  
 celebrar contratos  
 ciento veinte días  
 comercialización de  
 comisiones permanentes  
 competencia de la  
 competencia de la justicia  
 competencia de los  
 complementaria a que se refiere el  
 complementaria establecerá  
 comprendido entre  
 comunicación social  
 comunidades indígenas  
 con arreglo a la ley  
 con autorización  
 con criterios de

con discapacidad  
 con el artículo \_num\_  
 con el objetivo de  
 con el voto de la  
 con excepción de  
 con excepción de las  
 con excepción de los  
 con fuerza de ley  
 con fundamento en  
 con independencia  
 con las disposiciones  
 con lo dispuesto en  
 con lo dispuesto en el artículo \_num\_  
 con los requisitos  
 con más de diez años de  
 con participación  
 conceder indultos  
 concejo municipal  
 concejos municipales  
 concesión o permiso  
 concurso público de  
 condiciones especiales  
 condiciones para el  
 confiscación de bienes  
 conforme a la ley  
 conforme a lo dispuesto en  
 conformidad con la ley  
 conformidad con lo dispuesto en el  
 artículo \_num\_  
 conformidad con lo establecido por la  
 ley  
 congreso nacional  
 congreso nacional y  
 consejo de defensa nacional  
 consejo de la república  
 consejo nacional  
 consejo nacional de justicia  
 constitución de la república  
 constitucional de la  
 constitucionales y legales  
 contados a partir de la  
 contados desde la  
 contra los derechos  
 contribuciones sociales  
 corresponde a la  
 corresponde a los  
 corresponde al estado  
 corresponderá al  
 correspondiente a  
 correspondientes a

crímenes de responsabilidad  
 cualquier autoridad  
 cualquier naturaleza  
 cualquier otra  
 cualquier persona  
 cualquier título  
 cualquiera de las  
 cualquiera de las cámaras  
 cualquiera de los  
 cualquiera de sus  
 cualquiera que sea su  
 cuando se trate de  
 cuentas de la unión  
 cumplimiento de la  
 cumplimiento de las  
 cumplimiento de los  
 cumplimiento de los requisitos  
 de \_mês\_  
 de \_mês\_ de \_ano\_  
 de \_num\_ a \_num\_ habitantes  
 de \_num\_ a \_num\_ y \_num\_ años  
 de \_num\_ años  
 de \_num\_ de septiembre de \_ano\_  
 de \_num\_ días  
 de acuerdo con la  
 de acuerdo con lo  
 de acuerdo con los  
 de administración y  
 de ambas cámaras  
 de amplia defensa  
 de bienes y servicios  
 de cada una de las  
 de competencia de  
 de comunicación social  
 de comunicación y  
 de conformidad con el  
 de conformidad con la ley  
 de conformidad con las  
 de conformidad con lo dispuesto en el  
 artículo \_num\_  
 de conformidad con lo establecido en  
 de conformidad con lo establecido por  
 la ley  
 de conformidad con lo que  
 de conformidad con los  
 de constitucionalidad  
 de cualquier naturaleza  
 de cualquier otra  
 de cualquier otro  
 de cualquiera de



de cuentas de la unión  
 de derecho público  
 de derechos humanos  
 de dieciocho años  
 de directrices presupuestarias  
 de dos tercios de  
 de economía mixta  
 de entre brasileños  
 de esta constitución  
 de este artículo  
 de financiamiento  
 de fiscalización  
 de funcionamiento  
 de identificación  
 de inconstitucionalidad  
 de infraestructura  
 de iniciativa de  
 de instituciones  
 de investigación  
 de la administración  
 de la administración directa  
 de la administración pública  
 de la aprobación  
 de la cámara de diputados  
 de la competencia  
 de la competencia de  
 de la constitución  
 de la defensoría pública  
 de la deuda pública  
 de la federación  
 de la institución  
 de la investigación  
 de la judicatura  
 de la magistratura  
 de la mayoría absoluta  
 de la nacionalidad  
 de la organización  
 de la persona humana  
 de la previsión social  
 de la producción  
 de la recaudación de  
 de la recaudación del impuesto  
 de la república federativa de Brasil  
 de la república para  
 de la respectiva  
 de la responsabilidad  
 de la seguridad social  
 de la sociedad y  
 de la unión y de  
 de las actividades  
 de las atribuciones  
 de las contribuciones  
 de las disposiciones  
 de las entidades  
 de las fuerzas armadas  
 de las funciones  
 de las instituciones  
 de las respectivas  
 de las retribuciones  
 de lo dispuesto en el  
 de lo dispuesto en los  
 de los \_num\_  
 de los artículos  
 de los beneficios  
 de los candidatos  
 de los concejales  
 de los derechos humanos  
 de los derechos políticos  
 de los derechos y garantías  
 de los diputados estatales  
 de los electores  
 de los empleados  
 de los estados y del distrito federal  
 de los funcionarios  
 de los gobernadores  
 de los impuestos  
 de los intereses  
 de los límites de  
 de los miembros de  
 de los ministros  
 de los ministros de estado  
 de los municipios  
 de los organismos  
 de los párrafos \_num\_  
 de los partidos políticos  
 de los poderes públicos  
 de los principios  
 de los productos  
 de los proyectos de ley  
 de los pueblos indígenas  
 de los recursos de  
 de los respectivos  
 de los seis meses  
 de los servicios  
 de los servicios públicos  
 de los territorios  
 de los trabajadores  
 de los tribunales  
 de los tribunales de  
 de los tribunales regionales

de los tribunales superiores  
 de los votos válidos  
 de más de la mitad  
 de notable saber jurídico  
 de oposición y méritos  
 de ordenamiento territorial  
 de padre o madre  
 de participación  
 de presidente y vicepresidente  
 de presupuesto anual  
 de presupuesto de  
 de previsión complementaria  
 de previsión social  
 de primera instancia  
 de reforma \_num\_/\_ano\_  
 de responsabilidad  
 de segunda instancia  
 de segundo grado  
 de seguridad social  
 de servicios públicos  
 de su competencia  
 de sus derechos y  
 de sus entidades  
 de sus funciones y  
 de sus habitantes  
 de sus respectivos  
 deber del estado  
 décimas por ciento  
 décimas por ciento)  
 declarar el estado de  
 declarar la guerra  
 decretos legislativos  
 defensa nacional  
 defensoría pública  
 del año anterior  
 del apartado \_num\_ del  
 del artículo \_num\_  
 del artículo \_num\_ de  
 del artículo \_num\_ de esta constitución  
 del artículo \_num\_ y  
 del congreso nacional  
 del conocimiento  
 del derecho a la  
 del distrito federal  
 del distrito federal y de los municipios  
 del distrito federal y de los territorios  
 del distrito federal y territorios  
 del enunciado de este artículo  
 del estado civil  
 del medio ambiente  
 del mes de julio de  
 del ministerio público  
 del ministerio público de la unión  
 del ministerio público y  
 del ordenamiento  
 del párrafo \_num\_  
 del patrimonio cultural  
 del período presidencial  
 del poder ejecutivo  
 del poder judicial  
 del poder legislativo  
 del poder público  
 del presente artículo  
 del presidente  
 del presidente de la república  
 del presidente de la república o  
 del presidente del  
 del procurador general de la  
 del producto de la recaudación  
 del proyecto de ley  
 del respectivo ente federativo  
 del senado federal  
 del sistema nacional de  
 del superior tribunal de justicia  
 del supremo tribunal federal y  
 del territorio nacional  
 del tribunal de cuentas de la unión  
 del vicepresidente de la república  
 dentro de los \_num\_  
 dentro de los \_num\_ días  
 dentro de los límites  
 dentro de un plazo  
 derecho a la vida  
 derechos e intereses  
 derechos humanos  
 derechos políticos  
 derechos sociales  
 derechos y deberes  
 derogado por la ec \_num\_/\_ano\_  
 desarrollo de la  
 desarrollo de las  
 desarrollo económico y social  
 desarrollo nacional  
 desarrollo regional  
 desde el momento  
 desde la fecha de  
 designado por el  
 designados por el  
 después de haber  
 determinación de

días a partir de  
 días después de la  
 diputados estatales  
 diputados y senadores  
 directa e indirecta  
 directa o indirecta  
 directa o indirectamente  
 directamente o por  
 directrices presupuestarias  
 disposiciones de  
 disposiciones de la  
 disposiciones de los  
 disposiciones del artículo \_ano\_  
 disposiciones del artículo \_num\_  
 disposiciones generales  
 disposiciones que  
 dispuesto en el artículo \_num\_  
 dispuesto en el párrafo \_ano\_  
 dispuesto en el párrafo \_num\_  
 distribución de los  
 distrito federal  
 distrito federal o  
 distrito federal y los municipios  
 distrito federal y territorios  
 dos tercios de los miembros  
 dos tercios de sus miembros  
 durante el período  
 durante el receso  
 durante la vigencia  
 e imprescriptible  
 e incompatibilidades  
 ec \_num\_/\_ano\_  
 ec de reforma \_num\_/\_ano\_  
 económica y social  
 económico y social  
 económico y social de  
 educación básica  
 efecto retroactivo  
 ejercer el control  
 ejercer funciones  
 ejercer las demás atribuciones  
 ejercicio de los  
 ejercicio financiero  
 el \_num\_ por ciento  
 el apartado \_num\_ del enunciado de  
 este artículo  
 el artículo \_num\_  
 el impuesto previsto en el apartado  
 \_num\_  
 el párrafo \_num\_ de este artículo  
 el párrafo \_num\_ del artículo \_num\_  
 empleados públicos  
 empleo o función  
 empleos públicos  
 empresas estatales  
 empresas públicas  
 en condiciones de  
 en cualquier caso  
 en cumplimiento de  
 en ejercicio de sus  
 en el apartado \_num\_ del  
 en el artículo \_num\_  
 en el artículo anterior  
 en el caso de las  
 en el caso de los  
 en el caso de que  
 en el cumplimiento de  
 en el distrito federal  
 en el ejercicio de  
 en el ejercicio de su  
 en el ejercicio de sus funciones y  
 en el enunciado de este artículo  
 en el extranjero  
 en el marco de la  
 en el párrafo \_num\_ de este artículo  
 en el párrafo \_num\_ del artículo \_num\_  
 en el registro electoral  
 en el respectivo  
 en el territorio nacional  
 en esta constitución  
 en este artículo  
 en la administración  
 en la forma establecida  
 en la forma prevista  
 en la forma que determine la ley  
 en la formulación  
 en la ley de directrices presupuestarias  
 en las elecciones de  
 en las entidades  
 en las operaciones  
 en lo que corresponda  
 en los apartados \_num\_  
 en los artículos \_num\_  
 en los artículos \_num\_ y \_num\_  
 en los beneficios  
 en los casos previstos en  
 en los casos previstos por  
 en los municipios  
 en los párrafos \_num\_  
 en los siguientes

en los términos establecidos  
 en los términos establecidos por la ley  
 en los términos que establezca la ley  
 en los términos y  
 en materia penal  
 en relación con el  
 en relación con los  
 en todo el territorio nacional  
 en todos los niveles de  
 en un plazo de num días  
 en un plazo de cinco días  
 en un plazo de noventa días  
 en única instancia  
 en votación secreta  
 entidades autárquicas  
 entidades autárquicas y  
 entidades de la administración  
 entidades públicas  
 entrada en vigencia de  
 entrada en vigor  
 entre brasileños  
 enunciado de este artículo  
 es deber del estado  
 es incompatible con  
 es una institución  
 esta constitución  
 esta constitución para  
 esta constitución y  
 está integrado por  
 establece la ley  
 establecerá reglas sobre  
 establecidas en el  
 establecidas en esta constitución  
 establecidas en la ley  
 establecidas por la ley  
 establecido en el  
 establecido en el artículo num  
 establecido en la ley  
 establecido por la ley  
 establecidos en el  
 establecidos en esta constitución  
 establecidos en la constitución  
 establecidos en la ley  
 establecidos por la ley  
 establecimiento de  
 establezca la ley  
 estado de defensa  
 estado de guerra  
 estados y del distrito federal  
 estados y el distrito federal  
 estados y municipios  
 están obligados a  
 estará a cargo de un  
 estarán sujetos a  
 este artículo se  
 excepto en los casos  
 exclusivamente a  
 exclusivamente por  
 federativa de 100rasil  
 fines específicos  
 fiscalización de los  
 flagrante delito  
 forman parte del  
 funcionamiento de  
 funcionamiento de los  
 funcionamiento del  
 funciones públicas  
 fundamentales de la  
 gastos de personal  
 general de la república  
 general de la república y  
 general de la unión  
 general de previsión social  
 gobierno federal  
 goce de los derechos políticos  
 hacienda pública  
 hombres y mujeres  
 horas siguientes  
 igualdad de oportunidades  
 impuesto previsto en el apartado  
num  
 incompatible con  
 incompatibles con  
 inconstitucionalidad de  
 incorporada por la ec num/ano  
 incorporado por la ec num/ano  
 incumplimiento de  
 independencia de  
 independientemente de  
 infracciones penales  
 ingresos corrientes  
 iniciativa de la  
 iniciativa de ley  
 instituciones de  
 integrantes de la  
 integridad física  
 interés nacional  
 internacionales que  
 internacionales sobre  
 internacionales y

intervención federal  
 intervenir en la  
 inversión pública  
 investigación científica y  
 inviolabilidad de  
 judicial competente  
 jueces de derecho  
 jueces federales  
 justicia electoral  
 justicia federal  
 justicia laboral  
 justicia militar  
 la \_num\_  
 la ec \_num\_/\_ano\_  
 la ec de reforma \_num\_/\_ano\_  
 las disposiciones del artículo \_num\_  
 le corresponderá  
 ley complementaria  
 ley de directrices presupuestarias  
 ley de presupuesto  
 ley presupuestaria  
 leyes complementarias  
 libre nombramiento y remoción  
 límites establecidos  
 lo dispuesto en el artículo \_num\_  
 lo dispuesto en el párrafo \_num\_  
 los \_num\_  
 los apartados \_num\_  
 los artículos \_num\_  
 los municipios que tengan de \_num\_ a  
 \_num\_ habitantes  
 los párrafos \_num\_ y \_num\_  
 mandato electivo  
 manifestación de  
 mantenimiento de  
 más de \_num\_  
 más de cinco años  
 más de diez años de  
 más de la mitad de  
 mayores de \_num\_  
 mayores de setenta años  
 mayoría absoluta  
 mayoría absoluta de  
 mayoría de sus miembros  
 mediante concurso  
 mediante sentencia  
 mediante una ley  
 medidas necesarias para  
 medidas provisionales  
 medio ambiente y  
 medios de comunicación social  
 mencionados en el  
 miembros del ministerio público  
 millón \_num\_  
 ministerio público  
 ministerio público de la unión  
 ministerio público y  
 ministro de estado  
 ministros de estado  
 ministros del supremo tribunal federal  
 misiones diplomáticas  
 mismas condiciones  
 modificada por la ec \_num\_/\_ano\_  
 modificado por la ec \_num\_/\_ano\_  
 municipios que tengan de \_num\_ a  
 \_num\_ habitantes  
 nacional de cultura  
 nacional de desarrollo  
 necesarias para su  
 necesarios para el  
 necesarios para la  
 necesarios para su  
 negociación colectiva  
 niños y adolescentes  
 no podrá exceder de  
 nombrados por el  
 nombrados por el presidente de la  
 república  
 nombrar y remover a los  
 normas generales  
 notable saber jurídico  
 número de diputados  
 obligaciones que  
 obligatorio para  
 observando las disposiciones  
 operaciones de crédito  
 oposición y méritos  
 orden constitucional  
 orden de los abogados de 101rasil  
 ordenamiento territorial  
 órdenes judiciales de pago  
 ordinaria de trabajo  
 organización de los  
 organización del  
 organización y funcionamiento  
 organizaciones de  
 órganos del poder judicial  
 órganos jurisdiccionales  
 órganos y entidades  
 otra nacionalidad

para asegurar el	patrimonio cultural
para el ejercicio de	personalidad jurídica
para el ejercicio de la	personas con discapacidad
para el ejercicio de sus	personas jurídicas
para garantizar el	pertenecientes a
para garantizar la	plataforma continental
para la designación	plazo de _num_ días
para la ejecución	pleno ejercicio de los derechos
para la ejecución de	poder legislativo
para la protección	poderes públicos
para la realización de	podrá establecer
para la seguridad	podrán establecer
para las personas	podrán participar
para los municipios que tengan de	podrán ser objeto de
_num_ a _num_ habitantes	por ciento de la retribución
para un mandato de	por ciento de los
párrafo _num_	por cualquiera de las
párrafo _num_ a los	por delitos políticos
párrafo _num_ al	por el congreso nacional
párrafo _num_ corresponde a	por el poder ejecutivo
párrafo _num_ de este artículo	por el poder público
párrafo _num_ del artículo _num_	por el presidente de la república
párrafo _num_ el estado	por el senado federal
párrafo _num_ el impuesto previsto en	por el supremo tribunal federal
el apartado _num_	por el territorio
párrafo _num_ el presidente	por el tiempo que
párrafo _num_ en los	por el vicepresidente
párrafo _num_ la ley complementaria	por la autoridad
párrafo _num_ la ley establecerá	por la cámara de diputados
párrafo _num_ la unión	por la ec _num_/_ano_
párrafo _num_ las disposiciones	por la ec de reforma _num_/_ano_
párrafo _num_ los estados	por la ley complementaria
párrafo _num_ no	por la mayoría absoluta
párrafo _num_ para	por la mayoría absoluta de
párrafo _num_ se prohíbe la	por las opiniones
párrafo _num_ si el	por ley complementaria
párrafo _num_ una ley	por los tribunales
párrafo anterior	por medio de los
párrafos _num_	por naturalización
párrafos _num_ y _num_	por un período de cuatro años
parte del patrimonio	presentado por el
participación de	preservación del
participación de los	presidente de la cámara de
participación en	presidente de la república
participación en la	presidente de la república de
participación en los	presidente de la república para
participar en el	presidente de la república se
participar en la	presidente del senado federal
partido político	presidente del tribunal
partidos políticos	

presidente y el vicepresidente de la república	que establezca la ley
presidente y vicepresidente de la república	que la ley determine
presidido por el	que se encuentren en
prestación de servicios	que se refiere el artículo _num_
prestaciones por	que se refiere el párrafo _num_
prestar servicios	que tengan de _num_ a _num_ habitantes
presten servicios	quedando prohibida la
presupuestarias y	quince por ciento)
presupuesto anual	receso del congreso
presupuesto de la	recursos humanos
previa autorización	recursos naturales
previa autorización del	recursos públicos
previsión complementaria	refiere el artículo _num_
previsión social	refiere el párrafo _num_
previstas en los	reforma _num_/_ano_
previsto en el apartado _num_	régimen general de previsión social
previstos en esta constitución	régimen jurídico
previstos en los	régimen propio de previsión social
previstos por la ley	regionales federales
principios de la	relacionadas con
privado de su libertad	relacionadas con la
procedimiento para	relacionados con
procedimientos para	relacionados con la
procedimientos que	relaciones exteriores
procesar y juzgar	relaciones internacionales
procurador general de la república	relaciones laborales
procuraduría general	rendición de cuentas
procuraduría general de la	renumerado a partir del
producto de la recaudación	representación de
producto de la recaudación de	representación del
programas de desarrollo	representación en
programas y proyectos	representación proporcional
promover el desarrollo	representantes del
promulgación de la	república federativa de 103rasil
propio de previsión social	reputación intachable
protección de las	residentes en el
protección del medio ambiente	respectiva cámara
protección especial	respectivo ente federativo
proyecto de ley de	respeto a los derechos
publicación de la	responsabilidad civil
públicas y privadas	responsabilidad de
públicos o privados	responsabilidad de los
públicos y privados	responsabilidad por
pueblos indígenas	responsable de la
que correspondan	responsables por
que determine la ley	retribuciones de los
que disponga la ley	salvo en los casos
que establece la ley	salvo lo dispuesto
	se aplicarán las

se fundamenta en  
 se garantiza el derecho  
 se refiere el artículo \_num\_  
 se refiere el párrafo \_num\_  
 se reunirá en sesiones  
 sección \_num\_  
 secretario de estado  
 sector público y  
 según corresponda  
 según lo determine  
 según lo dispuesto por  
 segunda instancia  
 seguridad nacional  
 seguridad pública  
 seguridad social  
 senadores y diputados  
 sentencia condenatoria  
 sentencia judicial  
 será nombrado por el  
 será obligatorio  
 serán nombrados por  
 serán regulados por  
 serán responsables  
 serán sancionados  
 servicio militar  
 servicio público  
 servicios auxiliares  
 servicios públicos  
 servidor público  
 sesión legislativa  
 si el presidente  
 siguientes funciones  
 siguientes principios  
 sin autorización  
 sin necesidad de  
 sin perjuicio de la  
 sin perjuicio de las  
 sin perjuicio de lo dispuesto en el  
 sin perjuicio de los  
 sin perjuicio del  
 sino en virtud de  
 sistema electoral  
 sistema financiero  
 sistema nacional de  
 soberanía nacional  
 sobre la base de  
 sobre la materia  
 sobre los asuntos  
 social y cultural  
 social y económico

sociedades de economía mixta  
 son iguales ante la ley  
 son inalienables  
 su funcionamiento  
 su organización y funcionamiento  
 su participación en  
 sujeto a lo dispuesto en  
 superior tribunal de justicia  
 supremo tribunal federal  
 supremo tribunal federal y  
 sus entidades autárquicas  
 sus instituciones  
 tecnología e innovación  
 tendrá derecho a  
 tendrán derecho a  
 tengan de \_num\_ a \_num\_ habitantes  
 tercios de los miembros  
 tercios de sus miembros  
 tiempo de cotización  
 tiempo de servicio  
 tiene el deber de  
 tienen derecho a  
 tienen derecho al  
 tienen el deber de  
 título \_num\_  
 todas las personas  
 todas las personas tienen derecho a  
 todos los ciudadanos  
 toma de posesión  
 tomarán posesión  
 total o parcialmente  
 transferencia de  
 tratados internacionales  
 tribunal competente  
 tribunal de cuentas de la unión  
 tribunal de justicia  
 tribunal regional  
 tribunal superior  
 tribunal superior del trabajo  
 tribunales de cuentas  
 tribunales de justicia  
 tribunales regionales  
 tribunales regionales federales  
 tribunales superiores  
 tribunales y jueces  
 un millón \_num\_  
 un plazo de \_num\_ días  
 un plazo máximo de  
 un procedimiento  
 un proyecto de ley



un régimen especial	y administrativas
un representante	y condiciones para
una _num_ parte de	y control social
una ley complementaria	y de acuerdo con
una ley especial	y de los servicios
una ley específica	y de los territorios
una sentencia judicial	y del distrito federal
unidades de la federación	y el distrito federal
votación directa	y el funcionamiento de
votación secreta	y la administración
y _num_ años	y las condiciones
y _num_ de este artículo	y los ministros de estado
y _num_ del artículo _num_	y reputación intachable
y a los municipios	y vicepresidente de la república
y administración	zona económica exclusiva
y administrativa	

## APÊNDICE C. Alinhamento entre unidades de tradução da Constituição Brasileira

Alinhamento das unidades de tradução da Constituição Brasileira mapeadas pelo *Grinder* até o artigo 3º.

Quadro 7 - Unidades de tradução da Constituição brasileira

#	Unidades de Tradução PT	#	Unidades de Tradução ES
1	constituição da república	1	constitución de la república
2	federativa do brasil.	2	federativa de brasil
3	preâmbulo	3	preámbulo
4	nós	4	nosotros,
5	representantes do	5	los representantes del
6	povo	6	pueblo
7	brasileiro,	7	brasileño,
8	reunidos	8	reunidos en
9	em		
10	assembleia nacional	9	asamblea nacional constituyente
11	constituente		
12	para	10	para
13	instituir	11	constituir
14	um estado	12	un estado
15	democrático,	13	democrático,
16	destinado		
17	a assegurar	14	que asegure
		15	el
18	o exercício dos direitos	16	ejercicio de los

#	Unidades de Tradução PT	#	Unidades de Tradução ES
19	sociais	17	derechos sociales
20	e	18	e
21	individuais,	19	individuales,
22	a	25	
23	liberdade,	20	la libertad,
24	a segurança,	22	seguridad,
25	o bem-estar,	23	el bienestar,
26	o desenvolvimento,	24	el desarrollo,
28	igualdade e	25	la igualdad
		26	y la
29	a justiça	27	justicia
30	como	28	como
31	valores	29	valores
32	supremos	30	supremos
		31	de una
33	de uma sociedade	32	sociedad
34	fraterna	33	fraterna
35	pluralista	34	pluralista
36	e		
37	sem	35	y sin
38	preconceitos	36	prejuicios
39	fundada	37	fundada en
40	na	38	la
41	harmonia	39	armonía
42	social e	40	social y
43	comprometida	41	comprometida
44	na		
45	ordem	42	en el orden
46	interna	43	interno
48	internacional,	45	internacional,
49	com a	46	con
50	solução	48	solución
51	pacífica	49	pacífica de
52	das		
53	controvérsias	50	las controversias,
54	promulgamos	51	promulgamos
55	sob a	52	bajo
56	proteção	54	protección de
57	de		
58	deus	55	dios,
60	seguinte	56	la siguiente
62	federativa do brasil.	58	federativa de brasil.
63	título i	59	título i
64	dos princípios	60	de los principios

#	Unidades de Tradução PT	#	Unidades de Tradução ES
65	fundamentais	61	fundamentales
66	art. 1º	62	artículo 1.
68	república federativa do brasil,	63	la república federativa de brasil,
69	formada	64	formada
		65	por la
70	pela união	66	unión
71	indissolúvel	67	indisoluble
72	dos	68	de los
73	estados	69	estados y municipios
75	municípios		
76	e do distrito federal,	70	y del distrito federal,
77	constitui-	71	se constituye
78	se		
79	em	72	en
80	estado democrático	73	estado
		74	democrático
		75	de
81	de direito	76	derecho y
83	tem como		
84	fundamentos	77	se fundamenta en:
85	i	78	i
86	a soberania;	79	la soberanía;
87	ii	80	ii
89	cidadania;	81	la ciudadanía;
90	iii	82	iii
91	a dignidade	84	dignidad de
92	da pessoa humana;	85	la persona humana;
93	iv	86	iv
94	os		
95	valores	87	los valores
96	sociais	88	sociales
97	do trabalho	89	del trabajo
98	e da	90	y de la
99	livre	91	libre
100	iniciativa;	92	iniciativa;
101	v	93	v
102	o pluralismo	94	el pluralismo
103	político.	95	político.
104	parágrafo único.	96	párrafo único.
105	todo o	97	todo el
106	poder	98	poder
107	emana	99	emana
108	do	100	del
109	povo,	101	pueblo,

#	Unidades de Tradução PT	#	Unidades de Tradução ES
110	que o	102	el cual
		103	lo
111	exerce	104	ejerce por
112	por meio de	105	medio de
113	representantes	106	representantes
114	eleitos	107	elegidos
115	ou	108	o
116	diretamente,	109	directamente,
117	nos termos	110	en los términos establecidos
		111	por
118	desta constituição.	112	esta constitución.
119	art. 2º	113	artículo 2.
120	são	116	son,
121	poderes da	114	los poderes
122	união,	115	de la unión
123	independentes e	117	independientes
		118	y
124	harmônicos	119	armónicos
125	entre	120	entre sí,
126	si		
127	o	121	el
128	legislativo,	122	legislativo,
130	executivo e	124	ejecutivo y
132	judiciário.	126	judicial.
133	art. 3º	127	artículo 3.
134	constituem	128	constituyen
135	objetivos	129	objetivos
136	fundamentais da	130	fundamentales de la
137	república federativa do brasil:	131	república federativa de brasil:
139	construir	133	construir
140	uma sociedade	134	una sociedad
141	livre,	135	libre,
142	justa e	136	justa
143	solidária;	137	y solidaria;
145	garantir	139	garantizar el
146	o desenvolvimento	140	desarrollo nacional;
147	nacional;		
149	erradicar	142	erradicar
150	a	143	la
151	pobreza	144	pobreza
152	e a	145	y la
153	marginalização	146	marginación
154	e	147	y
155	reduzir	148	reducir

#	Unidades de Tradução PT	#	Unidades de Tradução ES
156	as	149	las
157	desigualdades	150	desigualdades
158	sociais	151	sociales y
160	regionais	152	regionales;
162	promover o	154	promover
163	bem	155	el bien
164	de todos,	157	todos
165	sem	158	sin
167	de origem,	160	de origen,
168	raça,	161	raza,
169	sexo,	162	sexo,
170	cor,	163	color
171	idade	164	edad
173	quaisquer outras	165	o cualquier otra
174	formas de	166	forma de
175	discriminação.	167	discriminación.
176	art. 4º	168	artículo 4.
178	república federativa do brasil	169	la república federativa de brasil
179	rege	170	se rige por
181	nas suas	172	en sus
182	relações internacionais	173	relaciones internacionales:
183	pelos seguintes		
184	princípios:	171	los siguientes principios
186	independência nacional;	176	independencia
		177	nacional;
188	prevalência	180	prevalencia
189	dos direitos	181	de los derechos humanos;
190	humanos;		
192	autodeterminação	184	autodeterminación
		185	de los
193	dos povos;	186	pueblos
195	não-	188	la no
196	intervenção;	189	intervención;
198	igualdade entre	191	la igualdad
		192	entre los
199	os estados;	193	estados;
200	vi	194	vi
201	defesa da	195	la defensa de la
202	paz;	196	paz;
203	vii	197	vii
204		21	la
207	dos		
208	conflitos;	201	los conflictos;
209	viii	202	viii

#	Unidades de Tradução PT	#	Unidades de Tradução ES
210	repúdio	204	repudio
211	ao	205	al
212	terrorismo	206	terrorismo
213	e ao	207	y al
214	racismo;	208	racismo
215	ix	209	ix
216	cooperação	210	la cooperación
217	entre os povos	212	pueblos
218	para	213	para el
219	o progresso	214	progreso
220	da	215	de la
221	humanidade	216	humanidad;
222	x	217	x
223	concessão de	218	la concesión de
224	asilo	219	asilo
229	buscará	223	promoverá la
231	integração	224	integración
232	econômica,	225	económica,
233	política,	226	política,
234	social e cultural	227	social y cultural
235	dos povos	229	pueblos
236	da		
237	américa	230	de américa
238	latina	231	latina
239	visando	233	vistas
240	à	232	con
241	formação	234	a la formación de
242	de uma	235	una
243	comunidade	236	comunidad
244	latino	237	latinoamericana
245	americana		
246	de	238	de
247	nações	239	naciones

Fonte: Elaboração própria.