

# Esperanto and the Web: approaches and approximations

Cinthyana Renata Sachs C. de Barbosa<sup>1</sup>, João Coelho Neto<sup>2</sup>, Edio Roberto Manfio<sup>3</sup>

<sup>1</sup>Departamento de Computação – Universidade Estadual de Londrina (UEL)  
Caixa Postal 10.001 – 86.057-970 – Londrina – PR – Brazil

<sup>2</sup>Programa de Pós-Graduação em Ensino – Universidade Estadual do Norte de Paraná  
(UENP) 86.360-000 – Cornélio Procópio – PR – Brazil

<sup>3</sup>Faculdade de Tecnologia Deputado Júlio Julinho Marcondes de Moura (FATEC) –  
17.400-000 – Garça – SP – Brazil

cinthyana@uel.br, joaocoelho@uenp.edu.br, edio.manfio@fatec.sp.gov.br

***Abstract.** This article aims to analyze the approach of Esperanto in terms of Language Processing, in an attempt to specify its grammar, through the Lexicalized Context-Free Grammar, which is able to deal with linguistics issues of that language. The bibliographic method was used to identify that the Esperanto language can also achieve successful communication, especially on the Web, as millions of people have access to the network. Thus, the objective of this paper was to present the studies carried out for the comprehension of the Esperanto Language, collecting important details of its grammatical structure used in the implementation of a Planned Language Processor.*

***Resumo.** Este artigo visa analisar a aproximação do Esperanto no que tange ao Processamento da Língua, na tentativa de especificar sua gramática, por meio da Gramática Livre de Contexto Lexicalizada, capaz de tratar questões linguísticas da referida língua. O método utilizado foi o bibliográfico, identificando que a Língua Esperanto também pode alcançar êxito para a comunicação, em especial na Web, pois milhões de pessoas têm acesso à rede. Desse modo, o objetivo deste trabalho foi apresentar os estudos realizados para a compreensão da Língua Esperanto, reunindo detalhes importantes de sua estrutura gramatical usadas na implementação de um Processador de Língua Planificada.*

## 1. Introdução

Este trabalho fez parte das pesquisas de um projeto maior [Barbosa, Senefonte e Valente Neto 2011], que trata a Língua Esperanto quanto ao Processamento de Língua Planificada para especificar sua gramática por meio da Gramática Livre de Contexto Lexicalizada capaz de suportar o tratamento das questões linguísticas da referida língua.

A língua planejada Esperanto tem sido intensamente usada por meio de recursos disponíveis na web e algumas pessoas acreditam que essa possa alcançar êxito para a comunicação, em especial na Internet, pois o padrão Web é o esperanto que deu certo. É a língua universal que faz com que todos possam interagir [Teixeira Filho 2002]. A proposta do Esperanto não é a de substituir qualquer outra língua nacional, mas complementá-las, sendo assim utilizada como uma língua neutra quando falada com

alguém que não tenha a mesma língua do interlocutor [Liga Brasileira de Esperanto 2022].

Milhões de pessoas têm acesso à rede mundial de computadores e falando as mais distintas línguas. O Esperanto, nesse caso, seria fundamental, pois permitiria uma comunicação entre indivíduos de qualquer país e de qualquer idioma [Viana 2004]. Ainda que poucos, já existem falantes nativos do Esperanto, filhos de pais que falam Esperanto em casa por pertencerem a povos que manejam línguas diversas: pensam e falam apenas em Esperanto, que lhes serve excelentemente para a vida psíquica e social. O comportamento verbal desses indivíduos em nada se distingue do de outros que se expressam por línguas ditas naturais [Mattos 1987]. Muitas universidades em nosso país já possuem, de forma oficial e não oficial, os cursos da Língua Esperanto [Lopreato 1999]. É o caso da USP, UFJF, UFAL e UFC há mais de 50 anos.

Quanto aos trabalhos de Processamento de Linguagem Natural (PLN) que dão seus primeiros passos para a Língua Esperanto temos o programa Eoklavaro 1.2 que resolve os problemas para quem digita textos em Esperanto no sistema Windows [Servo 2021] e o software Lerni [Barbosa *et al.* 2018] criado com o objetivo de auxiliar docentes de Esperanto no ensino da língua para crianças de 8 a 12 anos. Tradutores automáticos para línguas como inglesa, alemã, polonesa, sueca e russa podem ser encontrados em Esperantilo (2013) que apesar de fazer um estudo inicial sobre sintaxe do Esperanto, marca sentenças corretas com falsas, como chama a atenção sobre esse fato na referida página do software. Assim, estudos envolvendo *parsing* em PLN para a Língua Esperanto são aqui propostos utilizando uma gramática de elevado poder computacional e linguístico que é a Gramática Livre de Contexto Lexicalizada (GLCL).

O procedimento metodológico para a coleta de dados escolhido para o desenvolvimento deste trabalho foi a pesquisa bibliográfica e documental. A pesquisa bibliográfica tem como principal característica o fato de que o campo onde será feita a coleta de dados é a própria bibliografia sobre o tema ou o objeto que se pretende investigar [Lourenço 2019].

Nesse contexto, este artigo analisa contribuições da Língua Esperanto, principalmente quanto à implementação de um sistema de Processamento de Linguagem Natural. Assim, o objetivo geral é apresentar estudos para a compreensão da Língua Esperanto, reunindo detalhes importantes de sua estrutura gramatical que estão sendo usadas na referida implementação do Processador da Língua Planificada. O uso de uma língua como Esperanto pode permitir uma comunicação neutra entre os diversos povos que se utiliza da Web e com isso, o aprendizado que se dá nessa troca de relações seria favorecido, gerando evidências para a área de Inteligência Artificial e Computacional. Lembrando que antes da web, aprender esperanto significava, geralmente, encomendar um livro de uma editora pouco conhecida ou talvez visitar escritórios de Esperanto que ainda estão abertos em algumas grandes cidades [DW 2008].

Na Seção 1 foi contextualizada a temática. Na Seção 2 serão abordadas as questões linguísticas do Esperanto. A Seção 3 tratará do aprendizado do Esperanto. Na Seção 4 um estudo rigoroso das regras gramaticais da referida língua será apresentado. A Seção 5 discutirá o Processamento dessa Língua Planificada e o uso da Gramática Livre de Contexto Lexicalizada para a geração de tal língua em sistema computacional. Por fim, as considerações finais estarão na Seção 6.

## 2. Diversidade Linguística versus Língua Franca

O inglês pode ser comparado ao latim para os povos periféricos a Roma na época do Império Romano [Cardoso 2001]. O domínio é, portanto, por interesses econômicos e políticos e não científicos. Porém, para diminuir essa barreira, outras interlínguas têm sido propostas, como por exemplo, a *Universal Networking Language* (UNL) que incorpora a capacidade de tradução automática de textos na *web*, por intermédio de ferramentas de software para codificação e decodificação de textos em línguas naturais [Chawla 2006]. Essas ferramentas são acopladas aos *browsers* disponíveis, de maneira a permitir que usuários produzam e leiam textos em sua própria língua [Oliveira Junior *et al.* 2001]. O projeto UNL inclui 15 línguas, cada uma das quais atribuídas a diferentes grupos de pesquisa e desenvolvimento espalhados pelos mundos árabe, alemão, chinês, espanhol, francês, hindi, indonésio, japonês, letão, mongol, português, russo e tailandês.

No entanto, segundo Ávila (2001), limitações têm sido encontradas no Sistema UNL, dentre essas, os fatos de não haver restrições de domínio, acarretando uma perda de expressividade semântica; do uso do inglês como língua franca para comunicação entre as equipes de desenvolvimento e como língua de representação do conhecimento; da ambiguidade inerente aos vocábulos da língua inglesa. A distinção de tais mapeamentos sintático-semânticos problemáticos é uma tarefa extremamente sensível ao contexto, muitas vezes ultrapassando os limites da sentença. As descrições da UNL não tratam desses níveis refinados de representação.

Diante de várias questões, dentre essas as que foram expostas anteriormente, sugere-se o uso da Língua Esperanto. É bom esclarecer que o Esperanto não pertence a nenhuma doutrina política ou religiosa. O movimento esperantista é composto de pessoas com ideias bem diversificadas [Santiago 1992].

## 3. O aprendizado do Esperanto

O médico oftalmologista L. L. Zamenhof (1859-1917) que nasceu na Lituânia e cresceu em Bliastyok criou o Esperanto. Ele conheceu uma cidade dividida religiosamente, economicamente e politicamente por quatro línguas diferentes: alemão, russo, polaco e ídiche. Por isso, não se compreendiam entre si, o que resultava em constrangimentos. Vendo esse fato, ele refletiu como poderia resolvê-lo. Assim em 1887 publicou com o pseudônimo “Doutor Esperanto”, o seu primeiro livro didático de sua “*Internacia Lingvo*” (Língua Internacional) com fins humanitários. O pseudônimo “Esperanto” foi forte e tornou-se o nome da língua mais bem sucedida da *língua planejada*, segundo Matthias (2003). O Esperanto é falado entre um e três milhões de pessoas em 120 países.

A engenhosidade do sistema de afixos é um traço das línguas naturais, marcantemente desenvolvida no Esperanto. Isso torna a língua ao mesmo tempo mais simples e mais rica, muito flexível, dando ampla margem de criatividade. Ao todo são uns 40 afixos. Jogando com eles, de um simples radical podem ser formadas mais de 50 palavras, e como essas também se formam pela composição de radicais, as possibilidades de criar e recriar modos de expressão parecem infinitas. Um vocabulário completo do Esperanto tem cerca de 8 mil radicais, o que parece bem pouco comparando com outras línguas. Porém não é, porque com esses elementos podem ser formadas não apenas as cerca de 80 mil palavras correspondentes às ditas “grandes

línguas”, como também muitas outras com significados bem definidos [Santiago 1992]. A consequência disso é que para compreender 84% de um texto em inglês, é necessário saber umas 1400 palavras [Santiago 1992]. Para o mesmo resultado com o mesmo texto em Esperanto bastam 550 radicais. Uma pessoa teria que saber 10 mil palavras para entender quase 100% do texto em inglês e no caso do Esperanto, 2800 radicais.

Neste contexto, o Esperanto é a mais falada das línguas planificadas. A intenção era criar uma língua de muito fácil aprendizagem, a qual servisse como “segunda” língua universal para toda a população mundial e não, como muitos supõem, para substituir todas as línguas existentes. É uma linguagem aglutinante, sem gêneros gramaticais para entidades assexuadas, sem conjugação de verbos variáveis por pessoa ou número e com três modos: indicativo, imperativo e subjuntivo. O vocabulário é baseado em várias línguas européias, principalmente latim, francês, alemão, inglês, e russo, com algumas palavras de outras línguas. A escrita é fonética e a morfologia é extremamente regular e fácil de aprender.

Com relação à educação do Esperanto, estudos comprovam que é a língua mais rápida de ser aprendida e na qual o estudante médio pode esperar alcançar o nível mais alto de proficiência [Janton 1993]. Sua aprendizagem tem demonstrado ser mais fácil que qualquer uma das línguas naturais, visto que, a língua apresenta “apenas 16 regras gramaticais” [Viana 2004], as quais serão descritas. Teixeira (2005) enfatiza que aprender Esperanto é relativamente mais fácil que aprender qualquer idioma nacional e equivale a romper de fato com as barreiras da diversidade linguística vigente no mundo.

O traço mais característico do Esperanto, e que torna seu aprendizado muito mais fácil quando comparado a outros idiomas, é a sua regularidade, onde cada letra possui um som e cada som apenas uma letra. Tais características exigem um esforço muito menor da memória auditiva quando se compara com o inglês, por exemplo, onde uma letra pode possuir diferentes sons [Pereira 2015].

Isso tudo é corroborado quando vimos em 2021 que foi realizado o 106º Congresso Universal de Esperanto [Revista Virtual de Esperanto 2021] com o tema Paz e Confiança: valores universais. Também no Brasil no mesmo ano houve o 55º Congresso Brasileiro de Esperanto junto com o 2º Congresso Virtual da Juventude Esperantista Brasileira [Esperanto 2021]. No cenário pandêmico em que estamos vivendo, o movimento esperantista precisou se reinventar: as diversas plataformas digitais cumpriram um papel essencial para que pudessem manter viva a tradição da multiculturalidade e da defesa de ideais de fraternidade universal entre todos os povos.

#### **4. Língua Esperanto e as Regras Gramaticais**

Desde o lançamento do primeiro vocabulário em 1905, com cerca de 2600 palavras intercombináveis, o Esperanto vem aumentando o repertório lexical paulatinamente, até apresentar, hoje, em seu principal dicionário de referência, o *Nova PIV 2005 – Plena Ilustrita Vortaro* (NPIV – Novo Dicionário Completo Ilustrado 2005), elaborado pela SAT – *Sennacieca Asocio Tutmonda* (Associação Apátrida Mundial), um total de 17000 verbetes, com 47000 unidades lexicais. É chamada a atenção [Dias 2007] que nem todas as palavras do Esperanto são empréstimos, sendo várias delas recombinações entre radicais já existentes para abarcar novos conceitos.

O alfabeto do Esperanto tem 28 letras onde cada uma delas tem um som. São elas: A, B, C, Ĉ, D, E, F, G, Ĝ, H, Ĥ, I, J, Ĵ, K, L, M, N, O, P, R, S, Ŝ, T, U, Ŭ, V, Z e em todas as palavras completas (isto é, não apostrofadas) a sílaba tônica é a penúltima, ou seja, são paroxítonas. Nas palavras apostrofadas, passa a ser a última. Cada vogal corresponde a uma sílaba na divisão silábica.

Os substantivos têm a terminação **o** em Esperanto, como *knabo* (garoto) e *tablo* (mesa). O artigo definido é **la**, *the* (o, a, os, as): *la knabo*: o garoto. Entretanto não há artigo indefinido nem para o plural nem para o singular [Reed 2004]. Já a utilização de adjetivos ocorre de forma análoga à língua inglesa [Reed 2004] como em *la granda cxevalo* (a casa grande) ou *forta kanbo* (um garoto forte). Os adjetivos são terminados em **a**, concordando em número e caso com os substantivos. O comparativo de superioridade é formado com o vocábulo *pli*, o superlativo com *plej*. O comparativo é seguido da conjunção *ol*, o superlativo da preposição *el*.

Chamamos de “afixos” no Esperanto o conjunto dos prefixos e sufixos. Em esperanto não temos “infíxos”, como acontece na Língua Portuguesa. Embora não seja estudado comumente, o infixo existe como na palavra “chaleira”, onde a letra “L” é um infixo e na palavra *motorneiro* que a letra “N” é um infixo. Eles foram necessários para a formação dessa palavra derivada de *motor*, que utilizaram o sufixo “eir-a/o” (masculino/feminino). No Esperanto isso não acontece. Os sufixos e prefixos são aplicados diretamente no radical das palavras. Um estudo bem detalhado desses são encontrados em Fernandes (1957).

Em Esperanto o sujeito sempre deve estar claramente indicado. As orações são formadas por sujeito, verbo e objeto [Saliba 2010]. Sempre o receptor da ação, normalmente o objeto direto, recebe a terminação **-n** (acusativo). Essa marcação cai sobre os substantivos, no adjetivo, no singular ou plural e, ainda nos pronomes pessoais e possessivos. Já a ordem dos elementos da oração é bem flexível como podemos ver em *la patro vokis la filon* (o pai chamou o filho). Pode-se reescrever essa sentença de pelo menos seis modos diferentes: *la patro la filon vokis*, *la filon la patro vokis*, *vokis la patro la filon*, *vokis la filon la patro*, *la filon vokis la patro*, *la patro vokis la filon*. Entretanto o acusativo só recai sobre orações que tem verbo transitivo. Alguns outros verbos não necessitam o acusativo, terminando a ação em si mesmo, que são os casos de verbos intransitivos, como em *la homo vivas* (o homem vive) ou *la birdo kantas* (o pássaro canta).

A terminação do modo infinitivo é **-i**: *ami* (amar), *paroli* (falar), *iri* (ir), *manĝi* (comer) e de todos os verbos são [Soledade 2008]: **-as** (presente), **-is** (passado), **-os** (futuro), **-us** (condicional), **-u** (imperativo). Todas as pessoas em todos os verbos são conjugadas da mesma forma. Não há verbos irregulares nem diferenças entre as pessoas verbais. O Esperanto é absolutamente regular e não há qualquer exceção de conjugação. Como todas as pessoas gramaticais têm a mesma conjugação, é imprescindível que o sujeito ou pronome pessoal sempre acompanhe o verbo como em *mi iris* (caminhei) ou *ni dancis* (dançávamos). Em Esperanto existem dois modos participiais (ativo e passivo) e cada um deles tem três tempos verbais (presente, passado e futuro). Nos participios ativo e passivo respectivamente o presente tem terminação **-anta** e **-ata**, o passado **-inta** e **-ita** e o futuro **-onta** e **-ota** [Soledade 2008].

Particípios são usados conforme os exemplos [Soledade 2008]: **Particípios Ativos:** *Mi estas manĝonta la pomo* (=eu estou para comer a maçã; a ação ocorrerá em um futuro próximo), *Mi estas manĝanta la pomo* (=eu estou comendo a maçã; a ação ocorre durante a enunciação), *Mi estas manĝinta la pomo* (=eu acabei de comer a maçã; a ação ocorreu em um passado próximo), *Mi estos manĝonta la pomo* (=eu estarei para comer a maçã; a ação ocorrerá em um futuro posterior a uma outra ação, também no futuro), *Mi estos manĝanta la pomo* (=eu estarei comendo a maçã; a ação ocorrerá, no futuro, simultaneamente a uma outra ação), *Mi estos manĝinta la pomo* (=eu terei comido a maçã; a ação ocorre no passado em relação a uma ação futura), *Mi estis manĝonta la pomo* (=eu estava para comer a maçã; a ação estava em iminência de ocorrer enquanto outra estava ocorrendo no passado), *Mi estis manĝanta la pomo* (=eu estava comendo a maçã; a ação ocorreu simultaneamente a outra, no passado), *Mi estis manĝinta la pomo* (=eu já tinha comido a maçã; a ação ocorreu em um passado anterior a outra, quem também ocorreu no passado) **Particípios Passivos:** *La pomo estas manĝota de mi* (=a maçã está para ser comida por mim), *La pomo estas manĝata de mi* (=a maçã está sendo comida por mim), *La pomo estas manĝita de mi* (=a maçã acabou de ser comida por mim), *La pomo estos manĝota de mi* (=a maçã está para ser comida por mim), *La pomo estos manĝata de mi* (=a maçã estará sendo comida por mim), *La pomo estos manĝita de mi* (=a maçã terá sido comida por mim), *La pomo estis manĝota de mi* (=a maçã estava para ser comida por mim), *La pomo estis manĝata de mi* (=a maçã estava sendo comida por mim), *La pomo estis manĝita de mi* (=a maçã tinha sido comida por mim). O único verbo auxiliar do esperanto é *esti*, e a preposição usada para o agente da voz passiva é sempre *de*.

Para distinguir os sujeitos dos objetos diretos, esses recebem a terminação **-n**. Essa é acrescentada a todas as palavras variáveis (substantivos, adjetivos, pronomes), tanto no singular quanto no plural que estejam nessa função [Soledade 2008]: *la kato manĝis la muson* (=o gato comeu o rato), *li legis bonajn librojn* (=ele leu bons livros), *mi trinkas varman kafon kun malvarman lakton* (=eu bebo café quente com leite frio), *mi amas vin* (=eu amo você) ou *ŝi troviĝis miajn ŝuojn* (=ela trouxe os meus sapatos).

Os verbos de ligação *ser*, *estar*, *permanecer*, *ficar* e *parecer* não chamam o acusativo, pois não retratam uma ação [Reed 2004] [Saliba 2010] como vimos em *la birdo estas bela* (o pássaro é belo) ou *la maro ŝajnas trankvila* (o mar parece tranquilo).

Os pronomes pessoais do Esperanto correspondem ao caso reto do Português que são *mi* (eu), *vi* (tu ou você), *li* (ele), *ŝi* (ela), *ĝi* (ele ou ela para coisa ou animal), *ni* (nós), *vi* (vós ou vocês), *ili* (eles ou elas). No Esperanto existe o pronome indefinido de terceira pessoa **oni**, que não tem correspondente direto em Português, mas se assemelha ao pronome *on* do francês. Exemplos de uso (Soledade, 2008): **oni donas donacoj en la Kristnasko** (dão-se presentes no Natal), **oni parolas en Esperanto en la klasoĉambro** (fala-se em esperanto na sala de aula; a gente fala em esperanto na sala de aula), **oni agos malsame kiam oni koni la vero** (agir-se-á diferente quando se souber da verdade; as pessoas agirão diferente quando souberem da verdade) e **oni rompis la fenestro** (quebraram a janela; alguém quebrou a janela). Existe também o pronome reflexivo de terceira pessoa *si*, que deve ser usado nas seguintes situações [Soledade 2008]: *li sin vundis* (ele se feriu; feriu-se a si mesmo) e *li portos la dokumento kun si* (ele levará o documento consigo).

Acrescenta-se o sufixo **-a** aos pronomes. No plural, acrescenta-se ainda a terminação **-j**. Não variam em gênero [Soledade 2008] como *mia* (meu ou minha), *via* (teu, tua ou seu, sua (de você)), *ŝia* (=seu, sua (dela)), *ĝia* (=seu, sua (do animal, da coisa)), *nia* (=nosso, nossa), *via* (=vosso, vossa (de vocês)), *ilia* (=seu, sua (dele, delas)) *sia* (reflexivo em 3ª pessoa; =seu próprio, sua própria (dele próprio, dela própria)). Os pronomes interrogativos são: *kio*, *kia* e *kiu*. (Associação Portuguesa de Esperanto, 2002): **Kio** pergunta (o que?), **Kia** pergunta (de que tipo?), **Kiu** pergunta (quem, quais?), **kiam** pergunta (quando?), **kioma** pergunta (que (quantas)?) *kioma horo estas? Estas la kvara* (Que (quantas) horas são?), **kiom** pergunta (quanto, quantos?).

Entretanto, quando se pergunta sobre a quantidade de um substantivo deve-se usar a preposição **da** como em *Kiom kostis la libro? Dek dolaroj* (Quanto custou o livro? 10 dólares) ou *Kiom da libroj vi aĉetis? Mi aĉetis tri librojn*.

**Ĉu** é usada para pergunta como Acaso, por acaso? Entretanto não é traduzido normalmente. Também é utilizado tornar as sentenças interrogativas como em *ĉu vi estas la instruisto?* (você é o professor?), *ĉu ilis scias esperanton?* (eles sabem esperanto? *ĝi estas nova, ĉu ne?* (ele/ela é novo/a, não é?), *li estas tre bela, ĉu vere?* (ele é muito bonito, não é verdade?).

Algumas ideias complementares são formadas com partículas especiais, como por exemplo, a partícula **ĉi** que reforça a ideia de proximidade e são usadas com as partículas **ti** e **ĉi**: *Tie* (ali, aí ou lá), ou *ĉi tie* ou *tie ĉi* (aqui ou cá). E a partícula **ajn** que aumenta a indefinição sugerindo o termo "quem/qual quer que seja": *Iu* (alguém ou alguém), *iu ajn* (qualquer ou qualquer um), *kies ajn* (de quem quer que seja) ou *kiel ajn* (onde quer que seja).

## 5. Processamento da Língua Planificada Esperanto

Para muitos a Língua Esperanto não é uma língua natural e sim uma língua planificada com o intuito de servir para a comunicação universal. Desse modo, seria mais adequado falarmos aqui de Processamento de Linguagem Planificada (PLP) e não Processamento de Linguagem Natural (PLN). Porém, mantendo embora a regularidade da sua origem artificial, o Esperanto tem se tornado uma língua natural [Mattos 1987]. Assim, este trabalho apresenta uma parcela do processo de construção de um sistema de PLP, a qual, como em qualquer projeto de PLN, é necessária a compreensão da linguagem para a construção da gramática.

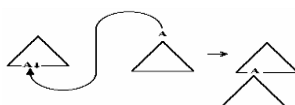
Para uma validação das técnicas de *parsing* e reconhecimento da Língua Esperanto foi utilizada uma Gramática Livre de Contexto Lexicalizada (GLCL), a qual pode ser encontrada nas regras gramaticais de Barbosa (2004) e Barbosa and Renna e Souza (2004), onde há bastante similaridade com a Língua Portuguesa no quesito de construções gramaticais. Essa apresenta um contexto adequado para tratar sintaxe, semântica e o seu relacionamento. Ela somente pode gerar uma linguagem livre de contexto (LLC), mantendo limites de complexidade de ordem cúbica no pior caso, para o *parsing* e reconhecimento [Schabes and Waters 1993b]. Gramáticas Livres de Contexto (GLCs) têm sido uma ferramenta bem aceita por linguistas computacionais há um longo tempo [Schabes 1993]. As GLCs com algumas restrições são as que tiveram maior progresso na descrição da Linguagem Natural.

Neste trabalho a motivação para o estudo das GLCL é de natureza linguística bem como formal. Usando objetos estruturados como objetos elementares de um formalismo, é possível, segundo Joshi and Schabes (1992), construir formalismos cujas propriedades relacionam diretamente à capacidade gerativa forte (descrição estrutural) a qual é mais relevante para descrições linguísticas do que capacidade gerativa fraca (conjunto de frases).

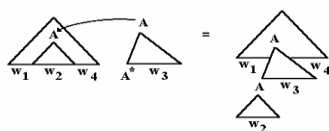
Do ponto de vista formal, a lexicalização permite associar cada estrutura elementar em uma gramática com um item lexical. O método apresentado a seguir nos permite obter lexicalização enquanto preserva a estrutura (lexicalização *forte*), o que é linguisticamente muito significativo. Já uma motivação linguística bastante importante para o uso de gramáticas da família *Tree Adjoining Grammar* (TAG), como é o caso da GLCL, é que essas gramáticas permitem fatorar a recursão proveniente da declaração das restrições linguísticas (dependências), tornando tais restrições estritamente locais, e assim simplificando a descrição linguística [Kroch and Joshi 1985].

Uma gramática *lexicalizada* é uma gramática onde toda regra contém um item lexical explícito (por exemplo, uma palavra no léxico) e essa consiste de: **a)** um conjunto finito de estruturas, cada uma associada com um item lexical; cada item lexical será chamado de *âncora* (necessariamente uma palavra não vazia) da estrutura correspondente; **b)** uma operação (ou operações) para compor as estruturas [Kroch and Joshi 1985]. O *léxico* consiste de um conjunto finito de estruturas onde cada uma é associada com uma âncora. Utiliza-se também o termo “lexicalizada” quando se fala sobre estruturas. Diz-se que uma estrutura é *lexicalizada* se há pelo menos um item lexical não vazio que aparece nesta [Joshi and Schabes 1992].

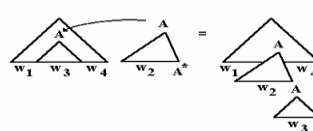
Informalmente falando, essa gramática consiste de dois conjuntos de árvores [Barbosa 1998]: as árvores iniciais, as quais são combinadas por *substituição* (Figura 1) e árvores auxiliares, as quais são combinadas por *adjunções* (Figuras 2 e 3). Quando as árvores são combinadas por essas operações, as representações semânticas são combinadas [Abeillé et al. 1992]. Uma GLCL é lexicalizada porque toda árvore inicial e auxiliar necessita conter uma palavra do dicionário em sua fronteira.



**Figura 1. Operação de Substituição na GLCL [Fonte: Schabes 1993]**



**Figura 2. Operação de Adjunção Recursiva à Esquerda na GLCL [Fonte: Schabes 1993]**

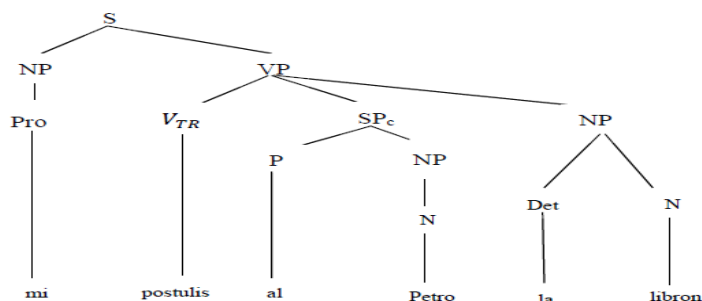


**Figura 3. Operação de Adjunção Recursiva à Direita na GLCL [Fonte: Schabes 1993]**

A GLCL é um sistema de geração de árvores ao invés de um sistema de geração de cadeias. O conjunto de árvores derivadas em uma GLCL constitui a linguagem objeto. De acordo com Joshi and Schabes (1992), essas árvores de derivação são importantes tanto sintática como semanticamente. A derivação nas GLCLs deve iniciar com uma árvore inicial rotulada em *S*. Depois disso, essa árvore pode ser repetidamente



estendida usando-se substituição e adjunção. Uma derivação é completa quando todo nó na fronteira é rotulado com um símbolo terminal [Schabes and Waters 1993b] formando uma árvore sintática com categorias semânticas como mostrada na Figura 4.



**Figura 4. Derivação da frase *mi postulis al Petro la libron* (Exigi de Pedro o livro)  
[Fonte: os autores]**

Alguém poderia dizer que a frase da Figura 4 seria ambígua olhando sua tradução no português, pois não se sabe se o livro pertence a Pedro ou se o sujeito da frase exigiu um livro qualquer que estava naquele momento com Pedro. Porém, isso não ocorreria no Esperanto, pois o uso da preposição “*al*” (ao invés, da preposição *de* que há também no esperanto) põe, definitivamente, fim à ambiguidade lexical, uma vez que a preposição “*al*” do Esperanto não admite significação de posse como admite a preposição “*de*”.

Quanto à ambiguidade estrutural, a Língua Esperanto sai privilegiada em relação à Língua Portuguesa e outras línguas [Barbosa, Senefonte e Coelho Neto 2013], pois a ordem dos elementos em esperanto é bastante livre e o acusativo ajuda muito, tanto na flexibilidade quanto na precisão. Ribeiro (2001) atenta para o fato da existência de ambiguidade sintática na Língua Esperanto ser com menor frequência devido à presença de marca de caso **n** (veja o objeto direto **libron** na árvore acima) que tem a função de destacar o objeto direto das orações.

Com relação ao corpus deste trabalho, assim como vimos em Minnaja and Paccagnella (2000), foram utilizadas várias fontes como de livros e artigos que abordam o Esperanto. Foi construído um *tagset*, de acordo com as características do Esperanto, onde na Figura 4, nó *NP* foi construído para sintagma nominal, *VP* para sintagma verbal, *Pro* para pronome, *Vtr* para verbo transitivo, *SP<sub>C</sub>* para sintagma preposicional, *N* para substantivo, *Det* para determinante e *S* para símbolo inicial.

## 6. Considerações Finais

Este artigo visou apresentar contribuições da Língua Esperanto principalmente no que tange a implementação de um sistema de PLN. Inicialmente foi descrita a diversidade linguística existente no mundo e as tentativas de se implantar uma língua de comunicação universal e qual o papel do Esperanto, nesse sentido. Fez-se um breve retrospecto da origem da Língua Esperanto, abordando seu processo histórico e a intenção de se fazer um processamento para essa língua. Finalmente, foi abordada a GLCL para a Língua Esperanto que foi a base do referido trabalho.

Desse modo, a língua planejada Esperanto não surgiu para substituir idiomas locais, mas sim para ser utilizada como uma língua padrão entre povos. A estrutura do

Esperanto é simples, sua gramática não apresenta irregularidades ou exceções e mesmo assim ela é muito rica e flexível.

O objetivo deste trabalho foi apresentar os estudos realizados para a compreensão da Língua Esperanto, reunindo detalhes importantes de sua estrutura gramatical que estão sendo usadas na implementação do Processador da Língua Planificada, pois o uso de uma língua como Esperanto permitirá uma comunicação neutra entre os diversos povos que utilizam a web e isso favoreceria o aprendizado nessa troca de relações.

Foi abordada a gramática para a descrição do Esperanto por meio da GLCL, onde operações de substituições são utilizadas para representações de sentenças simples, e de adjunções o são para estruturas recursivas de linguagem. Há algumas restrições impostas à adjunção que fazem com que essas gramáticas produzam somente LLCs e, em média, são mais poderosas que as GLCs. Shieber, Schabes & Pereira (1994) chamam à atenção sobre esse poder expressivo da adjunção, o qual permite que fenômenos da LN, como dependência de longa distância, sejam expressos localmente na gramática. Essas estruturas especificam *domínios de localidades estendidos* [Joshi and Schabes 1996] sobre quais restrições podem ser declaradas. Ou seja, uma árvore elementar codifica diretamente uma subestrutura sintática vista como uma árvore analisada parcialmente. Isso permite definir restrições em mais de um nível da análise da árvore quando comparada com regras livres de contexto e usar estruturas atômicas. Ainda, a noção de lexicalização na GLCL é também linguisticamente significativa por preservar não apenas os conjuntos de cadeias, mas também as estruturas, ou seja, apresenta *capacidade gerativa forte*.

## Referências

- Abeillé, A. and Schabes, Y. (1992). Non Compositional Discontinuous Constituents in Tree Adjoining Grammar. *Proceedings of IV European Summer School in Logic, Language and Information*, Colchester: University at Essex. p.1-20.
- Ávila, I. M. A. (2001). *Acessibilidade em Interfaces Computacionais da Web: Aspectos Lingüísticos*. Campinas: IC da UNICAMP.
- Barbosa, C. R. S. C. de. (2004). *Técnicas de Parsing da Gramática Livre de Contexto Lexicalizada da Língua Portuguesa*. São José dos Campos: CPG do ITA, 171p. Tese de Doutorado.
- Barbosa, C. R. S. C. de. (1998). *Gramática para Consultas Radiológicas em Língua Portuguesa*. Porto Alegre: Instituto de Informática da UFRGS, 143p. Dissertação de Mestrado.
- Barbosa, C. R. S. C., Renna e Souza, C. de. (2004). Reconhecedor de uma Gramática Livre de Contexto Lexicalizada. *Anais da XII Escola Regional de Informática da Sociedade Brasileira de Computação - Paraná*, Guarapuava: Unicentro.
- Barbosa, C. R. S. C., Senefonte, H. C. M. e Valente Neto, R. A. (2011). GLCL para a Língua Esperanto. *Anais do VII Seminário de Estudos sobre Linguagem e Significação*, Londrina: UEL.
- Barbosa, C. R. S. C., Senefonte, H. C. M. e Coelho Neto, J. (2013). Análise da Ambiguidade na Língua Esperanto. *Anais do VIII Seminário de Estudos sobre Linguagem e Significação*, Londrina: UEL.

- Barbosa, M. *et al.* (2018). Utilizando mecânicas de jogo para dinamizar o aprendizado da língua Esperanto. *Anais dos Workshops do VII Congresso Brasileiro de Informática na Educação*, Fortaleza: SBC. p.256-263.
- Cardoso, G. P. (2001). Questão da língua. In *Revista da Associação Médica Brasileira*. v.47, n.2, Scielo.
- Chawla, A. (2006). *UNL Punjabi Deconverter*. Patiala: Computer Science and Engineering Department of Thapar Institute of Engineering & Technology. 98p. Master Thesis.
- Dias, A. E. W. (2007). *Renovação Lexical do Esperanto: Mecanismos de Formação de Neologismos*. Brasília: Departamento de Linguística da UnB. 217p. Dissertação de Mestrado.
- DW – Deutsch Welle (2008). “Esperanto na Web”. <https://www.dw.com/pt-br/esperanto-encontra-nova-popularidade-atrav%C3%A9s-da-internet/a-3499611>
- Esperantilo (2013). “Esperantilo – text editor with particular Esperanto functions, spell and grammar checking and machine translation”. [http://www.esperantilo.org/index\\_en.html](http://www.esperantilo.org/index_en.html)
- Esperanto (2021). “55º Congresso Brasileiro de Esperanto e 2º Congresso Virtual da Juventude Esperantista Brasileira”. <https://agendaespiritabrazil.com.br/2021/07/03/55o-congresso-brasileiro-de-esperanto-e-o-2o-congresso-virtual-da-juventude-esperantista-brasileira-de-12-a-15-de-agosto/>, Agosto.
- Fernandes, A. C. e Mendes, H. M. (1957). *Curso de Esperanto em 20 lições*. Rio de Janeiro: Liga Brasileira de Esperanto.
- Janton, P. (1993). *Esperanto: Language, Literature and Community*. New York: State University of New York Press. 169p.
- Joshi, A. and Schabes, Y. (1996). *Tree-Adjoining Grammars*. Edited by Salomaa, A. and Rozenberg, G. In: *Handbook of Formal Languages and Automata*. Berlin, Springer-Verlag.
- Joshi. A. and Schabes. Y. (1992). *Tree-Adjoining Grammars and Lexicalized Grammars. IV European Summer School in Logic. Language and Information*, Colchester: University of Essex, p.1-23.
- Joshi. A., Vijay-Shanker. K. and Weir. D. (1991). “The Convergence of Mildly Context-Sensitive Grammatical Formalisms”. In: *Foundational Issues in Natural Language Processing*. Edited by Sells, P., Shieber, S.; Wasow, T. Cambridge, MA: MIT Press. p.31-81.
- Kroch, A. and Joshi, A. K. (1985). *Linguistic Relevance of Tree Adjoining Grammars*. Pennsylvania: Department of Computer and Information Science, University of Pennsylvania, Apr. Technical Report MS-CIS-85-18.
- Liga Brasileira de Esperanto (2022). “O que é o Esperanto?” <https://esperanto.org.br/info/index.php/18-disvastigado/5-o-que-esperanto>
- Lourenço, K. (2019). *Um Estudo Introdutório da Língua Esperanto*. Araguaína: Departamento de Letras da UFT. 99p. Trabalho de Conclusão de Curso.
- Matthias, U. (2003). *Esperanto: o novo latim da Igreja e do Ecumenismo*. Campinas: Pontes.105p.
- Mattos, G. (1987). A Organização do Esperanto. In: *Letras*. v.36, pages 26-41. UFPR.
- Minnaja, C. and Paccagella, L. G. (2000). A Part-of-Speech Tagger for Esperanto oriented to MT. *Proceedings of the International Conference on Machine*

- Translation and Multilingual Applications in the New Millennium: MT 2000*, Exeter: University of Exeter, UK. p.13-1–13-5.
- Oliveira Júnior, O. N, Martins, R. T., Rino, L. H. M. e Nunes, M, G, V. (2001). *O uso de Interlíngua para Comunicação via Internet: O Projeto UNL/Brasil*. São Carlos: ICMC da USP, 15p. Relatório NILC-TR-01-3.
- Oliveira Júnior, L. A. P. (1999). *Esperanto e Esperantista: na prática os sonhos não envelhecem ou A primeira fase do movimento esperantista (1887-1917)*. Uberlândia: Departamento de História da UFU, 53p. Trabalho de Conclusão de Curso.
- Pereira, M. G. (2015). O Esperanto como alternativa ao Imperialismo Linguístico do Inglês no Sistema ONU. *Anais do II Seminário de Relações Internacionais*, Caruaru: Faculdade ASCES – Associação Caruaruense de Ensino Superior e Técnico. <http://repositorio.asc.es.edu.br/handle/123456789/155>
- Reed, I. K. (2004). “Project Gutenberg: A complete Grammar of Esperanto”. <http://www.gutenberg.org/files/7787/7787-h/7787-h.htm>.
- Revista Virtual de Esperanto (2021). “O Esperanto Ilustrado”. <http://esperantoilustrado.blogspot.com/2021/04/congresso-universal-de-esperanto-sera.html>
- Ribeiro, M. B. (2001). *Propostas de Aplicações para Tratamento Lingüístico Computacional do Esperanto*. Campinas: IC da Unicamp, 36p. Notas de Trabalho da Disciplina Fatores Humanos em Sistemas de Informação.
- Saliba, A. (2010). “Minicurso: Esperanto: para um mundo moderno”. [www.scribd.com/.../esperanto-para-um-mundo-moderno-adonis-saliba](http://www.scribd.com/.../esperanto-para-um-mundo-moderno-adonis-saliba)
- Santiago, I. C. O. (1992). *O que é Esperanto: a questão da língua internacional*. São Paulo: Brasiliense. Coleção Primeiros Passos, número 185. 100p.
- Schabes, Y. (1993). *Lexicalized Context-Free Grammars*. Broadway, Cambridge. Mitsubishi Electric Research Laboratories, 16p. Technical Report TR 93-01.
- Schabes, Y. and Waters, R. C. (1993a). *Lexicalized Context-Free Grammars. Proceedings of XXXI Annual Meeting of the Association for Computational Linguistics*. Ohio: [s.n.]. p.121-129.
- Schabes, Y. and Waters, R. C. (1993b). *Lexicalized Context-Free Grammar: A Cubic-Time Parsable, Lexicalized Normal Form For Context-Free Grammar That Preserves Tree Structure*. Broadway, Cambridge: Mitsubishi Electric Research Laboratories. June. 30p. Technical Report 93-04.
- Servo, E. (2021). “Um programa para ajudar quem digita em Esperanto”. <https://esperanto.brazilo.org/archives/32035>
- Shieber, S. M., Schabes, Y. and Pereira, F. C. N. (1994). *Principles and Implementation of Deductive Parsing*. Broadway, Cambridge. Mitsubishi Electric Research Laboratories, 63p. Technical Report 94-08.
- Soledade, G. (2008). “Fundamenta Kurso de Esperanto”. <http://www.gabriel.soledade.nom.br/esperanto/licao5.htm>.
- Teixeira Filho, J. (2002). *Comunidades Virtuais: como as comunidades de práticas na Internet estão mudando os negócios*. Rio de Janeiro, SENAC. 183p.
- Teixeira, M. A. F. (2005). *Esperanto: baza kurso*. Florianópolis: UFSC. 391p.
- Viana, N. (2004). Linguagem, Poder e Relações Internacionais. In *Humanidades em Foco*, v.2, n.4, pages 10-19.